IFAU – INSTITUTE FOR
LABOUR MARKET POLICY
EVALUATION

Bounds on treatment effects on transitions

Geert Ridder
Johan Vikström

WORKING PAPER 2011:6

The Institute for Labour Market Policy Evaluation (IFAU) is a research institute under the Swedish Ministry of Employment, situated in Uppsala. IFAU's objective is to promote, support and carry out scientific evaluations. The assignment includes: the effects of labour market policies, studies of the functioning of the labour market, the labour market effects of educational policies and the labour market effects of social insurance policies. IFAU shall also disseminate its results so that they become accessible to different interested parties in Sweden and abroad.

IFAU also provides funding for research projects within its areas of interest. The deadline for applications is October 1 each year. Since the researchers at IFAU are mainly economists, researchers from other disciplines are encouraged to apply for funding.

IFAU is run by a Director-General. The institute has a scientific council, consisting of a chairman, the Director-General and five other members. Among other things, the scientific council proposes a decision for the allocation of research grants. A reference group including representatives for employer organizations and trade unions, as well as the ministries and authorities concerned is also connected to the institute.

Postal address: P.O. Box 513, 751 20 Uppsala
Visiting address: Kyrkogårdsgatan 6, Uppsala
Phone: +46 18 471 70 70
Fax: +46 18 471 70 71
ifau@ifau.uu.se
www.ifau.se

Papers published in the Working Paper Series should, according to the IFAU policy, have been discussed at seminars held at IFAU and at least one other academic forum, and have been read by one external and one internal referee. They need not, however, have undergone the standard scrutiny for publication in a scientific journal. The purpose of the Working Paper Series is to provide a factual basis for public policy and the public policy discussion.

ISSN 1651-1166

Bounds On Treatment Effects On Transitions*

Geert Ridder[†] Johan Vikström[‡]

March 18, 2011

Abstract

This paper considers the definition and identification of treatment effects on conditional transition probabilities. We show that even under sequential 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, that would narrow the bounds or even allow for point identification. 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, and seminar participants at IFAU-Uppsala, Uppsala University, University of Aarhus and University of Mannheim. Financial support of the Tom Hedelius Foundation and the NSF grants SES 0819612 and 0819638 is acknowledged.

[†]University of Southern California, ridder@usc.edu.
[‡]IFAU-Uppsala and UCLS, johan.vikstrom@ifau.uu.se.

1 Introduction

We consider the effect of an intervention where 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 den Berg (2003) who use the Mixed Proportional Hazard (MPH) model (their analysis is generalized to a multi-state 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. 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-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 show that additional (weak) assumptions like monotone treatment response and positively correlated outcomes tighten the bounds considerably.

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, Fougere, and Serandon (1997), Gritz (1993), Ham and LaLonde (1996), Abbring and 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. Fredriksson and Johansson (2008) have shown how the effect on the cdf, that is the unconditional survival probability, can be identified even if the treatment can start at any point in time. 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.

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. First, there 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. 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, Card, Fertig, Gra, Jacobi, Jensen, Leetmaa, Nima, Patacchini, Schmidt, Klaauw, and Weber (2007). In other cases a treatment consist of a sequence of sub-treatments assigned at pre-specified points in time to the survivors in the state. If one is interested in disentangling the sub-treatment effects, the treatment effect over the spell has to be investigated.

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. Section 5 illustrates the bounds for a job bonus experiment. Section 6 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 $\overline{d}(t) = \{d(s), 0 \le s \le t\}$ be the treatment status up to time t. The MPH model implies that the population distribution of the time to transition $T^{\overline{d}(T)}$ where the superscript is the relevant treatment history[1], has density

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

and distribution function

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

The hazard/transition rate given the treatment history is

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

To define treatment effects in the MPH model we compare groups with different treatment histories $\overline{d}(t)$. Let $\overline{d}_0(t)$ and $\overline{d}_1(t)$ be two such histories. We can compare either the average time-to-transition distribution functions in t, i.e. $F(t|\overline{d}_0(t))$ and $F(t|\overline{d}_1(t))$, or the average transition rates in t, i.e. $\theta(t|\overline{d}_0(t))$ and $\theta(t|\overline{d}_1(t))$. 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 in 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 $\mathbb{E}_V\left[V|T^{\bar{d}_0(T)} \geq t\right]$ and $\mathbb{E}_V\left[V|T^{\bar{d}_1(T)} \geq t\right]$ 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

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. When defining a causal effect 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.

