cemmap

centre for microdata methods and practice

Bounds on Treatment Effects on Transitions

Johan Vikström
Geert Ridder
Martin Weidner

The Institute for Fiscal Studies
Department of Economics, UCL

cemmap working paper CWP01/15

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

An ESRC Research Centre

Institute for Fiscal Studies, 7 Ridgmount Street, London WC1E 7AE **Tel:** 44 (0)20 7291 4800 **Email:** info@cemmap.ac.uk **www.cemmap.ac.uk**

Bounds On Treatment Effects On Transitions*

Johan Vikström† Geert Ridder‡ Martin Weidner§

January 8, 2015

Abstract

This paper considers identification of treatment effects on conditional transition probabilities. We show that even under random assignment only the instantaneous average treatment effect is point identified. Because treated and control units drop out at different rates, randomization only ensures the comparability of treatment and controls at the time of randomization, so that long run average treatment effects are not point identified. Instead we derive informative bounds on these average treatment effects. Our bounds do not impose (semi)parametric restrictions, as e.g. proportional hazards. We also explore various assumptions such as monotone treatment response, common shocks and positively correlated outcomes.

Keywords: Partial identification, duration model, randomized experiment, treatment effect
JEL classification: C14, C41

*We are grateful for helpful suggestions from John Ham, Per Johansson, Michael Svarer, Gerard van den Berg, Adam Rosen, and seminar participants at IFAU-Uppsala, Uppsala University, University of Aarhus and University of Mannheim. We acknowledge support from the Jan Wallander and Tom Hedelius Foundation, NSF grants SES 0819612 and 0819638, and from the the ESRC Centre for Microdata Methods and Practice grant RES-589-28-0001.

†Corresponding author: IFAU-Uppsala and UCLS at Uppsala University, Box 513, 751 20 Uppsala, Sweden; Email: johan.vikstrom@ifau.uu.se.

‡Department of Economics, Kaprilian Hall, University of Southern California, Los Angeles, CA 90089, U.S.A.; Email: ridder@usc.edu.

§Department of Economics, University College London, Gower Street, London WC1E 6BT, U.K., and CeMMAP; Email: m.weidner@ucl.ac.uk.

1 Introduction

We consider the effect of an intervention if the outcome is a transition from an initial to a destination state. The population of interest is a cohort of units that are in the initial state at the time origin. Treatment is assigned to a subset of the population either at the time origin or at some later time. Initially we assume that the treatment assignment is random. One main point of this paper is that even if the treatment assignment is random, only certain average effects of the treatment are point identified. This is because the random assignment of treatment only ensures comparability of the treatment and control groups at the time of randomization. At later points in time treated units with characteristics that interact with the treatment to increase/decrease the transition probability relative to similar control units leave the initial state sooner/later than comparable control units, so that these characteristics are under/over represented among the remaining treated relative to the remaining controls and this confounds the effect of the treatment.

The confounding of the treatment effect by selective dropout is usually referred to as dynamic selection. Existing strategies that deal with dynamic selection rely heavily on parametric and semi-parametric models. An example is the approach of Abbring and Van den Berg (2003) who use the Mixed Proportional Hazard (MPH) model (their analysis is generalized to a multistate model in Abbring, 2008). In this model the instantaneous transition or hazard rate is written as the product of a time effect, the effect of the intervention and an unobservable individual effect. As shown by Elbers and Ridder (1982) the MPH model is nonparametrically identified, so that if the multiplicative structure is maintained, identification does not rely on arbitrary functional form or distributional assumptions beyond the assumed multiplicative specification. A second example is the approach of Heckman and Navarro (2007) who start from a threshold crossing model for transition probabilities. Again they establish semi-parametric identification, although their model requires the presence of additional covariates besides the treatment indicator that are independent of unobservable errors and have large support.

In this paper we ask what can be identified if the identifying assumptions of the semi-parametric models do not hold. We show that, because of dynamic selection, even under (sequential) random assignment we cannot point identify most average treatment effects of interest. However, we derive sharp bounds on various non-point-identified treatment effects, and show under what conditions they are informative. Our bounds are general, since beyond random assignment, we make no assumptions on functional form and additional covariates, and we allow for arbitrary heterogenous treatment effects as well as arbitrary unobserved heterogeneity. The bounds can also be applied if the treatment assignment is unconfounded by creating bounds conditional on the covariates (or the propensity score) that are averaged over the distribution of these covariates (or propensity score).

Besides these general bounds we derive bounds under additional (weak) assumptions like monotone treatment response and positively correlated outcomes. We relate these assumptions to the assumptions made in the MPH model and to assumptions often made in discrete duration models and structural models. The additional assumptions often tighten the bounds considerably. We also discuss how to apply our various identification results to construct asymptotically valid confidence intervals for the respective treatment effects.

There are many applications in which we are interested in the effect of an intervention on transition probabilities/rates. The Cox (1972) partial likelihood estimator is routinely used to estimate the effect of an intervention on the survival rate of subjects. Transition models are used in several fields. Van den Berg (2001) surveys the models used and their applications. These models also have been used to study the effect of interventions on transitions. Examples are Ridder (1986), Card and Sullivan (1988), Bonnal et al. (2007) , Gritz (1993), Ham and LaLonde (1996), Abbring and Van den Berg (2003), and Heckman and Navarro (2007). A survey of models for dynamic treatment effects can be found in Abbring and Heckman (2007).

An alternative to the effect of a treatment on the transition rate is its effect on the cdf of the time to transition or its inverse, the quantile function. This avoids the problem of dynamic selection. From the effect on the cdf we can recover the effect on the average duration, but we cannot obtain the effect on the conditional transition probabilities, so that the effect on the cdf is not informative on the evolution of the treatment effect over time. This is a limitation since there are good reasons why we should be interested in the effect of an intervention on the conditional transition probability or the transition/hazard rate. One important reason is the close link between the hazard rate and economic theory (Van den Berg (2001)). Economic theory often predicts how the hazard rate changes over time. For example, in the application to a job bonus experiment considered in this paper labor supply and search models predict that being eligible for a bonus if a job is found, increases the hazard rate from unemployment to employment. According to these models there is a positive effect only during the eligibility period, and the effect increases shortly before the end of the eligibility period. The timing of this increase depends on the arrival rate of job offers and is an indication of the control that the unemployed has over his/her reemployment time. Any such control has important policy implications. This can only be analyzed by considering how the effect on the hazard rate changes over time.

The evolution of the treatment effect over time is of key interest in different fields. For instance, consider two medical treatments that have the same effect on the average survival time. However, for one treatment the effect does not change over time while for the other the survival rate is initially low, e.g. due to side effects of the treatment, while after that initial period the survival rate is much higher. As another example research on the effects of active labor market policies (ALMP), often documents a large negative lock-in effect and a later positive effect once the program has been completed, see e.g. the survey by Kluve et al. (2007).

We apply our bounds and confidence intervals to data from a job bonus experiment previously analyzed by e.g. Meyer (1996). As discussed above economic theory has specific predictions for the dynamic effect of a re-employment bonus with a finite eligibility period. Meyer (1996) estimates these dynamic effects using an MPH model. We study what can be identified if we rely solely on random assignment and some additional (weak) assumptions. We confirm the effects predicted by theory so that these are not an artefact of the MPH assumption.

In section 2 we define the treatment effects that are relevant if the outcome is a transition. Section 3 discusses their point or set identification in the case that the treatment is randomly assigned. This requires us to be precise on what we mean by random assignment in this setting. In section 4 we explore additional assumptions that tighten the bounds. In section 5 we derive the confidence intervals. Section 6 illustrates the bounds for the job bonus experiment. Section 7 concludes.

2 Treatment effects if the outcome is a transition

2.1 Parametric outcome models

To set the stage for the definition of a treatment effect on a transition, we consider the effect of an intervention in the Mixed Proportional Hazards (MPH) model. The MPH model specifies the individual hazard or transition rate $\theta(t, d(t), V)$

$$\theta(t, d(t), V) = \lambda(t)\gamma(t-\tau, \tau)^{d(t)}V$$

with t the time spent in the origin state, $\lambda(t)$, the baseline hazard, $d(t)$, the treatment indicator function at time t, and V, a scalar nonnegative unobservable that captures population heterogeneity in the hazard/transition rate and has a population distribution with mean 1. If treatment starts at time τ then $d(t) = I(t > \tau)$, i.e. we assume that treatment is an absorbing state. The nonnegative function $\gamma(t-\tau, \tau)$ captures the effect of the intervention, an effect that depends on the time until the treatment starts τ and the time treated $t-\tau$. Finally, although γ is common to all units, the effect of the intervention differs between the units, because it is proportional to the individual V. The ratio of the treated and non-treated transition rates for a unit with unobservable V is $\gamma(t-\tau, \tau)$ for $t > \tau$, so that in the MPH model $\gamma(t-\tau, \tau)$ is the proportional effect of the intervention on the individual transition rate.

Let $\bar{d}(t) = \{d(s), 0 \le s \le t\}$ be the treatment status up to time t. The MPH model implies that the population distribution of the time to transition $T^{\bar{d}(T)}$ where the superscript is the relevant treatment history[1], has density

$$f(t|\bar{d}(t)) = \mathbb{E}_V\left[V\lambda(t)\gamma(t-\tau,\tau)^{d(t)}e^{-\int_0^t \lambda(s)\gamma(s-\tau,\tau)^{d(s)}V\mathrm{d}s}\right],$$

and distribution function

$$F(t|\bar{d}(t)) = 1 - \mathbb{E}_V\left[e^{-\int_0^t \lambda(s)\gamma(s-\tau,\tau)^{d(s)}V\mathrm{d}s}\right].$$

The hazard/transition rate given the treatment history is

$$\theta(t|\bar{d}(t)) = \lambda(t)\gamma(t-\tau,\tau)^{d(t)}\mathbb{E}_V\left[V|T^{\bar{d}(T)} \ge t\right].$$

To define treatment effects in the MPH model we compare groups with different treatment histories $\bar{d}(t)$. Let $\bar{d}_0(t)$ and $\bar{d}_1(t)$ be two such histories. We can compare either the average time-to-transition distribution functions in t, i.e. $F(t|\bar{d}_0(t))$ and $F(t|\bar{d}_1(t))$, or the average transition rates in t, i.e. $\theta(t|\bar{d}_0(t))$ and $\theta(t|\bar{d}_1(t))$. The comparison of the average transition rates is conditional on survival in the initial state up to time t and the comparison of the average distribution functions is not conditional on survival. As a consequence if we compare distribution functions we average over the population distribution of V, but if we compare transition rates we average over the distribution of V for the subpopulation of survivors up to time t.

[1] In this case the treatment history is fully characterized by τ, but we use the more general notation to accommodate other dynamic treatments.

Let us take $\overline{d}_0(t) = 0$, i.e. the unit is in the control group during $[0, t]$, and $\overline{d}_1(t)$ such that treatment starts at time τ. Then $F(t|\overline{d}_1(t)) > F(t|\overline{d}_0(t))$ if and only if

$$\frac{1}{\int_\tau^t \lambda(s)\mathrm{d}s} \int_\tau^t \lambda(s)\gamma(s-\tau,\tau)\mathrm{d}s > 1, \tag{1}$$

i.e. if a λ weighted average of the effect on the individual transition rate is greater than 1. This time average hides the change in the treatment effect over the spell. Note that the comparison of the distribution functions is not confounded by differences in the distribution of the unobservable V between the treatment and control groups. However, if we compare the transition rates in $t > \tau$

$$\theta(t|\overline{d}_0(t)) = \lambda(t)\mathbb{E}_V\left[V|T^{\overline{d}_0(T)} \geq t\right],$$

and

$$\theta(t|\overline{d}_1(t)) = \lambda(t)\gamma(t-\tau,\tau)\mathbb{E}_V\left[V|T^{\overline{d}_1(T)} \geq t\right],$$

then because

$$\mathbb{E}_V\left[V|T^{\overline{d}_0(T)} \geq t\right] > \mathbb{E}_V\left[V|T^{\overline{d}_1(T)} \geq t\right],$$

if and only if (1) holds, we have that under that condition

$$\frac{\theta(t|\overline{d}_1(t))}{\theta(t|\overline{d}_0(t))} < \gamma(t-\tau,\tau).$$

Therefore if the intervention increases the transition rate on average (as in (1)), then the ratio of the average treated and control transition rates is strictly smaller than that of the individual treated and control transition rates. If the intervention decreases the transition rate on average, then the ratio of the average treated and control transition rates is strictly larger than that of the individual rates. Hence, the effect of the intervention on the transition rate is confounded by its differential effect on the distribution of the unobservable among the treated and controls. The intuition behind this result is that the difference between the treated and control transition rates is proportional to V and this difference determines the survival probability. Therefore if (1) holds, for all values of V the survival probability is smaller for the treated than for the controls and the difference is largest for large values of V. Therefore the average V among the survivors will be smaller for the treated than for the controls and this makes that the comparison of the average transition rates of the treated and controls is confounded by the dynamic selection. This dynamic selection or survivor bias is not just a feature of the MPH model. It is present in any population where the treatment and the individual characteristics interact to increase or decrease the transition probability.

Parametric and semi-parametric models for the transition can be used to correct for the survivor bias in the average treatment effect. In a fully specified MPH model we specify a distribution for V , so that we can estimate $\mathbb{E}_V\left[V|T^{\overline{d}_0(T)} \geq t\right]$ and $\mathbb{E}_V\left[V|T^{\overline{d}_1(T)} \geq t\right]$ to obtain the correction factor. The MPH model is nonparametrically identified so that the parametric assumptions can be relaxed. However, that requires that we maintain the multiplicative specification with a proportional unobservable. As argued by e.g. Van den Berg (2001) economic models for the hazard rate usually are not multiplicative. In general, such models have multiple unobservables that enter in a nonseparable way. Other (semi)parametric models for dynamic selection as that of Heckman and Navarro (2007) also require strong maintained assumptions, i.e. the inclusion of additional covariates that are assumed to be independent of the unobservables (and this assumption cannot be justified by a reference to randomization) and that have large support. Given the strong assumptions that are needed to correct for dynamic selection using parametric or semi-parametric models, it is important to know whether the causal effect of a treatment can be identified without these maintained assumptions.

2.2 Average treatment effect on transitions

We discuss the definition and identification of treatment effects on transition rates in discrete time. The definition of causal effects in continuous time adds technical problems (see e.g. Gill and Robins (2001)) that would distract from the conceptual issues. From now on we assume that transitions occur at times $t = 1, 2, \ldots$.

We denote the treatment indicator in period t by d_t and the treatment history up to and including period t by $\overline{d}_t$. Let the potential outcome $Y_t^{\overline{d}_t}$ be an indicator of a transition in period t if the treatment history up to and including t is $\overline{d}_t$. If treatment is an absorbing state, $\overline{d}_t$ is a sequence of 0-s until treatment starts in period τ and the remaining values are 1. It is possible that $\tau = \infty$, the unit is never treated, or $\tau = 1$, the unit is always in the treated state.

In any definition of the causal effect of a treatment on the transition probability/rate we must account for dynamic selection. If we do not specify a model for the transition probability/rate we need to find another way to maintain the comparability of the treatment and control groups over the spell. The approach that we take in this paper is to consider average transition probabilities/rates where the average is taken in the same population for both treated and controls (or in general for different treatment arms). The (semi)parametric models implicitly do this as well. For instance, in the MPH model the average treatment effect is $\gamma(t-\tau,\tau) = \gamma(t-\tau,\tau)\mathbb{E}(V)$. This is the average treatment effect if the composition of the population would not change over time due to drop out. Because the population composition does not change, in this hypothetical population the initial balance between the treated and controls is maintained as well.

As emphasized we are interested in conditional treatment effects, i.e. treatment effects defined for the survivors in t. Let $\overline{d}_{1t}$ and $\overline{d}_{0t}$ be two specific treatment histories. If we do not maintain comparability of the treatment and control groups by hypothetically shutting down any dynamic selection, i.e. by averaging over the population at time 0, we have to define a subpopulation of the treated and controls that has the same composition. To define the average treatment effect on the transition probability/rate at t we, initially, propose to average over the subpopulation of individuals who would have survived until time t under $\overline{d}_{1t}$. This is the analogue of the average effect on the treated considered in the static treatment effect literature. This leads to the following definition

Definition 1 *The causal effect of $\overline{d}_{1t}$ relative to $\overline{d}_{0t}$ on the transition probability in t is the*

Average Treatment Effect on Treated Survivors (ATETS) defined by

$$\begin{aligned}&\text{ATETS}_t^{\overline{d}_{1t},\overline{d}_{0t}} \\ &\quad= \mathbb{E}\left(Y_t^{\overline{d}_{1t}}|Y_{t-1}^{\overline{d}_{1,t-1}}=0,\ldots,Y_1^{\overline{d}_{11}}=0\right)-\mathbb{E}\left(Y_t^{\overline{d}_{0t}}|Y_{t-1}^{\overline{d}_{1,t-1}}=0,\ldots,Y_1^{\overline{d}_{11}}=0\right).\end{aligned} \tag{2}$$

Obvious choices for $\overline{d}_{1t}$ and $\overline{d}_{0t}$ are $\overline{d}_{1t}=(0,\ldots,0,1,\ldots,1)$ with the first 1 at position τ, and $\overline{d}_{0t}=(0,\ldots,0)$. If we make the usual assumption that there is no effect of the treatment before it starts[2], then for these two treatments $\text{ATETS}_t^{\overline{d}_{1t},\overline{d}_{0t}}=0, t=1,\ldots,\tau-1$. The differential selection only starts after the treatment begins, so that this property of the ATETS_t is consistent with that fact. After the treatment starts there is dynamic selection and the $\text{ATETS}_t^{\overline{d}_{1t},\overline{d}_{0t}}$ controls for that by comparing the transition rates for individuals with a common survival experience.

Note that we are only concerned with the comparability of the treatment and control groups over the spell, i.e. with the different levels of dynamic selection in the two groups. If we keep the treatment and control groups comparable over time, there is still the question how to interpret the time path of the average treatment effect over the spell. In this paper we do not try to decompose this path into the average treatment effect for a population of unchanging composition and a selection effect relative to this population. We do not define the treatment effect for this population, but rather for a population that changes over time due to dynamic selection, but the dynamic selection is made equal in the treatment and control group, so that the treatment effect is not confounded by dynamic selection. Again this is analogous to the difference between the Average Treatment Effect and the Average Treatment Effect on the Treated in the case of a static treatment effect where the latter is defined for the population selected for treatment and the treatment effect is for this selective population.

We also consider the average effect when averaging over the subpopulation of individuals who would have survived until t under both $\overline{d}_{0t}$ and $\overline{d}_{1t}$

Definition 2 *The causal effect of $\overline{d}_{1t}$ relative to $\overline{d}_{0t}$ on the transition probability in t is the Average Treatment Effect on Survivors (ATES) defined by*

$$\begin{aligned}\text{ATES}_t^{\overline{d}_{1t},\overline{d}_{0t}} &= \mathbb{E}\left(Y_t^{\overline{d}_{1t}}|Y_{t-1}^{\overline{d}_{1,t-1}}=0,\ldots,Y_1^{\overline{d}_{11}}=0,Y_{t-1}^{\overline{d}_{0,t-1}}=0,\ldots,Y_1^{\overline{d}_{01}}=0\right) \\ &\quad-\mathbb{E}\left(Y_t^{\overline{d}_{0t}}|Y_{t-1}^{\overline{d}_{1,t-1}}=0,\ldots,Y_1^{\overline{d}_{11}}=0,Y_{t-1}^{\overline{d}_{0,t-1}}=0,\ldots,Y_1^{\overline{d}_{01}}=0\right).\end{aligned} \tag{3}$$

This average effect is discussed in subsection 3.2.3. Below we derive partial identification results for both ATETS and ATES.

3 Identification of average treatment effects on transitions under random assignment

We now consider identification of the $\text{ATETS}_t^{\overline{d}_{1t},\overline{d}_{0t}}$ under random treatment assignment. We first need to define what we mean by random assignment in this case. Let D_t be the indicator

[2] Abbring and Van den Berg (2003) call this the no-anticipation assumption.

that treatment is assigned in period t, i.e. the unit is not treated in periods $1,\ldots,t-1$, selected for treatment in period t and, because treatment is assumed to be an absorbing state, remains in the treated state in the subsequent periods. We assume that the treatment is assigned at the beginning of the period, so that the treated responses are observed in periods $t, t+1, \ldots$. The control treatment $\overline{d}_{0t}$ is no treatment up to and including t. We distinguish between three types of randomized assignment

(i) **Random assignment of the time of treatment** For all t and $\overline{d}_s, s = 1, 2, \ldots$,

$$D_t \perp \left\{ Y_s^{\overline{d}_s}, s = 1, 2, \ldots \right\}.$$

(ii) **Sequential randomization** For all t and $\overline{d}_s, s \geq t$, with the first $t-1$ components equal to 0,

$$D_t \perp \left\{ Y_s^{\overline{d}_s}, s = t, t+1, \ldots \right\} \Big| D_{t-1} = 0.$$

(iii) **Sequential randomization among survivors** For all t and $\overline{d}_s, s \geq t$, with the first $t-1$ components equal to 0,

$$D_t \perp \left\{ Y_s^{\overline{d}_s}, s = t, t+1, \ldots \right\} \Big| D_{t-1} = 0,\ Y_{t-1}^0 = \cdots = Y_1^0 = 0.$$

Under (i), the period in which the unit enters the treated state is randomly assigned. This can be implemented at time 0 and a consequence is that some units may have left the initial state by the time their treatment starts. Under (ii) treatment is assigned randomly in period t to units that have not been treated before. Again this will select units for treatment that have left the initial state. Under (iii) the randomization is among the non-treated survivors. Note for $t = 1$ this assumption implies that $D_1 \perp Y_s^{\overline{d}_s}, s \geq 1$. Random assignment of the time of treatment implies sequential randomization, which implies sequential randomization among survivors. In this paper, we focus on identification of average treatment effects under sequential randomization.

Assumption 1 *Treatment assignment is by sequential randomization among survivors.*

Initially we consider the two period case where the transition occurs in period 1, period 2 or after period 2. The main results of this paper can be illustrated in this setting. We discuss the extension to an arbitrary number of periods in section 3.2.2. For every member of the population we have a vector of potential outcomes $Y_1^1, Y_1^0, Y_2^{11}, Y_2^{01}, Y_2^{00}$, and a vector of treatment indicators D_1, D_2. Let Y_t be the observed indicator of a transition in period t. These observed outcomes Y_1, Y_2 are related to the potential outcomes by the observation rules

$$Y_1 = D_1 Y_1^1 + (1 - D_1) Y_1^0, \tag{4}$$

and

$$Y_2 = D_1 Y_2^{11} + (1 - D_1) D_2 Y_2^{01} + (1 - D_1)(1 - D_2) Y_2^{00}. \tag{5}$$

Because treatment is an absorbing state we have

$$D_1 = 1 \quad \Rightarrow \quad D_2 = 1.$$

Assumption 1 is in this case

$$D_1 \perp Y_1^1, Y_1^0, Y_2^{11}, Y_2^{01}, Y_2^{00}, \tag{6}$$

and

$$D_2 \perp Y_2^{01}, Y_2^{00} \;\big|\; D_1 = 0, Y_1^0 = 0. \tag{7}$$

Hence, under this assumption we can relate the observed and potential transition probabilities.

Lemma 1 *If Assumption 1 holds, then*

$$\mathbb{E}(Y_1|D_1 = 1) = \mathbb{E}(Y_1^1), \tag{8}$$

$$\mathbb{E}(Y_1|D_1 = 0) = \mathbb{E}(Y_1^0), \tag{9}$$

$$\mathbb{E}(Y_2|Y_1 = 0, D_1 = 1) = \mathbb{E}(Y_2^{11}|Y_1^1 = 0), \tag{10}$$

$$\mathbb{E}(Y_2|Y_1 = 0, D_1 = 0, D_2 = 0) = \mathbb{E}(Y_2^{00}|Y_1^0 = 0), \tag{11}$$

$$\mathbb{E}(Y_2|Y_1 = 0, D_1 = 0, D_2 = 1) = \mathbb{E}(Y_2^{01}|Y_1^0 = 0). \tag{12}$$

Proof See Appendix.

3.1 Identification of instantaneous treatment effects

The interpretation of the $\text{ATETS}_t^{\overline{d}_{1t},\overline{d}_{0t}}$ depends on the treatments $\overline{d}_{0t}, \overline{d}_{1t}$. We distinguish between instantaneous or short-run effects and dynamic or long-run effects. Throughout $\overline{d}_{0t}$ means no treatment up to and including t. The instantaneous effect is the ATE in the first period of treatment. With two periods in which the treatment can start the two instantaneous treatment effects are

$$\text{ATETS}_1^{1,0} = \mathbb{E}(Y_1^1) - \mathbb{E}(Y_1^0),$$

and

$$\text{ATETS}_2^{01,00} = \mathbb{E}(Y_2^{01}|Y_1^0 = 0) - \mathbb{E}(Y_2^{00}|Y_1^0 = 0).$$

Under Assumption 1 it follows from equations (8) and (9) that we can point identify the first period instantaneous treatment effect

$$\text{ATETS}_1^{1,0} = \text{ATE}_1^{1,0} = \mathbb{E}(Y_1^1) - \mathbb{E}(Y_1^0) = \mathbb{E}(Y_1|D_1 = 1) - \mathbb{E}(Y_1|D_1 = 0),$$

and from equations (11) and (12) that we can point identify the second period instantaneous treatment effect

$$\begin{aligned}\text{ATETS}_2^{01,00} &= \mathbb{E}(Y_2^{01}|Y_1^0 = 0) - \mathbb{E}(Y_2^{00}|Y_1^0 = 0) \\ &= \mathbb{E}(Y_2|Y_1 = 0, D_1 = 0, D_2 = 1) - \mathbb{E}(Y_2|Y_1 = 0, D_1 = 0, D_2 = 0).\end{aligned}$$

3.2 Bounds on dynamic treatment effects on transitions

3.2.1 Two periods

With two periods the dynamic treatment effect is the effect in period 2 of a treatment started in period 1 relative to no treatment in both periods. The relevant ATETS is therefore

$$\text{ATETS}_2^{11,00} = \mathbb{E}(Y_2^{11}|Y_1^1 = 0) - \mathbb{E}(Y_2^{00}|Y_1^1 = 0),$$

that is the average treatment effect in the second period from treatment started in the first period for those who survive under treatment in the first period.

Because all that can be deduced from the data is in equations (8)-(12), which hold under Assumption 1, $\text{ATETS}_2^{11,00}$ is, in general, not point identified. However, the observed transition probabilities place restrictions on the potential ones. We use these restrictions to derive sharp bounds on $\text{ATETS}_2^{11,00}$. The bounds are sharp in the sense that there exist feasible joint distributions of the potential outcomes which are consistent with the upper bound and the lower bound.

The first step is to characterize the joint distribution of the potential outcomes. Note that because treatment is an absorbing state, Y_2^{10} is not defined. This means that the joint distribution of $Y_1^0, Y_1^1, Y_2^{00}, Y_2^{01}, Y_2^{11}$ can be fully characterized by the probabilities

$$\begin{aligned} p(y_1^1, y_1^0) &\equiv \Pr(Y_1^1 = y_1^1, Y_1^0 = y_1^0), \qquad\qquad y_1^1, y_1^0 = 0,1, \\ p(y_2^{11}, y_2^{01}, y_2^{00}|y_1^1, y_1^0) &\equiv \Pr(Y_2^{11} = y_2^{11}, Y_2^{01} = y_2^{01}, Y_2^{00} = y_2^{00}|Y_1^1 = 0, Y_1^0 = 0), \\ &\qquad\qquad y_1^1, y_1^0, y_2^{11}, y_2^{01}, y_2^{00} = 0,1. \end{aligned}$$

The $\text{ATETS}_2^{11,00}$ can be expressed as a function of these probabilities.

By Assumption 1 the observed first period transition probabilities impose the restrictions

$$\Pr(Y_1 = y_1|D_1 = 1) = \sum_{y_1^0=0}^{1} p(y_1, y_1^0), \tag{13}$$

and

$$\Pr(Y_1 = y_1|D_1 = 0) = \sum_{y_1^1=0}^{1} p(y_1^1, y_1). \tag{14}$$

By Assumption 1 the observed second period transition probabilities impose the restrictions

$$\begin{aligned} &\Pr(Y_2 = y_2|D_1 = 1, Y_1 = 0) \qquad (15) \\ &= \frac{\sum_{y_2^{01}=0}^{1}\sum_{y_2^{00}=0}^{1} p(y_2, y_2^{01}, y_2^{00}|0,0)p(0,0) + \sum_{y_2^{01}=0}^{1}\sum_{y_2^{00}=0}^{1} p(y_2, y_2^{01}, y_2^{00}|0,1)p(0,1)}{\sum_{y_1^0=0}^{1} p(0, y_1^0)}, \end{aligned}$$

and

$$\begin{aligned} &\Pr(Y_2 = y_2|D_1 = 0, D_2 = 0, Y_1 = 0) \qquad (16) \\ &= \frac{\sum_{y_2^{11}=0}^{1}\sum_{y_2^{01}=0}^{1} p(y_2^{11}, y_2^{01}, y_2|0,0)p(0,0) + \sum_{y_2^{11}=0}^{1}\sum_{y_2^{01}=0}^{1} p(y_2^{11}, y_2^{01}, y_2|1,0)p(1,0)}{\sum_{y_1^1=0}^{1} p(y_1^1, 0)}, \end{aligned}$$

and

$$\Pr(Y_2 = y_2|D_1 = 0, D_2 = 1, Y_1 = 0) \tag{17}$$
$$= \frac{\sum_{y_2^{11}=0}^{1} \sum_{y_2^{00}=0}^{1} p(y_2^{11}, y_2, y_2^{00}|0,0)p(0,0) + \sum_{y_2^{11}=0}^{1} \sum_{y_2^{00}=0}^{1} p(y_2^{11}, y_2, y_2^{00}|1,0)p(1,0)}{\sum_{y_1^1=0}^{1} p(y_1^1, 0)}.$$

Theorem 1 provides closed form expressions for the sharp bounds on $\text{ATETS}_2^{11,00}$. These bounds require no assumptions beyond sequential random assignment among survivors. They allow for arbitrary heterogeneity in treatment response. We explicitly show that the bounds are sharp. The bounds exist if $\Pr(Y_1 = 0|D_1 = 1) > 0$, because if this probability is 0 the subpopulation for which $\text{ATETS}_2^{11,00}$ is defined has no members.

Theorem 1 (Bounds on ATETS) *Suppose that Assumption 1 holds. If* $\Pr(Y_1 = 0|D_1 = 1) = 0$*, then* $\text{ATETS}_2^{11,00}$ *is not defined. If* $\Pr(Y_1 = 0|D_1 = 1) > 0$*, then we have the following sharp bounds*

$$\begin{aligned} &\Pr(Y_2 = 1|D_1 = 1, Y_1 = 0) \\ &- \min\left\{1, \frac{1 - (1 - \Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0)) \Pr(Y_1 = 0|D_1 = 0)}{\Pr(Y_1 = 0|D_1 = 1)}\right\} \\ &\qquad \le \text{ATETS}_2^{11,00} \le \\ &\qquad\qquad \Pr(Y_2 = 1|D_1 = 1, Y_1 = 0) \\ &\qquad\qquad - \max\left\{0, \frac{\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0) \Pr(Y_1 = 0|D_1 = 0) - 1}{\Pr(Y_1 = 0|D_1 = 1)} + 1\right\}. \end{aligned} \tag{18}$$

Proof See Appendix.

As a special case

Corollary 1 *If* $\Pr(Y_1 = 0|D_1 = 1) > 0$ *and* $\Pr(Y_1 = 0|D_1 = 1) + \Pr(Y_1 = 0|D_1 = 0) - 1 \le 0$*, then*

$$\Pr(Y_2 = 1|D_1 = 1, Y_1 = 0) - 1 \le \text{ATETS}_2^{11,00} \le \Pr(Y_2 = 1|D_1 = 1, Y_1 = 0). \tag{19}$$

Inspection of the bounds and the proof in the appendix shows that $\mathbb{E}(Y_2^{11}|Y_1^1 = 0)$ is point identified. It also shows that the upper and lower bound on $\mathbb{E}(Y_2^{00}|Y_1^1 = 0)$ are equal if all treated and all controls survive the first period. Also, corollary 2 shows that if there is no dynamic selection, i.e. if $\Pr(Y_1 = 0|D_1 = 0) = 1$ and $\Pr(Y_1 = 0|D_1 = 1) = 1$, the dynamic treatment effect $\text{ATETS}_2^{11,00}$ is point identified. If everyone survives the first period we have under random treatment assignment two directly comparable groups even in the second period.