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 propose to average over the subpopulation of individuals who would have survived until time t under both treatment arms. The individuals in this population have the same survival experience and any difference between the transition rates must be due to the effect of the treatment.[2]

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.

As emphasized we are interested in conditional treatment effects, i.e. treatment effects defined for the survivors in t. Let $\overline{d}_{0t}$ and $\overline{d}_{1t}$ be two specific treatment histories. To make the survival experience the same for all survivors we average over the hypothetical subpopulation of individuals who

[2] We could have restored comparability by averaging the transition probability of the controls at t over the subpopulation of the treated survivors to t. This is problematic because the transition probability is only defined for survivors up to t in the control group. For this reason we average over the subpopulation that would have survived under treatment and under no treatment.

would have survived until t under both $\overline{d}_{0t}$ and $\overline{d}_{1t}$. 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 Survivors (ATES) defined by*

$$\mathrm{ATES}_t^{\overline{d}_{1t},\overline{d}_{0t}} = \tag{2}$$
$$\mathbb{E}\left(Y_t^{\overline{d}_{1t}} | Y_{t-1}^{\overline{d}_{1,t-1}} = 0, \dots, Y_1^{\overline{d}_{11}} = 0, Y_{t-1}^{\overline{d}_{0,t-1}} = 0, \dots, Y_1^{\overline{d}_{01}} = 0\right) -$$
$$\mathbb{E}\left(Y_t^{\overline{d}_{0t}} | Y_{t-1}^{\overline{d}_{1,t-1}} = 0, \dots, Y_1^{\overline{d}_{11}} = 0, Y_{t-1}^{\overline{d}_{0,t-1}} = 0, \dots, Y_1^{\overline{d}_{01}} = 0\right).$$

Obvious choices for $\overline{d}_{1t}$ and $\overline{d}_{0t}$ are $\overline{d}_{1t} = (0, \dots, 0, 1, \dots, 1)$ with the first 1 at position τ, and $\overline{d}_{0t} = (0, \dots, 0)$. If we make the usual assumption that there is no effect of the treatment before it starts[3], then for these two treatments $\mathrm{ATES}_t^{\overline{d}_{1t},\overline{d}_{0t}} = 0, t = 1, \dots, \tau - 1$. The differential selection only starts after the treatment begins, so that this property of the ATES_t is consistent with that fact. After the treatment starts there is dynamic selection and the $\mathrm{ATES}_t^{\overline{d}_{1t},\overline{d}_{0t}}$ controls for that by comparing the transition rates for individuals with a common survival experience.

3 Identification of average treatment effects on transitions under random assignment

We now consider identification of the $\mathrm{ATES}_t^{\overline{d}_{1t},\overline{d}_{0t}}$ under random treatment assignment. We first need to define what we mean by random assignment in this case. Let D_t be the indicator that treatment is assigned in period t, i.e. the unit is not treated in periods $1, \dots, t-1$, 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 $t, t+1, \dots$. The control treatment $\overline{d}_{0t}$ is no treatment up to and including t. We distinguish between three types of randomized assignment

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

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

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

Assumption 2 (Sequential randomization) *For all t and $\overline{d}_s, s \geq t$ with the first $t-1$ components equal to 0*

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

Assumption 3 (Sequential randomization among survivors) *For all t and $\overline{d}_s, s \geq t$ with the first $t-1$ components equal to 0*

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

Under Assumption 1, 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 Assumption 2 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 Assumption 3 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 Assumption 3.

Initially we consider the two period case where the transition occurs in period 1, period 2 or after period 2. The main points 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

$$Y_1 = D_1 Y_1^1 + (1 - D_1) Y_1^0 \tag{3}$$

and

$$Y_2 = D_1 Y_2^{11} + (1 - D_1) D_2 Y_2^{01} + (1 - D_1)(1 - D_2) Y_2^{00}. \tag{4}$$

Because treatment is an absorbing state

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

Assumption 3 is in this case

$$D_1 \perp Y_1^1, Y_1^0, Y_2^{11}, Y_2^{01}, Y_2^{00} \tag{5}$$

and

$$D_2 \perp Y_2^{01}, Y_2^{00} | D_1 = 0, Y_1^0 = 0. \tag{6}$$

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

Lemma 1 *If Assumption 3 holds, then*

$$\mathbb{E}(Y_1|D_1 = 1) = \mathbb{E}(Y_1^1) \tag{7}$$

$$\mathbb{E}(Y_1|D_1 = 0) = \mathbb{E}(Y_1^0) \tag{8}$$

$$\mathbb{E}(Y_2|Y_1 = 0, D_1 = 1) = \mathbb{E}(Y_2^{11}|Y_1^1 = 0) \tag{9}$$

$$\mathbb{E}(Y_2|Y_1 = 0, D_1 = 0, D_2 = 0) = \mathbb{E}(Y_2^{00}|Y_1^0 = 0) \tag{10}$$

$$\mathbb{E}(Y_2|Y_1 = 0, D_1 = 0, D_2 = 1) = \mathbb{E}(Y_2^{01}|Y_1^0 = 0). \tag{11}$$

Proof See Appendix. □

3.1 Identification of instantaneous treatment effects

The interpretation of the $\text{ATES}_t^{\overline{d}_{1t},\overline{d}_{0t}}$ 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

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

and

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

Note that for $\text{ATES}_2^{01,00}$ the treatment in the first period is the same (no treatment) in both treatment arms, so that survival in period 1 in both treatment arms is equivalent to surviving the first period under no treatment. Therefore the definition in (2) applies.

From equations (7) and (8) it follow that under Assumption 3 we can point identify the first period instantaneous treatment effect

$$\text{ATES}_1^{1,0} = \text{ATE}_1^{1,0} = \mathbb{E}(Y_1^1) - \mathbb{E}(Y_1^0) = \mathbb{E}(Y_1|D_1 = 1) - \mathbb{E}(Y_1|D_1 = 0), \quad (12)$$

and from equations (10) and (11) that we can also point identify the second period instantaneous treatment effect

$$\text{ATES}_2^{01,00} = \mathbb{E}(Y_2^{01}|Y_1^0 = 0) - \mathbb{E}(Y_2^{00}|Y_1^0 = 0) = \quad (13)$$

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

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 ATES is therefore

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

that is the average treatment effect in the second period from treatment started in the first period for those who survive under both treatment and no treatment in the first period. Because all that can be deduced from the data is in equations (7)-(11), which hold under Assumption 3, $\text{ATES}_2^{11,00}$ 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 $\text{ATES}_2^{11,00}$. 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 $Y_1^0, Y_1^1, Y_2^{00}, Y_2^{01}, Y_2^{11}$. Note that because treatment is an absorbing state, Y_2^{10} is not defined. In addition if $Y_1^1 = 1$, then Y_2^{11} is not defined, and if $Y_1^0 = 1$, then neither Y_2^{01} nor Y_2^{00} is 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

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

The dynamic treatment effect can be expressed as a function of these probabilities

$$ATES_2^{11,00} = \sum_{y_2^{00}=0}^{1} \sum_{y_2^{01}=0}^{1} p(1, y_2^{01}, y_2^{00}|0,0) - \sum_{y_2^{11}=0}^{1} \sum_{y_2^{01}=0}^{1} p(y_2^{11}, y_2^{01}, 1|0,0) \quad (14)$$

because

$$\mathbb{E}(Y_2^{11}|Y_1^1 = 0, Y_1^0 = 0) = \sum_{y_2^{00}=0}^{1} \sum_{y_2^{01}=0}^{1} p(1, y_2^{01}, y_2^{00}|0,0) \quad (15)$$

and

$$\mathbb{E}(Y_2^{00}|Y_1^1 = 0, Y_1^0 = 0) = \sum_{y_2^{11}=0}^{1} \sum_{y_2^{01}=0}^{1} p(y_2^{11}, y_2^{01}, 1|0,0). \quad (16)$$

By Assumption 3 the observed first period transition probabilities impose the restrictions

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

and

$$\Pr(Y_1 = y_1|D_1 = 0) = \sum_{y_1^1=0}^{1} p(y_1^1, y_1). \quad (18)$$

By Assumption 3 the observed second period transition probabilities impose the restrictions

$$\Pr(Y_2 = y_2|D_1 = 1, Y_1 = 0) = \quad (19)$$

$$\frac{\sum_{y_2^{01}=0}^{1} \sum_{y_2^{00}=0}^{1} p(y_2, y_2^{01}, y_2^{00}|0,0) p(0,0) + p(y_2|0,1) p(0,1)}{\sum_{y_1^0=0}^{1} p(0, y_1^0)}$$

and

$$\Pr(Y_2 = y_2|D_1 = 0, D_2 = 0, Y_1 = 0) = \tag{20}$$

$$\frac{\sum_{y_2^{11}=0}^{1}\sum_{y_2^{01}=0}^{1} p(y_2^{11}, y_2^{01}, y_2|0,0)p(0,0) + \sum_{y_2^{01}=0}^{1} p(y_2^{01}, y_2|1,0)p(1,0)}{\sum_{y_1^1=0}^{1} p(y_1^1, 0)}$$