Corollary 2 (Point identification) $\text{ATETS}_2^{11,00}$ *is point identified if and only if both* $\Pr(Y_1 = 0|D_1 = 0) = 1$ *and* $\Pr(Y_1 = 0|D_1 = 1) = 1$.

The information in the bound depends on its width. The best case is that none of the 0 or 1 restrictions is binding and in that case the width of the bounds is

$$\frac{2 - \Pr(Y_1 = 0|D_1 = 1) - \Pr(Y_1 = 0|D_1 = 0)}{\Pr(Y_1 = 0|D_1 = 1)}.$$

3.2.2 Arbitrary number of periods

In the case of an arbitrary number of periods we only need to consider the effect in period t of a treatment that starts in period 1 relative to a treatment that starts in a later period before period t or after period t. We only consider the latter case here, but the bounds for the case that treatment starts between periods 1 and t can be derived in the same way. The relevant Average Treatment Effect on Survivors is $\text{ATETS}_t^{1,0}$ where 1 and 0 stand for t vectors of 1 and 0, i.e. treatment in all periods and control in all periods, and is defined by

$$\text{ATETS}_t^{1,0} = \mathbb{E}\left[Y_t^1 \,\middle|\, \overline{Y}_{t-1}^1 = 0\right] - \mathbb{E}\left[Y_t^0 \,\middle|\, \overline{Y}_{t-1}^1 = 0\right],$$

with the notation $\overline{Y}_t = (Y_t, \ldots, Y_1)'$ that also applies to other variables, the event of survival treatment up to and including t. Again, note that the superscripts 1 and 0 stand for vectors of 1 and 0 of appropriate length. The bounds are given in the next theorem.

Theorem 2 (Bounds on ATETS) *Suppose that Assumption 1 holds. If* $\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1\right) = 0$ *then* $\text{ATETS}_t^{1,0}$ *is not defined. If* $\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1\right) > 0$ *then*

$$\begin{aligned} &\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 1) \\ &- \min\left\{1, \frac{1 - [1 - \Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 0)]\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right)}{\Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1)}\right\} \\ &\quad \leq \text{ATETS}_t^{1,0} \leq \\ &\qquad \Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 1) \\ &\qquad - \max\left\{0, \frac{\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 0)\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right) - 1}{\Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1)} + 1\right\}. \end{aligned} \tag{20}$$

Proof See Appendix.

As a special case

Corollary 3 *If* $\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1\right) + \Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right) - 1 \leq 0,$ *then*

$$\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 1) - 1 \leq \text{ATETS}_2^{11,00} \leq \Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 1). \tag{21}$$

Note the similar structure of the bounds in the two period case and the arbitrary t case.

3.2.3 Average treatment effect on survivors

We now present bounds on the average effect on survivors. As in the previous subsection we only consider the effect in period t of a treatment that started in period 1 relative to a treatment that starts after period t (if ever). Therefore the relevant Average Treatment Effect on Survivors $\text{ATES}_t^{1,0}$ is

$$\text{ATES}_t^{1,0} = \mathbb{E}\left[Y_t^1 \middle| S_{t-1}\right] - \mathbb{E}\left[Y_t^0 \middle| S_{t-1}\right],$$

with $S_t = \{Y_t^1 = 0, \ldots, Y_1^1 = 0, Y_t^0 = 0, \ldots, Y_1^0 = 0\}$, the event of survival up to and including t. The bounds are given in Theorem 3.

Theorem 3 (Bounds on ATES) *Suppose that Assumption 1 holds. If* $\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=1)+\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=0)-1\leq 0$*, then* $\text{ATES}_t^{1,0}$ *is not defined. If* $\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=1)+\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=0)-1>0$*, then we have the following sharp bounds*

$$
\begin{aligned}
&\max\left\{0,\frac{(\Pr(Y_t=1|\overline{Y}_{t-1}=0,\overline{D}_t=1)-1)\Pr\left(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=1\right)}{\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=1)+\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=0)-1}+1\right\}\\
&-\min\left\{1,\frac{\Pr(Y_t=1|\overline{Y}_{t-1}=0,\overline{D}_t=0)\Pr\left(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=0\right)}{\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=1)+\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=0)-1}\right\}\\
&\leq \text{ATES}_t^{1,0}\leq\\
&\min\left\{1,\frac{\Pr(Y_t=1|\overline{Y}_{t-1}=0,\overline{D}_t=1)\Pr\left(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=1\right)}{\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=1)+\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=0)-1}\right\}\\
&-\max\left\{0,\frac{(\Pr(Y_t=1|\overline{Y}_{t-1}=0,\overline{D}_t=0)-1)\Pr\left(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=0\right)}{\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=1)+\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=0)-1}+1\right\}.
\end{aligned}
\tag{22}
$$

Proof see Appendix.

One difference compared to $\text{ATETS}_t^{1,0}$ is that, in general, neither the outcome under treatment ($\mathbb{E}\left[Y_t^1|S_{t-1}\right]$) nor the outcome under no-treatment ($\mathbb{E}\left[Y_t^0|S_{t-1}\right]$) is point identified. One similarity is that both $\text{ATES}_t^{1,0}$ and $\text{ATETS}_t^{1,0}$are point identified if $\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=1)=1$ and $\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=0)=1$.

4 Bounds on treatment effects on transitions under additional assumptions

4.1 Monotone Treatment Response, Common Shocks, and Positively Correlated Outcomes

The sharp bounds in the previous section did not impose any assumptions beyond random assignment. In this section, we explore the identifying power of additional assumptions. The assumptions that we make are implicit in parametric models as the MPH model, and also in the discrete duration models and structural models presented in this section. A general discrete duration model for the control and treated outcomes is

$$
\begin{aligned}
Y_{it}^0 &= I(\alpha_t+V_i-\varepsilon_{it}\geq 0),\\
Y_{it}^1 &= I(\alpha_t+\gamma_{it}+V_i-\varepsilon_{it}\geq 0).
\end{aligned}
\tag{23}
$$

This discrete duration model has a composite error that is the sum of unobserved heterogeneity V_i and a random shock ε_{it}. The model restricts the joint distribution of the potential outcomes. A less restrictive model has different random shocks $\varepsilon_{0it},\varepsilon_{1it}$ that are independent, but even in this case the potential outcomes are positively correlated through their dependence on V_i. In the sequel we consider assumptions on the joint distribution of potential outcomes in different treatment arms, that are in line with the assumptions implicit in this model, but do not assume that the potential outcomes are exactly as in this model. These assumptions will be used in combination with a weaker version of the constant treatment effect assumption. In the above model the treatment has a positive effect on the survival time if $\gamma_{it} \leq 0$ for all i, t. This is essentially the Monotone Treatment Response (MTR) assumption introduced by Manski (1997) and Manski and Pepper (2000). Since the assumptions introduced in this section do not rely on a particular discrete duration model they are consistent with nonproportional structural hazard models suggested by Van den Berg (2001).

The Monotone Treatment Response (MTR) is a weaker assumption than homogeneous treatment effect. As before we denote the event of survival under both $\overline{d}_0(t)$ and $\overline{d}_1(t)$ by S_t.

Assumption 2 (Monotone Treatment Response (MTR)) *For treatment paths $\overline{d}_{0t}, \overline{d}_{1t}$ we have that for all i either*

$$\Pr\left(Y_{it}^{\overline{d}_{1t}} = 1 \middle| S_{i,t-1}\right) \geq \Pr\left(Y_{it}^{\overline{d}_{0t}} = 1 \middle| S_{i,t-1}\right),$$

for all t, or

$$\Pr\left(Y_{it}^{\overline{d}_{1t}} = 1 \middle| S_{i,t-1}\right) \leq \Pr\left(Y_{it}^{\overline{d}_{0t}} = 1 \middle| S_{i,t-1}\right),$$

for all t.

For $t = 1$ Assumption 2 implies that for all i

$$\Pr(Y_{i1}^1 = 1) \geq \Pr(Y_{i1}^0 = 1),$$

or

$$\Pr(Y_{i1}^1 = 1) \leq \Pr(Y_{i1}^0 = 1).$$

Note that it is assumed that the effect is either positive or negative for all t. This assumption can be relaxed at the expense of more complicated bounds.

Assumption 2 refers to the individual transition probability and not to the transition indicators. These individual transition probabilities are defined with respect to the distribution of individual idiosyncratic shocks, e.g. ε in the MPH model. The population transition probabilities that appear in the definition of the ATETS and in Theorem 1 are individual transition probabilities averaged over the distribution of the individual heterogeneity among the survivors in both treatment arms.

The next assumption restricts the joint distribution of potential outcomes in the treatment arms. The assumption essentially imposes that the outcomes in all treatment arms involve the same random shocks. Consider the discrete duration model in (23). If $\gamma_{it} \leq 0$ then the treated have a larger survival probability in t. Therefore the event that i survives in t if not treated, i.e. $Y_{it}^0 = 0$, is equivalent to $\varepsilon_{it} \geq \alpha_t + V_i$, so that this event implies that $\varepsilon_{it} \geq \alpha_t + \gamma_{it} + V_i \geq 0$, i.e. $Y_{it}^1 = 0$. Note that we assume that the random shock ε_{it} is invariant under a change in treatment status. This is stronger than the assumption that the *distribution* of the random shocks is the same whether i is treated or not. The latter assumption can have random shocks $\varepsilon_{it}, \tilde{\varepsilon}_{it}$ in the model above, if we assume that they have the same distribution. In a structural model the random shocks are often invariant, as is illustrated in a simple job search model below.

Assumption 3 (Common Shocks (CS)) *For all i, t and treatment paths $\overline{d}_0(t)$ and $\overline{d}_1(t)$*

$$\Pr(Y_{it}^{\overline{d}_{1t}} = 0|S_{i,t-1}) \geq \Pr(Y_{it}^{\overline{d}_{0t}} = 0|S_{i,t-1}) \quad \Rightarrow \quad \Pr(Y_{it}^{\overline{d}_{1t}} = 0|S_{i,t-1}, Y_{it}^{\overline{d}_{0t}} = 0) = 1, \tag{24}$$

and

$$\Pr(Y_{it}^{\overline{d}_{1t}} = 0|S_{i,t-1}) \leq \Pr(Y_{it}^{\overline{d}_{0t}} = 0|S_{i,t-1}) \quad \Rightarrow \quad \Pr(Y_{it}^{\overline{d}_{0t}} = 0|S_{i,t-1}, Y_{it}^{\overline{d}_{1t}} = 0) = 1. \tag{25}$$

Because the right-hand side of (24) is equivalent to $\Pr(Y_{it}^{\overline{d}_{1t}} = 1|S_{i,t-1}, Y_{it}^{\overline{d}_{0t}} = 0) = 0$, it is also equivalent to $\Pr(Y_{it}^{\overline{d}_{1t}} = 1, Y_{it}^{\overline{d}_{0t}} = 0|S_{i,t-1}) = 0$, which in turn is equivalent to $\Pr(Y_{it}^{\overline{d}_{1t}} \geq Y_{it}^{\overline{d}_{0t}}|S_{i,t-1}) = 0$.

The assumption is satisfied in structural models. Consider for instance a non-stationary job search model for an unemployed individual as in Van den Berg (1990) or Meyer (1996). The treatment is a re-employment bonus as discussed in Section 5 below. In each period a job offer is obtained with probability $p(t, V_i)$. Let $Y_{of,it}$ be the indicator of an offer in period t and $Y_{of,it} = I(\varepsilon_{of,it} \in A(t, V_i))$ with $A(t, V_i)$ a set. If the job offer is not under control of i, the arrival process is the same under treatment and control. The reservation wage is denoted by ξ_{it}^1 for the treated and ξ_{it}^0 for the controls. In general (see Meyer (1996)) $\xi^1(t, V_i) \leq \xi^0(t, V_i)$, so that if H is the wage offer c.d.f. we have the acceptance probabilities $1 - H(\xi^1(t, V_i)) \geq 1 - H(\xi^0(t, V_i))$. The acceptance indicators are $Y_{ac,it}^0 = I(\varepsilon_{w,it} \geq \xi^0(t, V_i))$ and $Y_{ac,it}^1 = I(\varepsilon_{w,it} \geq \xi^1(t, V_i))$ with $\varepsilon_{w,it}$ the wage offer. Because $Y_{it}^0 = Y_{of,it} Y_{ac,it}^0$ and $Y_{it}^1 = Y_{of,it} Y_{ac,it}^1$, we see that

$$Y_{it}^1 = 0 \Rightarrow Y_{it}^0 = 0.$$

Note that the dimension of V_i is arbitrary and that we have two random shocks that have a structural interpretation and are invariant under a change in treatment status.

If we compare the transition probability $\Pr(Y_2^{00} = 1|Y_1^1 = 0, Y_1^0 = 0)$ to $\Pr(Y_2^{00} = 1|Y_1^1 = 1, Y_1^0 = 0)$, i.e. the probability of a transition in period 2 if no treatment was received in periods 1 and 2 given survival with or without treatment in period 1 to the same probability given survival without but not with treatment in period 1, then it is reasonable to assume that the former probability is not larger than the latter. Individuals with $Y_1^1 = 0, Y_1^0 = 0$ have characteristics that make them not leave the initial state as opposed to individuals with $Y_1^1 = 1, Y_1^0 = 0$ that have characteristics that make them leave the initial state if treated in period 1. If the variables that affect the transition out of the initial state are positively correlated between periods, then

$$\Pr(Y_2^{00} = 1|Y_1^1 = 0, Y_1^0 = 0) \leq \Pr(Y_2^{00} = 1|Y_1^1 = 1, Y_1^0 = 0). \tag{26}$$

To motivate this consider the discrete duration model for those not treated in periods $1, \ldots, t$

$$Y_{it}^0 = I(\alpha_t + V_i - \varepsilon_{it} \geq 0),$$

and for those who are treated in these periods

$$Y_{it}^1 = I(\alpha_t + \gamma_{it} + V_i - \tilde{\varepsilon}_{it} \geq 0).$$

Note that the Common Shocks assumption is not made. Now $Y_{it}^0 = 1$ if and only if

$$V_i - \varepsilon_{it} \geq -\alpha_t.$$

Let $k = 1, \ldots, t-1$. The conditioning events are $Y_{is}^0 = 0, s = 1, \ldots, t-1$ and $Y_{is}^1 = 0, s = 1, \ldots, t-1$, thus

$$\begin{aligned} &V_i - \varepsilon_{is} < -\alpha_s, && s = 1, \ldots, t-1, \\ &V_i - \tilde{\varepsilon}_{is} < -\alpha_s - \gamma_{is}, && s = 1, \ldots, t-1, \end{aligned}$$

and $Y_{is}^0 = 0$, for $s = 1, \ldots, t-1$, $Y_{is}^1 = 0$, for $s = 1, \ldots, k-1$, and $Y_{ik}^1 = 1$, thus

$$\begin{aligned} &V_i - \varepsilon_{is} < -\alpha_s, && s = 1, \ldots, t-1, \\ &V_i - \tilde{\varepsilon}_{is} < -\alpha_s - \gamma_{is}, && s = 1, \ldots, k-1, \\ &V_i - \tilde{\varepsilon}_{ik} \geq -\alpha_k - \gamma_{ik}. \end{aligned}$$

For example, for $t = 2$ and $k = 1$ the conditioning events are

$$V_i - \varepsilon_{i1} < -\alpha_1, \qquad V_i - \tilde{\varepsilon}_{i1} < -\alpha_1 - \gamma_{i1},$$

and

$$V_i - \varepsilon_{i1} < -\alpha_1, \qquad V_i - \tilde{\varepsilon}_{i1} \geq -\alpha_1 - \gamma_{i1}.$$

Hence if $V_i - \varepsilon_{i1}$ and $V_i - \tilde{\varepsilon}_{i1}$ are positively related with $V_i - \varepsilon_{i2}$ then (26) will in general hold. Individuals with $Y_1^1 = 1, Y_1^0 = 0$ are assumed to be more susceptible to a transition in period 2 than individuals with $Y_1^1 = 0, Y_1^0 = 0$.

In the general case we have by the same reasoning

$$\begin{aligned} &\Pr(Y_t^0 = 1 | Y_k^1 = 1, Y_{k-1}^1 = 0, \ldots, Y_1^1 = 0, Y_{t-1}^0 = 0, \ldots, Y_1^0 = 0) \\ &\quad \geq \Pr(Y_t^0 = 1 | Y_k^1 = 0, Y_{k-1}^1 = 0, \ldots, Y_1^1 = 0, Y_{t-1}^0 = 0, \ldots, Y_1^0 = 0) \\ &\quad\quad \geq \Pr(Y_t^0 = 1 | Y_{t-1}^1 = 0, \ldots, Y_1^1 = 0, Y_{t-1}^0 = 0, \ldots, Y_1^0 = 0). \end{aligned}$$

An analogous argument can be made for $\Pr(Y_t^1 = 1 | Y_{t-1}^1 = 0, \ldots, Y_1^1 = 0, Y_k^0 = 1, Y_{k-1}^0 = 0, \ldots, Y_0^1 = 0)$. These arguments lead to the following assumption

Assumption 4 (Positively Correlated Outcomes (PCO)) *For all $k = 1, \ldots, t-1$ we have*

$$\begin{aligned} &\Pr(Y_t^0 = 1 | Y_k^1 = 1, Y_{k-1}^1 = 0, \ldots, Y_1^1 = 0, Y_{t-1}^0 = 0, \ldots, Y_1^0 = 0) \\ &\quad \geq \Pr(Y_t^0 = 1 | Y_{t-1}^1 = 0, \ldots, Y_1^1 = 0, Y_{t-1}^0 = 0, \ldots, Y_1^0 = 0), \end{aligned}$$

and

$$\begin{aligned} &\Pr(Y_t^1 = 1 | Y_{t-1}^1 = 0, \ldots, Y_1^1 = 0, Y_k^0 = 1, Y_{k-1}^0 = 0, \ldots, Y_0^1 = 0) \\ &\quad \geq \Pr(Y_t^1 = 1 | Y_{t-1}^1 = 0, \ldots, Y_1^1 = 0, Y_{t-1}^0 = 0, \ldots, Y_1^0 = 0), \end{aligned}$$

and

$$\begin{aligned} &\Pr(Y_t^0 = 1 | Y_{t-1}^1 = 0, \ldots, Y_1^1 = 0, Y_k^0 = 1, Y_{k-1}^0 = 0, \ldots, Y_0^1 = 0) \\ &\quad \geq \Pr(Y_t^0 = 1 | Y_{t-1}^1 = 0, \ldots, Y_1^1 = 0, Y_{t-1}^0 = 0, \ldots, Y_1^0 = 0), \end{aligned}$$

and

$$\begin{aligned} &\Pr(Y_t^1 = 1 | Y_k^1 = 1, Y_{k-1}^1 = 0, \ldots, Y_1^1 = 0, Y_{t-1}^0 = 0, \ldots, Y_1^0 = 0) \\ &\quad \geq \Pr(Y_t^0 = 1 | Y_{t-1}^1 = 0, \ldots, Y_1^1 = 0, Y_{t-1}^0 = 0, \ldots, Y_1^0 = 0). \end{aligned}$$

The motivating example shows that PCO does not imply nor is implied by MTR or CS. The CS assumption is on the contemporaneous correlation of random shocks while PCO relates to a (positive) relation of the combined random error over time. Since the latter in general contains an important individual effect, positive correlation is not a strong assumption.

4.2 Bounds under the additional assumptions

We now obtain bounds on ATETS for arbitrary t when we compare a treatment started in period 1 to no treatment in all periods. Bounds under MTR and CS are given in Theorem 4 and Theorem 5 provides bounds under PCO. Bounds under all three additional assumptions are in Theorem 6.

Theorem 4 (Bounds on ATETS under MTR and CS for t periods) *Suppose Assumptions 1, 2, and 3 hold. If* $\Pr\left(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=1\right)=0$ *then* $\text{ATETS}_t^{1,0}$ *is not defined. If* $\Pr\left(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=1\right)>0$*, then*

$$\begin{aligned}
&\Pr(Y_t=1|\overline{Y}_{t-1}=0,\overline{D}_t=1)\\
&-\min\left\{1, 1+\frac{\Pr(Y_t=1|\overline{Y}_{t-1}=0,\overline{D}_t=0)\Pr\left(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=0\right)}{\Pr(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=1)}\right.\\
&\qquad\qquad\left.-\frac{\min\left\{\Pr(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=1),\Pr(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=0)\right\}}{\Pr(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=1)}\right\}\\
&\qquad\qquad\leq \text{ATETS}_t^{1,0}\leq\\
&\qquad\qquad\qquad\Pr(Y_t=1|\overline{Y}_{t-1}=0,\overline{D}_t=1)\\
&\qquad\qquad\qquad-\max\left\{0, \frac{[\Pr(Y_t=1|\overline{Y}_{t-1}=0,\overline{D}_t=0)-1]\Pr\left(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=0\right)}{\Pr(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=1)}\right.\\
&\qquad\qquad\qquad\qquad\left.+\frac{\min\left\{\Pr(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=1),\Pr(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=0)\right\}}{\Pr(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=1)}\right\}.
\end{aligned}$$

Proof See Appendix.

Theorem 5 (Bounds on ATETS under PCO for t periods) *Let Assumptions 1 and 4 hold. If* $\Pr\left(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=1\right)=0$ *then* $\text{ATETS}_t^{1,0}$ *is not defined. If* $\Pr\left(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=1\right)>0$ *and* $\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=1)+\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=0)-1>0$ *for all*

$s = 1, \ldots, t-1$, *then*

$$\Pr(Y_t = 1|\overline{D}_t = 1, \overline{Y}_{t-1} = 0) - \min\left\{1, 1 - \frac{1 - \Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 0)}{\Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1)} * \right.$$
$$\left. * \prod_{s=1}^{t-1} \max\left\{0, \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 1) + \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 0) - 1\right\}\right\}$$
$$\leq \text{ATETS}_t^{1,0} \leq$$
$$\Pr(Y_t = 1|\overline{D}_t = 1, \overline{Y}_{t-1} = 0)$$
$$- \max\left\{0, \frac{\left(\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 0) - 1\right) \Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right)}{\prod_{s=1}^{t-1} \max\{0, \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 1) + \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 0) - 1\}} + 1\right\}.$$

If $\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1\right) > 0$ *and* $\Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 1) + \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 0) - 1 \leq 0$ *for some* $s \leq t$, *then*

$$\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 1) - 1 \leq \text{ATETS}_t^{1,0} \leq \Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 1).$$

Proof See Appendix.

Theorem 6 (Bounds on ATETS under MTR, CS and PCO for t periods) *Let the Assumptions 1-4 hold. If* $\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1\right) = 0$ *then* $\text{ATETS}_t^{1,0}$ *is not defined. If* $\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1\right) > 0$, *then*

$$\Pr(Y_t = 1|\overline{D}_t = 1, \overline{Y}_{t-1} = 0) - \min\left\{1, 1 - \frac{1 - \Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 0)}{\Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1)} * \right.$$
$$\left. * \min\left\{\prod_{s=1}^{t-1} \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 1), \prod_{s=1}^{t-1} \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 0)\right\}\right\}$$
$$\leq \text{ATETS}_t^{1,0} \leq$$
$$\Pr(Y_t = 1|\overline{D}_t = 1, \overline{Y}_{t-1} = 0)$$
$$- \max\left\{0, \frac{\left(\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 0) - 1\right) \Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right)}{\min\left\{\prod_{s=1}^{t-1} \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 1), \prod_{s=1}^{t-1} \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 0)\right\}} + 1\right\}.$$

Proof See Appendix.

5 Inference

Initially, consider inference on $\theta_0 = \text{ATETS}_2^{11,00}$. We assume that $\Pr(Y_1 = 0|D_1 = 1) > 0$. From Theorem 1 we then find that the bounds on θ_0 can be expressed as

$$\max(a_1, a_2) =: \ell \leq \theta_0 \leq u := \min(a_3, a_4), \tag{27}$$

with

$$
\begin{aligned}
a_1 &= a_3 - 1, \\
a_2 &= a_3 - \frac{1 - [1 - \Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0)]\Pr(Y_1 = 0|D_1 = 0)}{\Pr(Y_1 = 0|D_1 = 1)}, \\
a_3 &= \Pr(Y_2 = 1|D_1 = 1, Y_1 = 0), \\
a_4 &= a_3 - 1 + \frac{1 - \Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0)\Pr(Y_1 = 0|D_1 = 0)}{\Pr(Y_1 = 0|D_1 = 1)}.
\end{aligned}
$$

If we observe an iid sample $\{(Y_{i1}, Y_{i2}, D_{i1}, D_{i2}), i \in 1, \ldots, n\}$, then the sample analog of $a = (a_1, a_2, a_3, a_4)'$ can easily be constructed, for example

$$
\widehat{a}_3 = \frac{\frac{1}{n}\sum_{i=1}^n \mathbb{1}(Y_{i2} = 1, D_{i1} = 1, Y_{i1} = 0)}{\frac{1}{n}\sum_{i=1}^n \mathbb{1}(D_{i1} = 1, Y_{i1} = 0)}, \qquad \widehat{a}_1 = \widehat{a}_3 - 1,
$$

and analogously for $\widehat{a}_2$ and $\widehat{a}_4$. It is easy to show that as the sample size n goes to infinity

$$
\sqrt{n}(\widehat{a} - a) \Rightarrow \mathcal{N}(0, \Sigma_a), \tag{28}
$$

and we can construct a consistent estimator $\widehat{\Sigma}_a$ of the 4×4 matrix Σ_a (for example, we use bootstrapping to calculate $\widehat{\Sigma}_a$ in our application in Section 6). In the following we assume that $\Sigma_{a,kk} > 0$ for all $k = 1, 2, 3, 4$.[3]

The identification results in Theorem 2 and 5 on $\theta_0 = \text{ATETS}_t^{1,0}$ can also be expressed in the form (27) for suitable $a = (a_1, a_2, a_3, a_4)'$ that can be estimated such that (28) holds asymptotically. The identified set for $\theta_0 = \text{ATES}_t^{1,0}$ in Theorem 3 can be expressed as $\max(a_1, a_2, a_3, a_4) \leq \theta_0 \leq \min(a_5, a_6, a_7, a_8)$, and the identified set for $\theta_0 = \text{ATETS}_t^{1,0}$ in Theorem 4 and 6 can be expressed as $\max(a_1, a_2, a_3) \leq \theta_0 \leq \min(a_4, a_5, a_6)$, but the inference problem is otherwise analogous, and it is straightforward to generalize the discussion below to these cases.

5.1 Connection to the Moment Inequality Literature

The inference problem for θ_0 that is summarized by (27) and (28) is asymptotically equivalent to an inference problem on a finite number of moment inequalities that is well-studied in the literature, for example in Chernozhukov, Hong, and Tamer (2007), Romano and Shaikh (2008), Rosen (2008), Andrews and Guggenberger (2009), Andrews and Soares (2010), and Andrews and Barwick (2012). To make this connection explicit we define

$$
m(\theta) := \begin{pmatrix} \Sigma_{a,11}^{-1/2}(a_1 - \theta) \\ \Sigma_{a,22}^{-1/2}(a_2 - \theta) \\ \Sigma_{a,33}^{-1/2}(\theta - a_3) \\ \Sigma_{a,44}^{-1/2}(\theta - a_4) \end{pmatrix}, \qquad \widehat{m}(\theta) := \begin{pmatrix} \widehat{\Sigma}_{a,11}^{-1/2}(\widehat{a}_1 - \theta) \\ \widehat{\Sigma}_{a,22}^{-1/2}(\widehat{a}_2 - \theta) \\ \widehat{\Sigma}_{a,33}^{-1/2}(\theta - \widehat{a}_3) \\ \widehat{\Sigma}_{a,44}^{-1/2}(\theta - \widehat{a}_4) \end{pmatrix}.
$$

[3]Since $\widehat{a}_1$ and $\widehat{a}_3$ are perfectly correlated we have $\Sigma_a v = 0$ for the vector $v = (1, -1, 0, 0)'$, implying that $\text{rank}(\Sigma_a) \leq 3$, but this rank deficiency turns out not to be important for our purposes.

The bounds (27) can then equivalently be expressed as $m(\theta_0) \leq 0$, which is analogous to imposing four moment inequalities.[4] For convenience we have normalized $m(\theta)$ such that each component of $\sqrt{n}\,[\widehat{m}(\theta) - m(\theta)]$ has asymptotic variance equal to one. Using (28) we obtain $\sqrt{n}\,[\widehat{m}(\theta) - m(\theta)] \Rightarrow \mathcal{N}(0, \Sigma_m)$, where $\Sigma_m = A\Sigma_a A$, with $A = \mathrm{diag}(\Sigma_{a,11}^{-1/2}, \Sigma_{a,22}^{-1/2}, -\Sigma_{a,33}^{-1/2}, -\Sigma_{a,44}^{-1/2})$. An estimator $\widehat{\Sigma}_m$ can be constructed analogously.

All the papers on moment inequalities cited above start from choosing an objective function (or test statistics), whose sample version we denote by $\widehat{Q}(\theta)$, and then construct a confidence set for θ_0 as

$$\widehat{\Theta}(C_{1-\alpha}) = \{\theta \in \mathbb{R} : n\widehat{Q}(\theta) \leq C_{1-\alpha}\}, \tag{29}$$

where $C_{1-\alpha} \geq 0$ is a critical value that is chosen such that confidence $1 - \alpha$ is achieved asymptotically, i.e. $\lim_{n\to\infty} \Pr(\theta_0 \in \widehat{\Theta}(C_{1-\alpha})) \geq 1 - \alpha$.[5] Various objective functions have been considered in the literature. For example, the objective function considered in Chernozhukov, Hong, and Tamer (2007) reads in our notation $\widehat{Q}(\theta) = \|[\widehat{m}(\theta)]_+\|^2$, where $\|.\|$ refers to the Euclidian norm, and $[\widehat{m}(\theta)]_+ := \max(0, \widehat{m}(\theta))$, applied componentwise to the vector $\widehat{m}(\theta)$.

5.2 Construction of Confidence Intervals

Our specific inference problem is easier than the general inference problem for moment inequalities, because in our case the parameter θ_0 is just a scalar, and the total number of inequalities is relatively small. Our goal in the following is therefore to outline a concrete method of how to construct a confidence interval in that special case.

We choose the objective function $\widehat{Q}(\theta) = \|[\widehat{m}(\theta)]_+\|_\infty^2$, where $\|.\|_\infty$ is the infinity norm, i.e. we have $\widehat{Q}(\theta) = \max\{0, \widehat{m}_1(\theta), \widehat{m}_2(\theta), \widehat{m}_3(\theta), \widehat{m}_4(\theta)\}^2$. This objective function is convenient for our purposes, because the confidence set defined above then takes the intuitive form

$$\begin{aligned}&\widehat{\Theta}(C_{1-\alpha}) \\ &= \left[\max\left(\widehat{a}_1 - \frac{c_{1-\alpha}\widehat{\Sigma}_{a,11}^{1/2}}{\sqrt{n}}, \widehat{a}_2 - \frac{c_{1-\alpha}\widehat{\Sigma}_{a,22}^{1/2}}{\sqrt{n}}\right), \min\left(\widehat{a}_3 + \frac{c_{1-\alpha}\widehat{\Sigma}_{a,33}^{1/2}}{\sqrt{n}}, \widehat{a}_4 + \frac{c_{1-\alpha}\widehat{\Sigma}_{a,44}^{1/2}}{\sqrt{n}}\right)\right],\end{aligned} \tag{30}$$

where $c_{1-\alpha} := \sqrt{C_{1-\alpha}}$. This confidence interval can be constructed very easily.

Most Robust Critical Value

The critical value $c_{1-\alpha}$ still needs to be chosen. The problem with choosing the critical value in moment inequality problems is that this choice depends on the unknown slackness vector

[4] $m(\theta)$ is not actually a moment function, but has a slightly more complicated structure (e.g. a_3 is a conditional probability that can be expressed as the ratio between two moments). This, however, does not matter for the asymptotic analysis since the estimator $\widehat{m}(\theta)$ has the same first order asymptotic properties as it would have in the moment inequality case. We can therefore fully draw on the insights of the existing literature.