and

$$\Pr(Y_2 = y_2|D_1 = 0, D_2 = 1, Y_1 = 0) = \tag{21}$$

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

Theorem 1 (Bounds on ATES) *Suppose that Assumption 3 holds. If* $\Pr(Y_1 = 0|D_1 = 1) + \Pr(Y_1 = 0|D_1 = 0) - 1 \leq 0$*, then* $\mathbb{E}[Y_2^{11}|Y_1^1 = 0, Y_1^0 = 0], \mathbb{E}[Y_2^{00}|Y_1^1 = 0, Y_1^0 = 0]$*, and* $ATES_2^{11,00}$ *are not defined. If* $\Pr(Y_1 = 0|D_1 = 1) + \Pr(Y_1 = 0|D_1 = 0) - 1 > 0$*, then we have the following sharp bounds*

$$\max\left\{\frac{[\Pr(Y_2 = 1|D_1 = 1, Y_1 = 0) - 1]\Pr(Y_1 = 0|D_1 = 1)}{\Pr(Y_1 = 0|D_1 = 1) + \Pr(Y_1 = 0|D_1 = 0) - 1} + 1, 0\right\} \tag{22}$$

$$\leq \mathbb{E}[Y_2^{11}|Y_1^1 = 0, Y_1^0 = 0] \leq$$

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

and

$$\begin{gathered}\max\left\{\frac{[\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0) - 1]\Pr(Y_1 = 0|D_1 = 0)}{\Pr(Y_1 = 0|D_1 = 1) + \Pr(Y_1 = 0|D_1 = 0) - 1} + 1, 0\right\} \\ \leq \mathbb{E}[Y_2^{00}|Y_1^1 = 0, Y_1^0 = 0] \leq \\ \min\left\{\frac{\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0)\Pr(Y_1 = 0|D_1 = 0)}{\Pr(Y_1 = 0|D_1 = 1) + \Pr(Y_1 = 0|D_1 = 0) - 1}, 1\right\}\end{gathered} \tag{23}$$

and

$$\max\left\{\frac{[\Pr(Y_2 = 1|D_1 = 1, Y_1 = 0) - 1]\Pr(Y_1 = 0|D_1 = 1)}{\Pr(Y_1 = 0|D_1 = 1) + \Pr(Y_1 = 0|D_1 = 0) - 1} + 1, 0\right\} - \tag{24}$$

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

$$\leq ATES_2^{11,00} \leq$$

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

$$\max\left\{\frac{[\Pr(Y_2=1|D_1=0,D_2=0,Y_1=0)-1]\Pr(Y_1=0|D_1=0)}{\Pr(Y_1=0|D_1=1)+\Pr(Y_1=0|D_1=0)-1}+1,0\right\}$$

Proof see Appendix. □

Theorem 1 provides closed form expressions for the sharp bounds on $\text{ATES}_2^{11,00}$. These bounds require no assumptions beyond sequential random assignment among survivors. They allow for arbitrary heterogeneity in treatment response. The bounds only exist if $\Pr(Y_1=0|D_1=1)+\Pr(Y_1=0|D_1=0)-1>0$, because if this holds we are sure that some members of the population survive the initial period irrespective of whether they were treated in that period or not. This follows because by the Bonferroni inequality the fraction of the population the survives the first period in both treatment arms is bounded by

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

Since the $\text{ATES}_2^{11,00}$ is defined for this subpopulation it cannot be defined if the subpopulation has no members. The condition puts an upper bound on the degree of dynamic selection. It can always be satisfied by choosing the initial period sufficiently short.

Inspection of the bounds shows that the upper and lower bound are equal if either all treated or controls or both survive the first period.

Corollary 1 (Point identification) *The lower and upper bounds on* $\mathbb{E}(Y_2^{00}|Y_1^1=0,Y_1^0=0)$ *are both equal* $\Pr(Y_2=1|D_1=0,D_2=0,Y_1=0)$ *if and only if* $\Pr(Y_1=0|D_1=1)=1$ *and analogously the lower and upper bounds on* $\mathbb{E}(Y_2^{11}|Y_1^1=0,Y_1^0=0)$ *are both equal to* $\Pr(Y_2=1|D_1=1,Y_1=0)$ *if* $\Pr(Y_1=0|D_1=0)=1$. *Therefore* $\text{ATES}_2^{11,00}$ *is point identified if and only if both* $\Pr(Y_1=0|D_1=0)=1$ *and* $\Pr(Y_1=0|D_1=1)=1$.

Corollary 1 shows that if there is no dynamic selection, i.e. if $\Pr(Y_1=0|D_1=1)=1$ and $\Pr(Y_1=0|D_1=1)=1$, the dynamic treatment effect $ATES_2^{11,00}$ is point identified. If everyone survives the first period we have under random treatment assignment two directly comparable groups even in the second period. Note that $\mathbb{E}(Y_2^{11}|Y_1^1=0,Y_1^0=0)$ is point identified if $\Pr(Y_1 = 0|D_1 = 0) = 1$, and $\mathbb{E}(Y_2^{00}|Y_1^1 = 0, Y_1^0 = 0)$ is point identified if $\Pr(Y_1 = 0|D_1 = 1) = 1$. For instance, the average transition probability in period 2 under no treatment for those who survive period 1 under both treatment and without treatment is point identified if no one exits in the treatment group, but there is no restriction of the probability of a transition in the control group. 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 is

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

3.2.2 Arbitrary number of periods

In the case of an arbitrary number of periods we only need to 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. Here 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 is $ATES_t^{1,0}$ 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

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

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. Note again that the superscripts 1 and 0 stand for vectors of 1 and 0 of appropriate length. In the sequel we use the notation $\overline{Y}_t = (Y_t, \ldots, Y_1)'$ that also applies to other variables. The bounds are given in the next theorem.

Theorem 2 (Bounds on ATES) *Suppose that Assumption 3 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* $\mathbb{E}[Y_t^0|S_{t-1}], \mathbb{E}[Y_t^1|S_{t-1}]$, *and* $ATES_t^{1,0}$ *are 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*

$$\max\left\{\frac{\left(\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 1) - 1\right)\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1\right)}{\prod_{s=1}^{t-1} \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 1) + \prod_{s=1}^{t-1} \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 0) - 1} + 1, 0\right\} \leq \tag{25}$$

$$\mathbb{E}[Y_t^1|S_{t-1}] \leq$$

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

and

$$\max\left\{\frac{(\Pr(Y_t=1|\overline{Y}_{t-1}=0,\overline{D}_t=0)-1)\Pr\left(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=0\right)}{\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=1)+\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=0)-1}+1,0\right\} \leq \tag{26}$$

$$\mathbb{E}[Y_t^0|S_{t-1}] \leq$$

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

and

$$\max\left\{\frac{(\Pr(Y_t=1|\overline{Y}_{t-1}=0,\overline{D}_t=1)-1)\Pr\left(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=1\right)}{\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=1)+\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=0)-1}+1,0\right\} - \tag{27}$$

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

$$ATES_t^{1,0} \leq$$

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

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

Proof see Appendix. □

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, but also in the discrete duration models and structural models that we consider in this section. A general discrete duration model for the control and treated outcomes is

$$\begin{aligned} Y_{it}^0 &= I(\alpha_t + V_i - \varepsilon_{it} \geq 0) \\ Y_{it}^1 &= I(\alpha_t + \gamma_{it} + V_i - \varepsilon_{it} \geq 0). \end{aligned} \tag{28}$$

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 onV_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 4 (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 4 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 4 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 ATES 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 assumes that the outcomes in all treatment arms involve the same random shocks. Consider the discrete duration model in (28). 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 5 (Common Shocks (CS)) *For all i, t and treatment paths $\overline{d}_0(t)$ and $\overline{d}_1(t)$*

$$\Pr(Y_{it}^{\overline{d}_{1t}} = 0|S_{i,t-1}) \geq \Pr(Y_{it}^{\overline{d}_{0t}} = 0|S_{i,t-1}) \Rightarrow \Pr(Y_{it}^{\overline{d}_{1t}} = 0|S_{i,t-1}, Y_{it}^{\overline{d}_{0t}} = 0) = 1 \tag{29}$$

and

$$\Pr(Y_{it}^{\overline{d}_{1t}} = 0|S_{i,t-1}) \leq \Pr(Y_{it}^{\overline{d}_{0t}} = 0|S_{i,t-1}) \Rightarrow \Pr(Y_{it}^{\overline{d}_{0t}} = 0|S_{i,t-1}, Y_{it}^{\overline{d}_{1t}} = 0) = 1 \tag{30}$$

Because the right-hand side of (29) is equivalent to $\Pr(Y_{it}^{\bar{d}_{1t}} = 1 | S_{i,t-1}, Y_{it}^{\bar{d}_{0t}} = 0) = 0$, it is also equivalent to $\Pr(Y_{it}^{\bar{d}_{1t}} = 1, Y_{it}^{\bar{d}_{0t}} = 0 | S_{i,t-1}) = 0$, which in turn is equivalent to $\Pr(Y_{it}^{\bar{d}_{1t}} \geq Y_{it}^{\bar{d}_{0t}} | S_{i,t-1}) = 0$.