[5] As discussed in e.g. Andrews and Soares (2010), it is important that the coverage probability is asymptotically bounded by $1 - \alpha$ uniformly over θ_0 and over the distribution of the observables. We have only formulated the pointwise condition here to keep the presentation simple.

$m(\theta_0)$, which indicates whether each inequality $m_k(\theta_0) \leq 0$ is binding, close to binding, or far from binding. It is known, however, that the largest ("worst case") critical value needs to be chosen if $m(\theta_0) = 0$, i.e. if all moment inequalities are binding at the true parameter. To find this critical value one can use the fact that in this worst case $n\widehat{Q}(\theta)$ is asymptotically distributed as $\|[Z]_+\|_\infty^2$, where $Z \sim \mathcal{N}(0, \Sigma_m)$ is a random four vector. Using the estimator $\widehat{\Sigma}_m$ one can simulate this distribution. However, it can easily be shown that the $1-\alpha$ quantile of $\|[Z]_+\|_\infty$ is always smaller or equal to the following conservative critical value

$$c_{1-\alpha} = \Phi^{-1}\left(1 - \frac{\alpha}{4}\right), \tag{31}$$

where Φ^{-1} is the quantile function (the inverse cdf) of the standard normal distribution. The factor 1/4 that appears here reflects the fact that we have four moment inequalities. Combining equations (30) and (31) provides a confidence interval that is uniformly valid, i.e. whose asymptotic size is bounded by α, independent of what the true values of a_1, a_2, a_3 and a_4 are.

Critical Value for the Case $\ell \ll u$

The critical values based on the "worst case" where all inequalities are binding ($m(\theta_0) = 0$) can be very conservative if one or multiple inequalities are far from binding ($m_k(\theta_0) \ll 0$).[6] Furthermore, for the inference on $\theta_0 = \text{ATETS}_2^{11,00}$ based on Theorem 1, with a's as given above, it can easily be shown that if $\Pr(Y_1 = 0|D_1 = 1) > 0$ and $\Pr(Y_1 = 0|D_1 = 0) < 1$, then we have $\max(a_1, a_2) =: \ell < u := \min(a_3, a_4)$, implying that $m(\theta_0) = 0$ is impossible. However, what matters for the coverage rate of the confidence interval at finite sample is not whether $\ell < u$, but whether the difference $u - \ell$ is large relative to the standard deviations $\Sigma_{a,kk}^{1/2}$ of the $\widehat{a}_k$, $k = 1, 2, 3, 4$. This is what we mean by $\ell \ll u$ in the subsection title above.

To formalize this one can consider a pretest of the hypothesis $H_0 : \ell = u$, against the alternative $H_a : \ell < u$, with pretest size α_n^{pre} chosen to be very small, e.g. $\alpha_n^{\text{pre}} = 0.001 \ll \alpha$.[7] If the pretest is not rejected, then the critical value (31) should be chosen. If the pretest is rejected, then the two problems of choosing a suitable lower and upper bound for the confidence interval $\widehat{\Theta}$ completely decouple, because with high confidence we know that for any θ only one of those bounds can be binding at the same time, implying that at most two of the moment inequalities $m(\theta_0) \leq 0$ can be binding. In this latter case we can therefore choose the less conservative critical value

$$c_{1-\alpha} = \Phi^{-1}\left(1 - \frac{\alpha}{2}\right), \tag{32}$$

when computing the confidence interval (30).

Critical Value for the Case $a_1 \ll a_2 \ll u$

Analogous to the discussion of (31), the critical value (32) is again potentially conservative because it is based on the case where two of the inequalities $m(\theta_0) \leq 0$ (for either the lower or

[6]In addition, the formula (31) only provides an upper bound for the optimal critical value at $m(\theta_0) = 0$, but this second issue is often not very severe. For example, for $\alpha = 0.05$ and $\Sigma_m = \mathbb{I}_4$ one finds by simulation that the 0.95 quantile of $\|[Z]_+\|_\infty$, with $Z \sim \mathcal{N}(0, \Sigma_m)$, is $c_{0.95} = 2.234$, while the much easier to computer conservative critical value in (31) is $\Phi^{-1}(0.9875) = 2.241$.

[7]Theoretically one can assume $\alpha_n^{\text{pre}} \to 0$ as $n \to \infty$ to avoid asymptotic size distortions due to the pretest.

the upper bound, respectively) are jointly binding.[8] For example, if we find that $a_1 \ll a_2 \ll u$ (by which we again mean that the null hypotheses $H_0 : a_1 = a_2$, vs. $H_a : a_1 < a_2$, and $H_0 : a_2 = u$, vs. $H_a : a_2 < u$, are rejected with very high confidence), then a natural confidence interval to report is

$$\widehat{\Theta} = \left[\widehat{a}_2 - \frac{\Phi^{-1}(1-\alpha)\,\widehat{\Sigma}_{a,22}^{1/2}}{\sqrt{n}}, \min\left(\widehat{a}_3 + \frac{\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\widehat{\Sigma}_{a,33}^{1/2}}{\sqrt{n}}, \widehat{a}_4 + \frac{\Phi^{-1}\left(1-\frac{\alpha}{2}\right)\widehat{\Sigma}_{a,44}^{1/2}}{\sqrt{n}}\right)\right].$$

Note that the lower bound of $\widehat{\Theta}$ now corresponds to inverting a standard one-sided t-test. Analogous confidence intervals can obviously be constructed in other cases, e.g. $\ell \ll a_3 \ll a_4$ or $a_2 \ll a_1 \ll a_4 \ll a_3$, etc.

The different critical values and corresponding confidence intervals discussed above correspond to cases where different subsets of the inequalities $m(\theta_0) \leq 0$ can be simultaneously binding, i.e. to a moment selection problem. A much more general discussion of moment selection is given e.g. in Andrews and Soares (2010). Different confidence intervals than those discussed here, e.g. based on different objective functions $\widehat{Q}(\theta)$, can of course also be considered.

It should be noted that pretesting is not required if we use the approach in Hahn and Ridder (2014) who obtain a confidence interval by inverting the Likelihood Ratio test for the composite null and composite alternative test. Their current results do not cover the case considered here and we did not attempt the non-trivial extension to the case considered here.

6 Application to the Illinois bonus experiment

6.1 The re-employment bonus experiment

Between mid-1984 and mid-1985, the Illinois Department of Employment Security conducted a randomized social experiment.[9] The goal of the experiment was to explore, whether re-employment bonuses paid to Unemployment Insurance (UI) beneficiaries (treatment 1) or their employers (treatment 2) reduced the length of unemployment spells.

Both treatments consisted of a \$ 500 re-employment bonus, which was about four times the average weekly unemployment insurance benefit. In the experiment, newly unemployed UI claimants were randomly divided into three groups:

1. The *Claimant Bonus Group*. The members of this group were instructed that they would qualify for a cash bonus of \$500 if they found a job (of at least 30 hours) within 11 weeks and, if they held that job for at least 4 months. A total of 4186 individuals were selected for this group, and 3527 (84%) agreed to participate.

2. The *Employer Bonus Group*. The members of this group were told that their next employer would qualify for a cash bonus of \$500 if they, the claimants, found a job (of at least 30 hours) within 11 weeks and, if they held that job for at least four months. A total of 3963 were selected for this group and 2586 (65%) agreed to participate.

[8] It is also conservative, because the information in the correlation matrix Σ_m is not used to construct (32). It corresponds to the the most extreme case where both lower bound estimators $\widehat{a}_1$ and $\widehat{a}_2$ (or both upper bound estimators $\widehat{a}_3$ and $\widehat{a}_4$) are perfectly negatively correlated.

[9] A complete description of the experiment and a summary of its results can be found in Woodbury and Spiegelman (1987).

3. The *Control Group*, i.e. all claimants not assigned to one of the treatment groups. This group consisted of 3952 individuals. The individuals assigned to the control group were excluded from participation in the experiment. In fact, they did not know that the experiment took place.

The descriptive statistics in Table 2 in Woodbury and Spiegelman (1987) confirm that the randomization resulted in three similar groups.

6.2 Results of previous studies

Woodbury and Spiegelman (1987) concluded from a direct comparison of the control group and the two treatment groups that the claimant bonus group had a significantly shorter average unemployment duration. The average unemployment duration was also shorter for the employer bonus group, but the difference was not significantly different from zero. In the USA UI benefits end after 26 weeks and since administrative data were used, all unemployment durations are censored at 26 weeks. Woodbury and Spiegelman ignore the censoring and take as outcome variable the number of weeks of insured unemployment.

Meyer (1996) analyzed the same data but focused on the treatment effects on conditional transition probabilities which allows him to properly account for censoring. Meyer focuses on the conditional transitions rates because both labor supply and search theory imply specific dynamic treatment effects. The bonus is only given to an unemployed individual if (s)he finds a job within 11 weeks and retains it for four months. The cash bonus is the same for all unemployed. Theory predicts that (i) the transition rate during the eligibility period (first 11 weeks) will be higher in the two treatment groups compared with the control group, and (ii) that the transition rate in the treatment groups will rise just before the end of the eligibility period, as the unemployed run out of time to collect the bonus.

To test these predictions, Meyer (1996) estimates a proportional hazard (PH) model with a flexible specification of the baseline hazard. He uses the treatment indicator as an explanatory variable. Since there was partial compliance with treatment his estimator can be interpreted as a intention to treat (ITT) estimator.[10] In his analysis Meyer controls for age, the logarithm of base period earnings, ethnicity , gender and the logarithm of the size of the UI benefits. He finds a significantly positive effect of the claimant bonus and a positive but insignificant effect of the employer bonus. A more detailed analysis of the effects for the claimant group reveals a positive effect on the transition rate during the first 11 weeks in unemployment, an increased effect during week 9 and 10, and no significant effect on the transition rate after week 11 as predicted by labor supply and search theory.

6.3 Estimates of bounds

In his study Meyer (1996) relies on the proportionality of the hazard rate to investigate his hypotheses. We now ask what can be said if the assumptions of the MPH model do not hold, that is what can be identified if we rely solely on random assignment and the

[10]The partial compliance is addressed in detail by Bijwaard and Ridder (2005). They introduce a new method to handle the selective compliance in the treatment group. If there is full compliance in the control group, their two-stage linear rank estimator is able to handle the selective compliance in the treatment group even for censored durations. In order to achieve this they assume a MPH structure for the transition rate. Their estimates indicate that the ITT estimates by Meyer (1996) underestimate the true treatment effect.

additional assumptions. As Meyer we consider the ITT effect, i.e. we do not correct for partial compliance. We divide the 24 month observation period into 12 subperiods: week 1-2, week 3-4, ... , week 23-24. The reason for this is that there is a pronounced even-odd week effect in the data, with higher transition rate during odd weeks. With these subperiods the predictions we wish to test are: (i) a positive treatment effect during periods 1-5, i.e.

$$\text{ATETS}_t^{1,0} > 0\ ,\ \ t = 1, \ldots, 5,$$

(ii) no effect after the bonus offer has expired in periods 6-12, i.e.

$$\text{ATETS}_t^{1,0} = 0\ ,\ \ t = 6, \ldots, 12,$$

and (iii) a larger effect of the bonus offer at the end of the eligibility period in period 5, i.e.

$$\text{ATETS}_5^{1,0} > \text{ATETS}_4^{1,0}.$$

Note that in this experiment the treatment assignment is in period 1, so that in $\text{ATETS}_t^{1,0}$ the superscripts 1 and 0 are t vectors with components equal to 1 and 0.

We report both the bounds that are obtained by simply replacing the population moments with their sample analogs, as well as the confidence intervals based on the approach described in section 5.[11] Table 1 presents the upper and the lower bound and the confidence interval on $\text{ATETS}_t^{1,0}$ for the claimant group assuming only random assignment. We find that the instantaneous treatment effect on the transition probability (week 1-2) is point identified and indicates a positive effect of the re-employment bonus. The transition probability is about 2 percentage points higher in the treatment group compared to the control group. This estimate is statistically significant. From week 3-4 and onwards the bounds are quite wide. In fact, without further assumptions we cannot rule out that the bonus actually has a negative impact on the conditional transition probability after week 3. However, the bounds are nevertheless informative on the average treatment effect in all time periods.

Table 1 also shows that the confidence intervals are marginally wider than the actual bounds. That is the uncertainty arising from the dynamic selection is far greater than the uncertainty due to sampling variation.

Next, Table 1 presents bounds under the additional assumptions in Section 4. As expected, if we impose additional assumptions the bounds are considerably narrower. Under MTR and CS we can rule out very large negative and very large positive dynamic treatment effects. Imposing MTR, CS as well as PCO further tightens the bounds. If these assumptions hold simultaneously we can, if we disregard sampling variation, rule out that the bonus offer has a negative effect on the transition rate out of unemployment up to week 20. This conclusions changes slightly when sampling variation is taken into account.

Let us return to the three hypotheses suggested by labor supply and search theory, and consider our most restrictive bounds under MTR, CS and PCO. We find that there is a positive effect of the bonus offer on the conditional transition rate up to week 11. This confirms the first hypothesis. The upper bound increases in time period 5 (weeks 9-10), but

[11] The covariance matrix Σ_a is estimated using bootstrap with 399 replications. Constructing confidence intervals furthermore requires moment selection, e.g. for the bounds under just random assignment we find that with very high confidence only one inequality is binding for the lower as well as the upper bound. Details are available from the authors upon request.

the lower bound does not increase enough, so that both an increase and no change (and even a small decrease) in the transition probability out of unemployment are consistent with the data. Now consider the third hypothesis that there is no effect on the transition rate after week 11. Again the bounds do not rule out that there is a positive effect on the conditional transition probability after week 11. These results illustrate that the evidence for the second and third hypotheses presented by a number of authors rely on the imposed structure, e.g. proportionality of the hazard or the restrictions implied by a particular discrete duration model.

7 Conclusions

In this paper, we have derived bounds on treatment effects on conditional transition probabilities under sequential randomization. The partial identification problem arises since random assignment only ensures comparability of the treatment and control groups at the time of randomization. In the literature this problem is often refereed to as the dynamic selection problem. For that reason only instantaneous or short-run effects are point identified, whereas dynamic or long-run effects in general are not point identified. Our weakest bounds impose no assumptions beyond sequential random assignment, so that they are not sensitive to arbitrary functional form assumptions, require no additional covariates and allow arbitrary heterogenous treatment effects as well as arbitrary unobserved heterogeneity. These non-parametric bounds offer an alternative to semi-parametric methods. They tend to be wide and therefore we have also derived more informative bounds under additional assumptions that often hold in semi-parametric reduced form and structural models.

An analysis of data from the Illinois re-employment bonus experiment shows that our bounds are informative about average treatment effects. It also demonstrates that previous results on the evolution of the average treatment effect require that assumptions as the proportionality of the hazard rate or those embodied in a particular (semi-)parametric discrete-time hazard model hold.

References

ABBRING, J. H., AND J. J. HECKMAN (2007): *Econometric evaluation of social programs, part III: Distributional treatment effects, dynamic treatment effects, dynamic discrete choice, and general equilibrium policy evaluation.*chap. Handbook of Econometrics, Volume 6. North Holland.

ABBRING, J. H., AND G. J. VAN DEN BERG (2003): "The non-parametric identification of treatment effects in duration models," *Econometrica*, 71, 1491–1517.

ANDREWS, D. W., AND P. J. BARWICK (2012): "Inference for parameters defined by moment inequalities: A recommended moment selection procedure," *Econometrica*, 80(6), 2805–2826.

ANDREWS, D. W., AND P. GUGGENBERGER (2009): "Validity of subsampling and plug-in asymptotic inference for parameters defined by moment inequalities," *Econometric Theory*, 25(03), 669–709.

ANDREWS, D. W., AND G. SOARES (2010): "Inference for parameters defined by moment inequalities using generalized moment selection," *Econometrica*, 78(1), 119–157.

BIJWAARD, G., AND G. RIDDER (2005): "Correcting for Selective Compliance in a Reemployment Bonus Experiment," *Journal of Econometrics*, 125, 77–111.

BONNAL, L., F. FOUGERE, AND A. SERANDON (1997): "Evaluating the impact of French employment policies on individual labour market histories," *Review of Economic Studies*, 64, 683–713.

CARD, D., AND D. SULLIVAN (1988): "Measuring the effect of subsidized training programs on movements in and out of unemployment," *Econometrica*, 56, 497–530.

CHERNOZHUKOV, V., H. HONG, AND E. TAMER (2007): "Estimation and confidence regions for parameter sets in econometric models1," *Econometrica*, 75(5), 1243–1284.