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 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

$$\Pr(Y_2^{00} = 1 | Y_1^1 = 0, Y_1^0 = 0) \leq \Pr(Y_2^{00} = 1 | Y_1^1 = 1, Y_1^0 = 0). \qquad (31)$$

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

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

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

$$V_i - \varepsilon_{is} < -\alpha_s \qquad s = 1, \ldots, t-1$$

$$V_i - \tilde{\varepsilon}_{is} < -\alpha_s - \gamma_{is} \qquad s = 1, \ldots, t-1$$

and $Y_{is}^0 = 0, s = 1, \ldots, t-1$ and $Y_{is}^1 = 0, s = 1, \ldots, k-1$, $Y_{ik}^1 = 1$, (for $k = 1, \ldots, t-1$) i.e.

$$V_i - \varepsilon_{is} < -\alpha_s \qquad s = 1, \ldots, t-1$$

$$V_i - \tilde{\varepsilon}_{is} < -\alpha_s - \gamma_{is} \qquad s = 1, \ldots, k-1 \qquad V_i - \tilde{\varepsilon}_{ik} \geq -\alpha_k - \gamma_{ik}.$$

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

$$V_i - \varepsilon_{i1} < -\alpha_1$$

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

and

$$V_i - \varepsilon_{i1} < -\alpha_1$$

$$V_i - \tilde{\varepsilon}_{i1} \geq -\alpha_1 - \gamma_{i1}.$$

Hence if $V_i - \varepsilon_{i1}$ and $V_i - \tilde{\varepsilon}_{i1}$ are positively related with $V_i - \varepsilon_{i2}$ then (31) 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

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

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

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

Because 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 6 (Positively Correlated Outcomes (PCO)) *For all* $k = 1, \ldots, t-1$ *we have*

$$\begin{aligned}\Pr(Y_t^0 = 1 | Y_k^1 = 1, Y_{k-1}^1 = 0, \ldots, Y_1^1 = 0, Y_{t-1}^0 = 0, \ldots, Y_1^0 = 0) \geq \\ \Pr(Y_t^0 = 1 | Y_{t-1}^1 = 0, \ldots, Y_1^1 = 0, Y_{t-1}^0 = 0, \ldots, Y_1^0 = 0)\end{aligned} \tag{32}$$

and

$$\begin{aligned}\Pr(Y_t^1 = 1 | Y_{t-1}^1 = 0, \ldots, Y_1^1 = 0, Y_k^0 = 1, Y_{k-1}^0 = 0, \ldots, Y_0^1 = 0) \geq \\ \Pr(Y_t^1 = 1 | Y_{t-1}^1 = 0, \ldots, Y_1^1 = 0, Y_{t-1}^0 = 0, \ldots, Y_1^0 = 0).\end{aligned} \tag{33}$$

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 $ATES_2^{11,00}$ in the two-period case under the additional assumptions.

Theorem 3 (Bounds under MTR and CS for 2 periods) *Suppose Assumptions 3, 4, and 5 hold. Then if* $\Pr(Y_1 = 0 | D_1 = 1) \geq \Pr(Y_1 = 0 | D_1 = 0)$ *and* $\Pr(Y_1 = 0 | D_1 = 0) > 0$

$$\begin{aligned}\max\left\{1 - \frac{(1 - \Pr(Y_2 = 1 | D_1 = 1, Y_1 = 0)) \Pr(Y_1 = 0 | D_1 = 1)}{\Pr(Y_1 = 0 | D_1 = 0)}, 0\right\} - \\ \Pr(Y_2 = 1 | D_1 = 0, D_2 = 0, Y_1 = 0) \leq ATES_2^{11,00} \leq \\ \min\left\{\frac{\Pr(Y_2 = 1 | D_1 = 1, Y_1 = 0) \Pr(Y_1 = 0 | D_1 = 1)}{\Pr(Y_1 = 0 | D_1 = 0)}, 1\right\} - \\ \Pr(Y_2 = 1 | D_1 = 0, D_2 = 0, Y_1 = 0)\end{aligned} \tag{34}$$

and if $\Pr(Y_1 = 0 | D_1 = 1) < \Pr(Y_1 = 0 | D_1 = 0)$ *and* $\Pr(Y_1 = 0 | D_1 = 1) > 0$

$$\Pr(Y_2 = 1 | D_1 = 1, Y_1 = 0) - \tag{35}$$

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

$$\leq ATES_2^{11,00} \leq$$

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

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

Proof See Appendix. □

The same result for arbitrary t when we compare a treatment started in period 1 to no treatment in all periods is

Theorem 4 (Bounds under MTR and CS for t periods) *Suppose Assumptions 3, 4, and 5 hold. Then if*

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

$$\max\left\{\frac{\left[\Pr(Y_t=1|\overline{Y}_{t-1}=0,\overline{D}_t=1)-1\right]\Pr(\overline{Y}_{t-1}=0|\overline{D}_{t-1}=1)}{\min\left\{\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=1),\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=0)\right\}}+1,0\right\}- \tag{36}$$

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

$$\leq ATES_t^{1,0} \leq$$

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

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

Proof See Appendix. □

For PCO we also give separate results for $t=2$ and $t\geq 3$

Theorem 5 (Bounds under PCO for 2 periods) *If Assumptions 3 and 6 hold, then if* $\Pr(Y_1 = 0|D_1 = 1) + \Pr(Y_1 = 0|D_1 = 0) - 1 > 0$

$$\max\left\{\frac{[\Pr(Y_2 = 1|D_1 = 1, Y_1 = 0) - 1]\Pr(Y_1 = 0|D_1 = 1)}{\Pr(Y_1 = 0|D_1 = 1) + \Pr(Y_1 = 0|D_1 = 0) - 1} + 1, 0\right\} - \tag{37}$$

$$\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0) \leq ATES_2^{11,00} \leq$$

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

$$\max\left\{\frac{[\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0) - 1]\Pr(Y_1 = 0|D_1 = 0)}{\Pr(Y_1 = 0|D_1 = 1) + \Pr(Y_1 = 0|D_1 = 0) - 1} + 1, 0\right\}$$

Proof See Appendix. □

Theorem 6 (Bounds under PCO for t periods) *If Assumptions 3 and 6 hold, then if* $\Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 1) + \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 0) - 1 > 0$ *for all* $s = 1, \ldots, t-1$

$$\max\left\{\frac{(\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 1) - 1)\Pr\left(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1\right)}{\prod_{s=1}^{t-1}\left[\Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 1) + \Pr(Y_s = 0|\overline{Y}_{s-1} = 0, \overline{D}_s = 0) - 1\right]} + 1, 0\right\} - \tag{38}$$

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

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

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

Proof See Appendix. □

The bounds can be improved if MTR, CS and PCO hold simultaneously.

Theorem 7 (Bounds under MTR, CS and PCO for t periods) *Suppose Assumptions 3, 4, 5 and 6 hold. If for all t we have*

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

then

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

$$\Pr(Y_t=1|\overline{Y}_{t-1}=0,\overline{D}_t=0)$$
$$\leq ATES_t^{1,0} \tag{39}$$
$$\Pr(Y_t=1|\overline{Y}_{t-1}=0,\overline{D}_t=1)-$$

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

Proof See Appendix. □

5 Application to the Illinois bonus experiment

5.1 The re-employment bonus experiment

Between mid-1984 and mid-1985, the Illinois Department of Employment Security conducted a randomized social experiment.[4] 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.

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

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.
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.

5.2 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.[5] 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.

5.3 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 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.

$$\mathrm{ATES}_t^{1,0} > 0 \quad , \ t = 1, \ldots, 5$$

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

$$\mathrm{ATES}_t^{1,0} = 0 \quad , \ t = 6, \ldots, 12$$

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

$$\mathrm{ATES}_5^{1,0} > \mathrm{ATES}_4^{1,0}.$$

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

[5]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.