COX, D. R. (1972): "Regression models and life-tables (with discussion)," *Journal of the Royal Statistical Society*, 34, 187–220.

ELBERS, C., AND G. RIDDER (1982): "True and spurious duration dependence: The identifiability of the proportional hazards model," *Review of Economic Studies*, 49, 402–411.

GILL, R. D., AND J. M. ROBINS (2001): "Causal Inference for Complex Longitudinal Data: The Continuous Case," *Annals of Statistics*, 29, 1785–1811.

GRITZ, R. M. (1993): "The impact of training on the frequency and duration of employment," *Journal of Econometrics*, 57, 21–51.

HAHN, J., AND G. RIDDER (2014): "Non-standard tests through a composite null and alternative in point-identified parameters," *Journal of Econometric Methods*, 4, 1–28.

HAM, J. C., AND R. J. LALONDE (1996): "The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training," *Econometrica*, 64, 175–205.

Heckman, J., and S. Navarro (2007): "Dynamic Discrete Choice and Dynamic Treatment Effects," *Journal of Econometrics*, 136, 341–396.

Kluve, J., D. Card, M. Fertig, M. Gora, L. Jacobi, P. Jensen, R. Leetmaa, L. Nima, E. Patacchini, S. Schaffner, C. Schmidt, B. v. d. Klaauw, and A. Weber (2007): *Active Labor Market Policies in Europe: Performance and Perspectives*. Springer.

Manski, C. F. (1997): "Monotone treatment response," *Econometrica*, 65, 1311–1334.

Manski, C. F., and J. Pepper (2000): "Monotone Instrumental Variables: With an Application to the Returns to Schooling," *Econometrica*, 68, 115–136.

Meyer, B. D. (1996): "What Have We Learned from the Illinois Reemployment Bonus Experiment?," *Journal of Labor Economics*, 14, 26–51.

Ridder, G. (1986): "An event history approach to the evaluation of training, recruitment and employment programmes," *Journal of Applied Econometrics*, 1, 109126.

Romano, J. P., and A. M. Shaikh (2008): "Inference for identifiable parameters in partially identified econometric models," *Journal of Statistical Planning and Inference*, 138(9), 2786–2807.

Rosen, A. M. (2008): "Confidence sets for partially identified parameters that satisfy a finite number of moment inequalities," *Journal of Econometrics*, 146(1), 107–117.

Van den Berg, G. J. (1990): "Nonstationarity in Job Search Theory," *Review of Economic Studies*, 57, 255–277.

——— (2001): *Duration models: specification, identification and multiple durations*chap. Handbook of Econometrics, vol. 6. North-Holland.

Woodbury, S. A., and R. G. Spiegelman (1987): "Bonusses to workers and employers to reduce unemployment: randomized trials in Illinois," *American Economic Review*, 77(4), 513–530.

Tables

Table 1: Bounds on $ATETS^{1,0}$ for the Illinois job bonus experiment

	No assumption bounds [A]				MTR+CS [B]			
	Lower-CI	LB	UB	Upper-CI	Lower-CI	LB	UB	Upper-CI
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Week								
1-2	0.012	0.023	0.023	0.034	0.012	0.023	0.023	0.034
3-4	-0.145	-0.137	0.094	0.102	0.000	0.011	0.038	0.050
5-6	-0.259	-0.251	0.074	0.082	-0.007	0.004	0.046	0.056
7-8	-0.346	-0.339	0.078	0.086	0.004	0.013	0.063	0.073
9-10	-0.452	-0.444	0.069	0.077	0.000	0.008	0.069	0.079
11-12	-0.552	-0.544	0.062	0.070	0.000	0.008	0.062	0.072
13-14	-0.655	-0.648	0.056	0.064	-0.010	-0.002	0.056	0.064
15-16	-0.750	-0.743	0.051	0.058	-0.004	0.003	0.051	0.058
17-18	-0.844	-0.836	0.049	0.057	-0.007	0.000	0.049	0.057
19-20	-0.943	-0.936	0.049	0.057	-0.011	-0.004	0.049	0.056
21-22	-0.994	-0.953	0.047	0.056	-0.028	-0.021	0.047	0.055
23-24	-0.989	-0.944	0.056	0.064	-0.011	-0.002	0.056	0.064
	PCO [C]				MTR+CS+PCO [D]			
	Lower-CI	LB	UB	Upper-CI	Lower-CI	LB	UB	Upper-CI
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
Week								
1-2	0.012	0.023	0.023	0.034	0.012	0.023	0.023	0.034
3-4	-0.131	-0.123	0.094	0.102	0.002	0.014	0.038	0.049
5-6	-0.209	-0.202	0.074	0.082	-0.004	0.007	0.046	0.055
7-8	-0.256	-0.247	0.078	0.087	0.008	0.016	0.063	0.072
9-10	-0.306	-0.299	0.069	0.077	0.004	0.012	0.069	0.078
11-12	-0.348	-0.340	0.062	0.070	0.004	0.012	0.062	0.071
13-14	-0.388	-0.379	0.056	0.064	-0.004	0.003	0.056	0.064
15-16	-0.419	-0.411	0.051	0.058	0.000	0.007	0.051	0.059
17-18	-0.445	-0.438	0.049	0.057	-0.003	0.005	0.049	0.058
19-20	-0.472	-0.464	0.049	0.057	-0.006	0.001	0.049	0.057
21-22	-0.504	-0.496	0.047	0.063	-0.022	-0.014	0.047	0.055
23-24	-0.523	-0.513	0.056	0.073	-0.006	0.003	0.056	0.065

Notes: CI is 95% confidence intervals. Variances and covariances used to obtain the CI are estimated using bootstrap (399 replications).

Appendix A: Proofs

Proof of Lemma 1

Because Assumption 1 implies random assignment in period 1 we have

$$\mathbb{E}(Y_1|D_1=1)=\mathbb{E}(Y_1^1|D_1=1)=\mathbb{E}(Y_1^1),$$

and

$$\mathbb{E}(Y_1|D_1=0)=\mathbb{E}(Y_1^0|D_1=0)=\mathbb{E}(Y_1^0).$$

By the observation rule and by (6)

$$\mathbb{E}(Y_2|Y_1=0,D_1=1)=\mathbb{E}(Y_2^{11}|Y_1^1=0,D_1=1)=\mathbb{E}(Y_2^{11}|Y_1^1=0).$$

For (11)

$$\mathbb{E}(Y_2|Y_1=0,D_1=0,D_2=0)=\mathbb{E}(Y_2^{00}|Y_1^0=0,D_1=0,D_2=0)=$$
$$\mathbb{E}(Y_2^{00}|Y_1^0=0,D_1=0)=\mathbb{E}(Y_2^{00}|Y_1^0=0),$$

where the first equality follows from the observation rules, the second from (7), and the third from (6). Analogously for (12)

$$\mathbb{E}(Y_2|Y_1=0,D_1=0,D_2=1)=\mathbb{E}(Y_2^{01}|Y_1^0=0,D_1=0,D_2=1)=$$
$$\mathbb{E}(Y_2^{01}|Y_1^0=0,D_1=0)=\mathbb{E}(Y_2^{01}|Y_1^0=0).$$

Proof of Theorem 1

We express $\text{ATETS}_2^{11,00}$ as a function of the joint distribution of the potential outcomes

$$\text{ATETS}_2^{11,00}= \tag{A.1}$$
$$\frac{\sum_{y_2^{00}=0}^1\sum_{y_2^{01}=0}^1 p(1,y_2^{01},y_2^{00}|0,0)p(0,0)+\sum_{y_2^{00}=0}^1\sum_{y_2^{01}=0}^1 p(1,y_2^{01},y_2^{00}|0,1)p(0,1)}{\sum_{y_1^0=0}^1 p(0,y_1^0)}$$
$$-\frac{\sum_{y_2^{11}=0}^1\sum_{y_2^{01}=0}^1 p(y_2^{11},y_2^{01},1|0,0)p(0,0)+\sum_{y_2^{00}=0}^1\sum_{y_2^{01}=0}^1 p(y_2^{11},y_2^{01},1|0,1)p(0,1)}{\sum_{y_1^0=0}^1 p(0,y_1^0)},$$

because

$$\mathbb{E}(Y_2^{11}|Y_1^1=0)= \tag{A.2}$$
$$\frac{\sum_{y_2^{00}=0}^1\sum_{y_2^{01}=0}^1 p(1,y_2^{01},y_2^{00}|0,0)p(0,0)+\sum_{y_2^{00}=0}^1\sum_{y_2^{01}=0}^1 p(1,y_2^{01},y_2^{00}|0,1)p(0,1)}{\sum_{y_1^0=0}^1 p(0,y_1^0)},$$

and

$$\mathbb{E}(Y_2^{00}|Y_1^1=0)= \tag{A.3}$$
$$\frac{\sum_{y_2^{11}=0}^1\sum_{y_2^{01}=0}^1 p(y_2^{11},y_2^{01},1|0,0)p(0,0)+\sum_{y_2^{00}=0}^1\sum_{y_2^{01}=0}^1 p(y_2^{11},y_2^{01},1|0,1)p(0,1)}{\sum_{y_1^0=0}^1 p(0,y_1^0)}.$$

In the remainder of the Appendix we use the following notation

$$\begin{aligned} p^1(y_1^1) &\equiv \Pr(Y_1^1 = y_1^1), \\ p^0(y_1^0) &\equiv \Pr(Y_1^0 = y_1^0), \\ p^{11}(y_2^{11}|y_1^1, y_1^0) &\equiv \Pr(Y_2^{11} = y_2^{11}|Y_1^1 = y_1^1, Y_1^0 = y_1^0) \ , \qquad y_1^1, y_1^0, y_2^{11} = 0, 1, \\ p^{00}(y_2^{00}|y_1^1, y_1^0) &\equiv \Pr(Y_2^{00} = y_2^{00}|Y_1^1 = y_1^1, Y_1^0 = y_1^0) \ , \qquad y_1^1, y_1^0, y_2^{00} = 0, 1. \end{aligned}$$

From (15) we have that

$$\mathbb{E}(Y_2^{11}|Y_1^1 = 0) = \Pr(Y_2 = 1|D_1 = 1, Y_1 = 0). \tag{A.4}$$

So that if $\Pr(Y_1 = 0|D_1 = 1) > 0$ then $\mathbb{E}(Y_2^{11}|Y_1^1 = 0)$ is point-identified. If $\Pr(Y_1 = 0|D_1 = 1) = 0$ then $\mathbb{E}(Y_2^{11}|Y_1^1 = 0)$, $\mathbb{E}(Y_2^{11}|Y_1^1 = 0)$ and $\text{ATETS}_2^{11,00}$ is not defined.

Using the notation above we have

$$\mathbb{E}(Y_2^{00}|Y_1^1 = 0) = \frac{p^{00}(1|0,0)p(0,0) + p^{00}(1|0,1)p(0,1)}{\sum_{y_1^0=0}^{1} p(0, y_1^0)}. \tag{A.5}$$

We consider the cases $\Pr(Y_1 = 0|D_1 = 0) = 0$ and $\Pr(Y_1 = 0|D_1 = 0) > 0$ separately. If $\Pr(Y_1 = 0|D_1 = 0) > 0$ we know from (16)

$$p^{00}(1|0,0) = \frac{\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0)\Pr(Y_1 = 0|D_1 = 0) - p^{00}(1|1,0)p(1,0)}{p(0,0)}.$$

Then

$$\mathbb{E}(Y_2^{00}|Y_1^1 = 0) = \frac{\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0)\Pr(Y_1 = 0|D_1 = 0)}{\sum_{y_1^0=0}^{1} p(0, y_1^0)} - \frac{p^{00}(1|1,0)p(1,0) - p^{00}(1|0,1)p(0,1)}{\sum_{y_1^0=0}^{1} p(0, y_1^0)}.$$

From (13) $p(0,1) = \Pr(Y_1 = 0|D_1 = 1) - p(0,0)$ and (14) $p(1,0) = \Pr(Y_1 = 0|D_1 = 0) - p(0,0)$ and upon substitution

$$\mathbb{E}(Y_2^{00}|Y_1^1 = 0) = \frac{\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0)\Pr(Y_1 = 0|D_1 = 0)}{\Pr(Y_1 = 0|D_1 = 1)} - \frac{p^{00}(1|1,0)[\Pr(Y_1 = 0|D_1 = 0) - p(0,0)] - p^{00}(1|0,1)[\Pr(Y_1 = 0|D_1 = 1) - p(0,0)]}{\Pr(Y_1 = 0|D_1 = 1)}. \tag{A.6}$$

The expected value $\mathbb{E}(Y_2^{00}|Y_1^1 = 0)$ depends on the unknown probabilities $p^{00}(1|1,0)$, $p^{00}(1|0,1)$ and $p(0,0)$. Now note that, because $p(0,0) = \Pr(Y_1^0 = 0, Y_1^1 = 0) \leq \Pr(Y_1^0 = 0) = \Pr(Y_1 = 0|D_1 = 0)$ and because $p(0,0) = \Pr(Y_1^0 = 0, Y_1^1 = 0) \leq \Pr(Y_1^1 = 0) = \Pr(Y_1 =$

$0|D_1 = 1)$, the function on the right hand side is decreasing in $p^{00}(1|1,0)$ and increasing in $p^{00}(1|0,1)$. Therefore

$$\frac{(\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0) - 1)\Pr(Y_1 = 0|D_1 = 0) + p(0,0)}{\Pr(Y_1 = 0|D_1 = 1)} \tag{A.7}$$

$$\leq \mathbb{E}(Y_2^{00}|Y_1^1 = 0) \leq$$

$$\frac{\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0)\Pr(Y_1 = 0|D_1 = 0) - p(0,0)}{\Pr(Y_1 = 0|D_1 = 1)} + 1$$

where the lower bound applies if $p^{00}(1|1,0) = 1$ and $p^{00}(1|0,1) = 0$, and the upper bound if $p^{00}(1|1,0) = 0$ and $p^{00}(1|0,1) = 1$. The lower bound is increasing and the upper bound decreasing in $p(0,0)$. By the Bonferroni inequality

$$p(0,0) \geq \max\{\Pr(Y_1 = 0|D_1 = 1) + \Pr(Y_1 = 0|D_1 = 0) - 1, 0\}.$$

We consider the cases that $\Pr(Y_1 = 0|D_1 = 1) + \Pr(Y_1 = 0|D_1 = 0) - 1 > 0$ and that $\Pr(Y_1 = 0|D_1 = 1) + \Pr(Y_1 = 0|D_1 = 0) - 1 \leq 0$ separately. If $\Pr(Y_1 = 0|D_1 = 1) + \Pr(Y_1 = 0|D_1 = 0) - 1 \leq 0$ then the lower bound on $p(0,0)$ is 0. Then, by (A.5) we have $\mathbb{E}(Y_2^{00}|Y_1^1 = 0) = p^{00}(1|0,1)$ and since $p^{00}(1|0,1)$ is not restricted by the observed outcomes we have

$$0 \leq \mathbb{E}(Y_2^{00}|Y_1^1 = 0) \leq 1. \tag{A.8}$$

If $\Pr(Y_1 = 0|D_1 = 1) + \Pr(Y_1 = 0|D_1 = 0) - 1 > 0$ we have upon substitution of the lower bound on $p(0,0)$ in (A.7)

$$\max\left\{\frac{\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0)\Pr(Y_1 = 0|D_1 = 0) - 1}{\Pr(Y_1 = 0|D_1 = 1)} + 1, 0\right\} \tag{A.9}$$

$$\leq \mathbb{E}(Y_2^{00}|Y_1^1 = 0) \leq$$

$$\min\left\{\frac{1 - (1 - \Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0))\Pr(Y_1 = 0|D_1 = 0)}{\Pr(Y_1 = 0|D_1 = 1)}, 1\right\}.$$

Next, note that $\Pr(Y_1 = 0|D_1 = 0) = 0$ implies that $\Pr(Y_1 = 0|D_1 = 1) + \Pr(Y_1 = 0|D_1 = 0) - 1 \leq 0$.

Finally, we combine these bounds with the results for $\mathbb{E}(Y_2^{11}|Y_1^1 = 0)$ to obtain bounds on $\text{ATETS}_2^{11,00}$.

We now prove that the bounds are the best possible, i.e. for each (lower or upper) bound we find the parameters of the joint distribution of the potential outcomes $p(y_1^1, y_1^0)$ and $p(y_2^{11}, y_2^{01}, y_2^{00}|y_1^1, y_1^0)$ such that the bound is binding and satisfy (13)-(17) (5 restrictions). First, consider the upper bound on $\text{ATES}_2^{11,00}$ and when $\Pr(Y_2^{00} = 1|Y_1^1 = 0) = 0$ and $\Pr(Y_1 = 0|D_1 = 0) > 0$. This is equivalent to the following restrictions on the parameters

$$0 = \sum_{y_2^{11}=0}^{1}\sum_{y_2^{01}=0}^{1} p(y_2^{11}, y_2^{01}, 1|0,0) \tag{A.10}$$

$$\Leftrightarrow \qquad p(y_2^{11}, y_2^{01}, 1|0,0) = 0 \quad, y_2^{11} = 0,1,\ y_2^{01} = 0,1,$$

and

$$0 = \sum_{y_2^{00}=0}^{1} \sum_{y_2^{01}=0}^{1} p(y_2^{11}, y_2^{01}, 1|0,1) \tag{A.11}$$

$$\Leftrightarrow \qquad p(y_2^{11}, y_2^{01}, 1|0,1) = 0 \quad , y_2^{11} = 0,1, \;\; y_2^{01} = 0,1.$$

If we set

$$p(1,0,0|0,1) = p(0,0,1|1,0) = p(0,1,1|1,0) = p(0,1,0|1,0) = 0, \tag{A.12}$$

the restrictions on the remaining parameters are

$$\Pr(Y_2 = 1|D_1 = 1, Y_1 = 0) \tag{A.13}$$

$$= \frac{\sum_{y_2^{01}=0}^{1} p(1, y_2^{01}, 0|0,0)p(0,0) + p(1,1,0|0,1)p(0,1)}{\sum_{y_1^0=0}^{1} p(0, y_1^0)},$$

and

$$\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0) \tag{A.14}$$

$$= \frac{\sum_{y_2^{01}=0}^{1} p(1, y_2^{01}, 1|1,0)p(1,0)}{\sum_{y_1^1=0}^{1} p(y_1^1, 0)},$$

and

$$\Pr(Y_2 = 1|D_1 = 0, D_2 = 1, Y_1 = 0) \tag{A.15}$$

$$= \frac{\sum_{y_2^{11}=0}^{1} p(y_2^{11}, 1, 0|0,0)p(0,0) + \sum_{y_2^{00}=0}^{1} p(1,1,y_2^{00}|1,0)p(1,0)}{\sum_{y_1^1=0}^{1} p(y_1^1, 0)},$$

and

$$\Pr(Y_1 = 0|D_1 = 1) = \sum_{y_1^0=0}^{1} p(0, y_1^0), \tag{A.16}$$

and

$$\Pr(Y_1 = 0|D_1 = 0) = \sum_{y_1^1=0}^{1} p(y_1^1, 0). \tag{A.17}$$

We substitute (A.16) and (A.17) into (A.13)-(A.15) to obtain

$$\Pr(Y_2 = 1|D_1 = 1, Y_1 = 0) \tag{A.18}$$

$$= \frac{(p(1,0,0|0,0) + p(1,1,0|0,0))p(0,0) + p(1,1,0|0,1)(\Pr(Y_1 = 0|D_1 = 1) - p(0,0))}{\Pr(Y_1 = 0|D_1 = 1)},$$

and

$$\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0) \tag{A.19}$$

$$= \frac{(p(1,1,1|1,0) + p(1,0,1|1,0))(\Pr(Y_1 = 0|D_1 = 0) - p(0,0))}{\Pr(Y_1 = 0|D_1 = 0)},$$

and

$$\Pr(Y_2 = 1|D_1 = 0, D_2 = 1, Y_1 = 0) \tag{A.20}$$

$$= \frac{(p(0,1,0|0,0) + p(1,1,0|0,0))p(0,0)}{\Pr(Y_1 = 0|D_1 = 0)}$$
$$+ \frac{(p(1,1,0|1,0) + p(1,1,1|1,0))(\Pr(Y_1 = 0|D_1 = 0) - p(0,0)}{\Pr(Y_1 = 0|D_1 = 0)}.$$

We now find a solution if $p(0,0) = \Pr(Y_1 = 0|D_1 = 1) + \Pr(Y_1 = 0|D_1 = 0) - 1$, i.e. $p(0,0)$ is at its lower bound. This implies that

$$p(0,1) = 1 - \Pr(Y_1 = 0|D_1 = 0), \qquad p(1,0) = 1 - \Pr(Y_1 = 0|D_1 = 1).$$

Because $p(0,0) + p(1,0) + p(0,1) = 1$ this implies that $p(1,1) = 0$. Because in all cases $p(0,0)$ will be at the lower bound, these values for $p(0,0), p(1,0), p(0,1)$ and $p(1,1)$ apply throughout.

By (A.19) the choice of $p(0,0)$ implies

$$p(1,0,1|1,0) + p(1,1,1|1,0) \tag{A.21}$$
$$= \frac{\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0)\Pr(Y_1 = 0|D_1 = 0)}{1 - \Pr(Y_1 = 0|D_1 = 1)},$$

with the right hand side less than or equal to 1 if and only if the lower bound in (A.9) is 0. Next, (A.19) holds if

$$p(1,0,0|0,0) + p(1,1,0|0,0) = p(1,1,0|0,1) = \Pr(Y_2 = 1|D_1 = 1, Y_1 = 0). \tag{A.22}$$

Finally, (A.20) holds if

$$p(0,1,0|0,0) + p(1,1,0|0,0) = p(1,1,0|1,0) + p(1,1,1|1,0) \tag{A.23}$$
$$= \Pr(Y_2 = 1|D_1 = 0, D_2 = 1, Y_1 = 0).$$

If we set

$$p(1,1,1|1,0)$$
$$= \min\left\{\frac{\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0)\Pr(Y_1 = 0|D_1 = 0)}{1 - \Pr(Y_1 = 0|D_1 = 1)},\right.$$
$$\left.\Pr(Y_2 = 1|D_1 = 0, D_2 = 1, Y_1 = 0)\right\},$$

and

$$p(1,1,0|0,0)$$
$$= \min(\Pr(Y_2 = 1|D_1 = 1, Y_1 = 0), \Pr(Y_2 = 1|D_1 = 0, D_2 = 1, Y_1 = 0)),$$

and

$$p(0,0,0|0,1) + p(0,0,0|1,0) + p(1,0,0|1,0) = 0,$$

we have obtained a set of parameters that satisfies all the restrictions with for the remaining parameters

$$p(0,1,0|0,1) = 1 - \Pr(Y_2 = 1|D_1 = 1, Y_1 = 0),$$

$$p(1,1,0|1,0) = 1 - \frac{\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0)\Pr(Y_1 = 0|D_1 = 0)}{1 - \Pr(Y_1 = 0|D_1 = 1)},$$
$$p(0,0,0|0,0) = 1 - \Pr(Y_2 = 1|D_1 = 0, D_2 = 1, Y_1 = 0),$$

and if e.g.

$$p(1,1,1|1,0) = \Pr(Y_2 = 1|D_1 = 0, D_2 = 1, Y_1 = 0),$$

and

$$p(1,1,0|0,0) = \Pr(Y_2 = 1|D_1 = 1, Y_1 = 0),$$

then $p(1,1,0|1,0) = 0$ and $p(0,1,0|0,0) = 0$ and

$$p(1,0,0|0,0) = \Pr(Y_2 = 1|D_1 = 0, D_2 = 1, Y_1 = 0) - \Pr(Y_2 = 1|D_1 = 1, Y_1 = 0),$$

and

$$p(1,0,1|1,0) = \frac{\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0)\Pr(Y_1 = 0|D_1 = 0)}{1 - \Pr(Y_1 = 0|D_1 = 1)} - \Pr(Y_2 = 1|D_1 = 0, D_2 = 1, Y_1 = 0).$$

The cases with

$$p(1,1,1|1,0) = \frac{\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0)\Pr(Y_1 = 0|D_1 = 0)}{1 - \Pr(Y_1 = 0|D_1 = 1)},$$
$$p(1,1,0|0,0) = \Pr(Y_2 = 1|D_1 = 1, Y_1 = 0),$$

and

$$p(1,1,1|1,0) = \Pr(Y_2 = 1|D_1 = 0, D_2 = 1, Y_1 = 0),$$
$$p(1,1,0|0,0) = \Pr(Y_2 = 1|D_1 = 0, D_2 = 1, Y_1 = 0),$$

and

$$p(1,1,1|1,0) = \frac{\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0)\Pr(Y_1 = 0|D_1 = 0)}{1 - \Pr(Y_1 = 0|D_1 = 1)},$$
$$p(1,1,0|0,0) = \Pr(Y_2 = 1|D_1 = 0, D_2 = 1, Y_1 = 0),$$

are dealt with analogously. Besides these two cases we have to find the joint distributions of the potential outcomes for the case that $\Pr(Y_2^{00} = 1|Y_1^1 = 0) > 0$ and $\Pr(Y_1 = 0|D_1 = 0) > 0$. The derivation is analogous to the one above. We also have to find the joint distributions consistent with the lower bound. Again the derivation is analogous. If $\Pr(Y_1 = 0|D_1 = 0) = 0$ the sharpness of the bounds follows directly.

Proof of Theorem 2

In the remainder of the Appendix we use the following notation.

$$\begin{aligned}
p_t^d(1|0,0) &= \Pr(Y_t^d = 1|Y_{t-1}^1 = 0, \ldots, Y_1^1 = 0, Y_{t-1}^0 = 0, \ldots, Y_1^0 = 0), \\
p_t^d(1|0,k) &= \Pr(Y_t^d = 1|Y_{t-1}^1 = 0, \ldots, Y_1^1 = 0, Y_k^0 = 1, \ldots, Y_1^0 = 0), \\
p_t^d(1|k,0) &= \Pr(Y_t^d = 1|Y_k^1 = 1, \ldots, Y_1^1 = 0, Y_{t-1}^0 = 0, \ldots, Y_1^0 = 0),
\end{aligned}$$

where $d = 0, 1$ and $k = 1, \ldots, t-1$.

Under Assumption 3 on random assignment

$$\mathbb{E}[Y_t^1|\overline{Y}_{t-1}^1 = 0] = \Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 1). \tag{A.24}$$

So that if $\Pr(\overline{Y}_{t-1}^1 = 0|D_1 = 1) > 0$ then $\mathbb{E}[Y_t^1|\overline{Y}_{t-1}^1 = 0]$ is point-identified, and if $\Pr(\overline{Y}_{t-1}^1 = 0|D_1 = 1) = 0$ then $\mathbb{E}[Y_t^1|\overline{Y}_{t-1}^1 = 0], \mathbb{E}[Y_t^0|\overline{Y}_{t-1}^1 = 0]$ and $\text{ATETS}_t^{1,0}$ is not defined.

Next, we have

$$\mathbb{E}\left[Y_t^0|\overline{Y}_{t-1}^1 = 0\right] = \frac{p_t^0(1|0,0)p_{t-1}(0,0) + \sum_{k=1}^{t-1} p_t^0(1|0,k)p_{t-1}(0,k)}{p_{t-1}(0,0) + \sum_{k=1}^{t-1} p_{t-1}(0,k)}. \tag{A.25}$$

We consider cases that $\Pr(\overline{Y}_{t-1} = 0|D_1 = 0) = 0$ and $\Pr(\overline{Y}_{t-1} = 0|D_1 = 0) > 0$ separately. If $\Pr(\overline{Y}_{t-1}^0 = 0|D_1 = 0) > 0$ then under Assumption 3 on random assignment

$$\Pr(Y_t = 1, \overline{Y}_{t-1} = 0|\overline{D}_t = 0) = \Pr(Y_t^0 = 1, \overline{Y}_{t-1}^0 = 0|\overline{D}_t = 0) = \Pr(Y_t^0 = 1, \overline{Y}_{t-1}^0 = 0).$$

Now by the law of total probability

$$\begin{aligned}\Pr(Y_t^0 = 1, \overline{Y}_{t-1}^0 = 0) &= \Pr(\overline{Y}_{t-1}^1 = 0, Y_t^0 = 1, \overline{Y}_{t-1}^0 = 0)\\ &+ \sum_{k=1}^{t-1} \Pr(Y_k^1 = 1, \ldots, Y_1^1 = 0, Y_t^0 = 1, \overline{Y}_{t-1}^0 = 0)\\ &= p_t^0(1|0,0)p_{t-1}(0,0) + \sum_{k=1}^{t-1} p_t^0(1|k,0)p_{t-1}(k,0).\end{aligned}$$

Therefore (and using again that the treated state is absorbing)

$$\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 0) = \frac{p_t^0(1|0,0)p_{t-1}(0,0) + \sum_{k=1}^{t-1} p_t^0(1|k,0)p_{t-1}(k,0)}{\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right)}$$

Solving for $p_t^0(1|0,0)$ gives

$$p_t^0(1|0,0) = \frac{\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 0)\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right) - \sum_{k=1}^{t-1} p_t^0(1|k,0)p_{t-1}(k,0)}{p_{t-1}(0,0)}. \tag{A.26}$$

Then

$$\begin{aligned}&\mathbb{E}\left[Y_t^0|\overline{Y}_{t-1}^1 = 0\right]\\ &= \frac{\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 0)\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right)}{p_{t-1}(0,0) + \sum_{k=1}^{t-1} p_{t-1}(0,k)}\\ &- \frac{\sum_{k=1}^{t-1} p_t^0(1|k,0)p_{t-1}(k,0) - \sum_{k=1}^{t-1} p_t^0(1|0,k)p_{t-1}(0,k)}{p_{t-1}(0,0) + \sum_{k=1}^{t-1} p_{t-1}(0,k)}.\end{aligned}$$

The expression on the right-hand side is decreasing in $p_t^0(1|k,0)$ for all k and increasing in $p_t^0(1|0,k)$ for all k. The lower bound is obtained by setting $p_t^0(1|k,0)$ at 1 and $p_t^0(1|k,0)$ at 0 and the upper bound by setting $p_t^0(1|k,0)$ at 0 and $p_t^0(1|k,0)$ at 1

$$\frac{\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 0)\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right) - \sum_{k=1}^{t-1} p_{t-1}(k,0)}{p_{t-1}(0,0) + \sum_{k=1}^{t-1} p_{t-1}(0,k)}$$

$$\leq \mathbb{E}\left[Y_t^0|\overline{Y}_{t-1}^1 = 0\right] \leq$$

$$\frac{\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 0) \Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right) + \sum_{k=1}^{t-1} p_{t-1}(0,k)}{p_{t-1}(0,0) + \sum_{k=1}^{t-1} p_{t-1}(0,k)}.$$

Because

$$\Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1) = p_{t-1}(0,0) + \sum_{k=1}^{t-1} p_{t-1}(0,k)$$

and

$$\Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0) = p_{t-1}(0,0) + \sum_{k=1}^{t-1} p_{t-1}(k,0)$$

we have

$$\frac{[\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 0) - 1] \Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right) + p_{t-1}(0,0)}{\Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1)} \tag{A.27}$$

$$\leq \mathbb{E}\left[Y_t^0|\overline{Y}_{t-1}^1 = 0\right] \leq$$

$$\frac{\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 0) \Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right) - p_{t-1}(0,0)}{\Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1)} + 1.$$

The upper bound is decreasing and the lower bound is increasing in $p_{t-1}(0,0)$. Next, by the Bonferroni inequality

$$p_{t-1}(0,0) \geq$$

$$\max\left\{\Pr(Y_{t-1}^1 = 0, \ldots, Y_1^1 = 0) + \Pr(Y_{t-1}^0 = 0, \ldots, Y_1^0 = 0) - 1, 0\right\}.$$

Also with $Y_0 \equiv 0$

$$\Pr(Y_{t-1}^1 = 0, \ldots, Y_1^1 = 0) = \prod_{s=1}^{t-1} \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 1)$$

and

$$\Pr(Y_{t-1}^0 = 0, \ldots, Y_1^0 = 0) = \prod_{s=1}^{t-1} \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 0)$$

so that

$$p_{t-1}(0,0) \geq$$

$$\max\left\{\prod_{s=1}^{t-1} \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 1) + \prod_{s=1}^{t-1} \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 0) - 1, 0\right\} =$$

$$\max\left\{\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1\right) + \Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right) - 1, 0\right\}.$$

If

$$\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1\right) + \Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right) - 1 \leq 0$$

the lower bound on $p_{t-1}(0,0)$ is 0, so that since $p_t^0(1|k,0)$ for all k is not restricted by the observed outcomes we have using (A.25) that

$$0 \leq \mathbb{E}\left[Y_t^0|\overline{Y}_{t-1}^1 = 0\right] \leq 1. \tag{A.28}$$

If

$$\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1\right) + \Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right) - 1 > 0$$

we have upon substitution of the lower bound on $p_{t-1}(0,0)$ into (A.27)

$$\frac{\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 0)\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right) - 1}{\Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1)} + 1 \tag{A.29}$$

$$\leq \mathbb{E}\left[Y_t^0|\overline{Y}_{t-1}^1 = 0\right] \leq$$

$$\frac{1 - [1 - \Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 0)]\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right)}{\Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1)}.$$

Next, note that $\Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0) = 0$ implies that $\Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1) + \Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0) - 1 \leq 0$.

Finally, we combine these bounds with the results for $\mathbb{E}[Y_t^1|\overline{Y}_{t-1}^1 = 0]$ to obtain bounds on $\text{ATETS}_2^{1,0}$.

Proof of Theorem 3

We first consider bounds on $\Pr(Y_t^1 = 1|S_{t-1})$. We observe

$$\Pr(Y_t = 1|Y_{t-1} = 0, \ldots, Y_1 = 0, D_t = 1, \ldots, D_1 = 1)$$

$$= \Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 1).$$

Note that because treatment is absorbing it would suffice to condition on $D_1 = 1$. We keep the whole t vector $\overline{D}_t$ in the notation, but observe that $\overline{D}_t = 1 \Leftrightarrow D_1 = 1$. The 0 in the condition is a $t-1$ vector. Under Assumption 3 on random assignment

$$\Pr(Y_t = 1, \overline{Y}_{t-1} = 0|\overline{D}_t = 1) = \Pr(Y_t^1 = 1, \overline{Y}_{t-1}^1 = 0|\overline{D}_t = 1)$$

$$= \Pr(Y_t^1 = 1, \overline{Y}_{t-1}^1 = 0).$$

Now by the law of total probability

$$\Pr(Y_t^1 = 1, \overline{Y}_{t-1}^1 = 0) = \Pr(Y_t^1 = 1, \overline{Y}_{t-1}^1 = 0, \overline{Y}_{t-1}^0 = 0)$$

$$+ \sum_{k=1}^{t-1} \Pr(Y_t^1 = 1, \overline{Y}_{t-1}^1 = 0, Y_k^0 = 1, \ldots, Y_1^0 = 0)$$

$$= p_t^1(1|0,0)p_{t-1}(0,0) + \sum_{k=1}^{t-1} p_t^1(1|0,k)p_{t-1}(0,k).$$

Therefore (and using again that the treated state is absorbing)

$$\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 1) = \frac{p_t^1(1|0,0)p_{t-1}(0,0)}{\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1\right)}$$

$$+\frac{\sum_{k=1}^{t-1} p_t^1(1|0,k)p_{t-1}(0,k)}{\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1\right)}.$$

Solving for $p_t^1(1|0,0)$ gives

$$p_t^1(1|0,0) = \frac{\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 1)\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1\right)}{p_{t-1}(0,0)} \tag{A.30}$$

$$-\frac{\sum_{k=1}^{t-1} p_t^1(1|0,k)p_{t-1}(0,k)}{p_{t-1}(0,0)}.$$

The expression on the right-hand side is decreasing in $p_t^1(1|0,k)$ for all k. The lower bound is obtained by setting $p_t^1(1|0,k)$ at 1 and the upper bound by setting $p_t^1(1|0,k)$ at 0.

$$\frac{\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 1)\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1\right) - \sum_{k=1}^{t-1} p_{t-1}(0,k)}{p_{t-1}(0,0)}$$

$$\leq p_t^1(1|0,0) \leq \frac{\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 1)\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1\right)}{p_{t-1}(0,0)}.$$

Because

$$\Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1) = \Pr(\overline{Y}^1_{t-1} = 0) = p_{t-1}(0,0) + \sum_{k=1}^{t-1} p_{t-1}(0,k)$$

we have

$$\frac{\left(\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 1) - 1\right)\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1\right)}{p_{t-1}(0,0)} + 1$$

$$\leq p_t^1(1|0,0) \leq \frac{\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 1)\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1\right)}{p_{t-1}(0,0)}.$$

The upper bound is decreasing and the lower bound is increasing in $p_{t-1}(0,0)$, which is the probability of survival up to and including $t-1$ in both treatment arms. The final step is therefore to obtain a lower bound on $p_{t-1}(0,0)$. From the proof of theorem 2 we have

$$p_{t-1}(0,0) \geq \max\left\{\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1\right) + \Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right) - 1, 0\right\}. \tag{A.31}$$

If

$$\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1\right) + \Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right) - 1 \leq 0$$

then we are sure that there are survivors in both treatment arms. If this condition holds then substitution of this lower bound gives the result.

By an analogous argument we obtain the bounds on $\Pr(Y_t^0 = 1|Y_{t-1}^1, \ldots, Y_1^1 = 0, Y_{t-1}^0 = 0, \ldots, Y_1^0 = 0)$. The bounds on $\text{ATES}_t^{1,0}$ follow directly.

Proof of Theorem 4

Using similar reasoning as above we have that if $\Pr(\overline{Y}^1_{t-1} = 0|D_1 = 1) = 0$ then $\text{ATETS}^{1,0}_t$ is not defined. If $\Pr(\overline{Y}^1_{t-1} = 0|D_1 = 1) > 0$ we have from (A.24) and (A.27)

$$\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 1)$$
$$-\frac{\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 0)\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right) - p_{t-1}(0,0)}{\Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1)} - 1$$
$$\leq \text{ATETS}^{1,0}_t \leq$$
$$\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 1)$$
$$-\frac{[\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 0) - 1]\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right) + p_{t-1}(0,0)}{\Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1)}.$$

Because the lower bound is increasing in $p_{t-1}(0,0)$ and the upper bound decreasing in $p_{t-1}(0,0)$ we need the lower bound on this probability. We have

$$p_{t-1}(0,0) = \Pr(Y^1_{t-1} = 0, \ldots, Y^1_1 = 0, Y^0_{t-1} = 0, \ldots, Y^0_1 = 0) =$$
$$\Pr(Y^1_{t-1} = 0, Y^0_{t-1} = 0|S_{t-2})\Pr(Y^1_{t-2} = 0, \ldots, Y^1_1 = 0, Y^0_{t-2} = 0, \ldots, Y^0_1 = 0).$$

By Assumption 2 either

$$\Pr\left(Y^1_{i,t-1} = 0 S_{i,t-2}\right) \leq \Pr\left(Y^0_{i,t-1} = 0 S_{i,t-2}\right), \tag{A.32}$$

or

$$\Pr\left(Y^1_{i,t-1} = 0 S_{i,t-2}\right) > \Pr\left(Y^0_{i,t-1} = 0 S_{i,t-2}\right), \tag{A.33}$$

for all i. Assume that (A.32) holds. By Assumption 3 this implies that

$$\Pr(Y^1_{i,t-1} = 0, Y^0_{i,t-1} = 1|S_{i,t-2}) = 0,$$

so that

$$\Pr(Y^1_{i,t-1} = 0|S_{i,t-2}) = \Pr(Y^1_{i,t-1} = 0, Y^0_{i,t-1} = 0|S_{i,t-2}) + \Pr(Y^1_{i,t-1} = 0, Y^0_{i,t-1} = 1|S_{i,t-2})$$
$$= \Pr(Y^1_{i,t-1} = 0, Y^0_{i,t-1} = 0|S_{i,t-2}).$$

Because Assumptions 2 and 3 hold for all t we find by recursion that under that assumption for all i

$$\Pr(Y^1_{i,t-1} = 0, \ldots, Y^1_{i1} = 0, Y^0_{i,t-1} = 0, \ldots, Y^0_{i1} = 0) = \prod_{s=1}^{t-1}\Pr(Y^1_{is} = 0|\overline{Y}^1_{i,s-1} = 0),$$

so that

$$p_{t-1}(0,0) = \prod_{s=1}^{t-1}\Pr(Y^1_s = 0|\overline{Y}^1_{s-1} = 0) = \prod_{s=1}^{t-1}\Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 1).$$

If Assumption 2 holds with (A.33), then

$$p_{t-1}(0,0)=\prod_{s=1}^{t-1}\Pr(Y_s^0=0|\overline{Y}_{s-1}^0=0)=\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=0).$$

We conclude that

$$p_{t-1}(0,0)\geq\min\left\{\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=1),\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=0)\right\}=$$

$$\min\left\{\Pr(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=1),\Pr(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=0)\right\}$$

and substitution gives the bounds.

Proof of Theorem 5

Using similar reasoning as above we have that if $\Pr(\overline{Y}_{t-1}^1=0|D_1=1)=0$ then $\text{ATETS}_t^{1,0}$ is not defined. If $\Pr(Y_1=0|D_1=1)>0$ we have under Assumption 3 on random assignment

$$\Pr(Y_t=1|\overline{Y}_{t-1}=0,\overline{D}_t=1)=\mathbb{E}[Y_t^1|\overline{Y}_{t-1}^1=0].$$

Next,

$$\mathbb{E}\left[Y_t^0|\overline{Y}_{t-1}^1=0\right]=\frac{p_t^0(1|0,0)p_{t-1}(0,0)+\sum_{k=1}^{t-1}p_t^0(1|0,k)p_{t-1}(0,k)}{p_{t-1}(0,0)+\sum_{k=1}^{t-1}p_{t-1}(0,k)},$$

which is an increasing function of $p_t^0(1|0,k)$ for all k. Since $p_t^0(1|0,k)$ is not restricted by the observed outcomes the upper bound on $\mathbb{E}\left[Y_t^0|\overline{Y}_{t-1}^1=0\right]$ is obtained if $p_t^0(1|0,k)=1$ for all k and by Assumption 4 the lower bound is obtained if $p_t^0(1|0,k)=p_t^0(1|0,0)$ for all k . Then

$$p_t^0(1|0,0)\leq\mathbb{E}\left[Y_t^0|\overline{Y}_{t-1}^1=0\right]\leq\frac{p_t^0(1|0,0)p_{t-1}(0,0)+\sum_{k=1}^{t-1}p_{t-1}(0,k)}{p_{t-1}(0,0)+\sum_{k=1}^{t-1}p_{t-1}(0,k)}.$$

By the proof of theorem 2 we have

$$\Pr(Y_t=1|\overline{Y}_{t-1}=0,\overline{D}_t=0)\Pr\left(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=0\right)\tag{A.34}$$

$$=p_t^0(1|0,0)p_{t-1}(0,0)+\sum_{k=1}^{t-1}p_t^0(1|k,0)p_{t-1}(k,0),$$

and upon substitution

$$\frac{\Pr(Y_t=1|\overline{Y}_{t-1}=0,\overline{D}_t=0)\Pr\left(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=0\right)-\sum_{k=1}^{t-1}p_t^0(1|k,0)p_{t-1}(k,0)}{p_{t-1}(0,0)}\tag{A.35}$$

$$\leq\mathbb{E}\left[Y_t^0|\overline{Y}_{t-1}^1=0\right]\leq$$

$$\frac{\Pr(Y_t=1|\overline{Y}_{t-1}=0,\overline{D}_t=0)\Pr\left(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=0\right)}{p_{t-1}(0,0)+\sum_{k=1}^{t-1}p_{t-1}(0,k)}$$

$$-\frac{\sum_{k=1}^{t-1} p_t^0(1|k,0)p_{t-1}(k,0)-\sum_{k=1}^{t-1} p_{t-1}(0,k)}{p_{t-1}(0,0)+\sum_{k=1}^{t-1} p_{t-1}(0,k)}.$$

Both the lower and upper bound is decreasing in $p_t^0(1|k,0)$ for all k. Therefore the lower bound on $\mathbb{E}[Y_t^0|\overline{Y}_{t-1}^1=0]$ is obtained if $p_t^0(1|k,0)=1$ for all k and by Assumption 4 the upper bound is obtained if $p_t^0(1|k,0)=p_t^0(1|0,0)$ for all k. Upon substitution of $p_t^0(1|k,0)=p_t^0(1|0,0)$ for all k in (A.34) and using that $\Pr(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=0)=p_{t-1}(0,0)+\sum_{k=1}^{t-1} p_{t-1}(k,0)$ we have

$$p_t^0(1|0,0)=\Pr(Y_t=1|\overline{Y}_{t-1}=0,\overline{D}_t=0). \tag{A.36}$$

Upon substitution of (A.36) into (A.35) and using that $\Pr(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=1)=p_{t-1}(0,0)+\sum_{k=1}^{t-1} p_{t-1}(0,k)$ and $\Pr(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=0)=p_{t-1}(0,0)+\sum_{k=1}^{t-1} p_{t-1}(k,0)$ we have

$$\frac{(\Pr(Y_t=1|\overline{Y}_{t-1}=0,\overline{D}_t=0)-1)\Pr\left(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=0\right)}{p_{t-1}(0,0)}+1 \tag{A.37}$$

$$\leq \mathbb{E}\left[Y_t^0|\overline{Y}_{t-1}^1=0\right]\leq$$

$$\frac{(\Pr(Y_t=1|\overline{Y}_{t-1}=0,\overline{D}_t=0)-1)p_{t-1}(0,0)}{\Pr(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=1)}+1.$$

The lower bound is increasing and the upper bound decreasing in $p_{t-1}(0,0)$. Assumption 4 also improves on the Bonferroni inequality for $p_{t-1}(0,0)$. We have

$$p_{t-1}(0,0)=\prod_{s=1}^{t-1}\Pr(Y_s^1=0,Y_s^0=0|S_{s-1}).$$

By the Bonferroni inequality and the results above

$$\Pr(Y_s^1=0,Y_s^0=0|S_{s-1})\geq \max\{1-\Pr(Y_s^1=1|S_{s-1})-\Pr(Y_s^0=1|S_{s-1}),0\}\geq$$
$$\max\{1-\Pr(Y_s=1|\overline{Y}_{s-1}=0,\overline{D}_s=1)-\Pr(Y_s=1|\overline{Y}_{s-1}=0,\overline{D}_s=0),0\}=$$
$$\max\{\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=1)+\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=0)-1,0\},$$

so that

$$p_{t-1}(0,0)\geq\prod_{s=1}^{t-1}\max\{\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=1)+\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=0)-1,0\}. \tag{A.38}$$

We compare this to the lower bound

$$\max\left\{\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=1)+\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=0)-1,0\right\}$$

that we obtained in the proof of Theorem 2. First, if there is an $1\leq s\leq t-1$ so that

$$\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=1)+\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=0)-1<0,$$

then

$$\prod_{s=1}^{t-1} \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 1) + \prod_{s=1}^{t-1} \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 0) - 1 < 0,$$

so that if the new lower bound is 0, so is the previous one. Finally, if for all $s = 1, \ldots, t-1$

$$\Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 1) + \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 0) - 1 > 0,$$

then

$$\prod_{s=1}^{t-1} \left[\Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 1) + \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 0) - 1\right] \geq$$

$$\prod_{s=1}^{t-1} \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 1) + \prod_{s=1}^{t-1} \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 0) - 1.$$

Upon substitution of (A.38) into (A.37) and considering cases $\Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 1) + \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 0) - 1 > 0$ for all $s = 1, \ldots, t-1$ and $\Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 1) + \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 0) - 1 \leq 0$ for some $s \leq t$ separately, and substitution gives the bounds.

Proof of Theorem 6

Under Assumption 3 on random assignment $\mathbb{E}[Y_t^1|\overline{Y}_{t-1}^1 = 0] = \Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 1)$, and by the proof of Theorem 5 we have under Assumption 4

$$\frac{(\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 0) - 1) \Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0\right)}{p_{t-1}(0,0)} + 1 \tag{A.39}$$

$$\leq \mathbb{E}\left[Y_t^0|\overline{Y}_{t-1}^1 = 0\right] \leq$$

$$\frac{(\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 0) - 1)p_{t-1}(0,0)}{\Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1)} + 1.$$

By the proof of Theorem 4 we have under Assumptions 2 and 3

$$p_{t-1}(0,0) \geq \min\left\{\Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1), \Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0)\right\}.$$

Collecting the results and substitution gives the bounds.