Table 1: Bounds on conditional transition probabilities for the Illinois job bonus experiment (claimant bonus)

	NO [1]		MTR+CS [2]	
	LB	UB	LB	UB
Week				
1-2	0.023	0.023	0.023	0.023
3-4	-0.097	0.111	0.011	0.038
5-6	-0.094	0.100	0.004	0.046
7-8	-0.101	0.121	0.013	0.063
9-10	-0.112	0.126	0.008	0.069
11-12	-0.121	0.139	0.008	0.062
13-14	-0.163	0.159	-0.002	0.056
15-16	-0.188	0.200	0.003	0.051
17-18	-0.301	0.302	0.000	0.049
19-20	-0.775	0.718	-0.004	0.049
21-22			-0.021	0.047
23-24			-0.002	0.056
	PCO [3]		MTR+CS+PCO [4]	
	LB	UB	LB	UB
1-2	0.023	0.023	0.023	0.023
3-4	-0.081	0.094	0.014	0.038
5-6	-0.067	0.074	0.007	0.046
7-8	-0.062	0.078	0.016	0.063
9-10	-0.057	0.069	0.012	0.069
11-12	-0.050	0.062	0.012	0.062
13-14	-0.053	0.056	0.003	0.056
15-16	-0.043	0.051	0.007	0.051
17-18	-0.045	0.049	0.005	0.049
19-20	-0.048	0.049	0.001	0.049
21-22	-0.061	0.047	-0.014	0.047
23-24	-0.053	0.056	0.003	0.056

Notes: NO are the bounds under random assignment, MTR those under for Monotone Treatment Response, CS under Common Shocks and PCO under Positively Correlated Outcomes.

Table 1 presents the upper and the lower bound on $ATES_t^{1,0}$ for the claimant group under just random assignment and under the additional assumptions in Section 4. Figure 1 displays the same bounds[6]. The general bounds, which impose no assumptions beyond random assignment are labeled NO. 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. 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, note that until week 20 the bounds are nevertheless informative on the average treatment effect.

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. PCO has the same effect. Imposing MTR, CS as well as PCO further tightens the bounds. If these assumptions hold simultaneously we can for the weeks up until week 20 rule out that the bonus offer has a negative effect on the transition rate out of unemployment.

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 the lower bound does not increase enough, so that both an increase and no change (end 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.

[6] We do not report confidence intervals. In principle these can be constructed using the approach in Hahn and Ridder (2010), but we leave the details to future work.

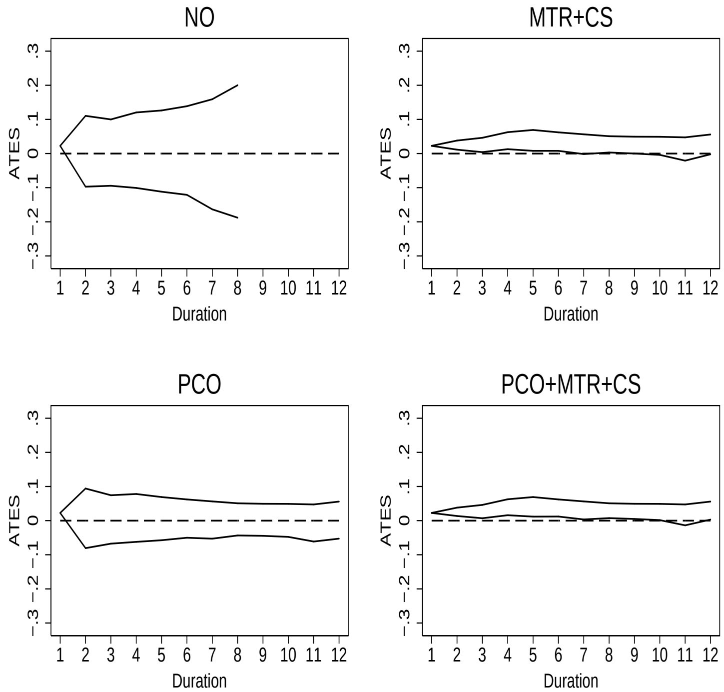

Figure 1: Bounds on conditional transition probabilities for the Illinois job bonus experiment (claimant bonus)

6 Conclusions

In this paper, we have derived bounds on treatment effects on conditional transitions 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 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 G. J. V. den Berg (2003): "The non-parametric identification of treatment effects in duration models," *Econometrica*, 71, 1491–1517.

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.

Bijwaard, G., and G. Ridder (2005): "Correcting for Selective Compliance in a Re-employment 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.

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.

Fredriksson, P., and P. Johansson (2008): "Dynamic Treatment Assignment: The Consequences for Evaluations Using Observational Data," *Journal of Business & Economic Statistics*, 26, 435–445.

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 (2010): "Partial identification and confidence intervals," Working Paper.

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. Gra, L. Jacobi, P. Jensen, R. Leetmaa, L. Nima, S. S. Patacchini, E., 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.

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.

Appendix: Proofs

Proof of Lemma 1

Because Assumption 3 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 (5)

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

For (10)

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

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

where the first equality follows from the observation rules, the second from (6), and the third from (5). Analogously for (11)

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

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

$\square$.

In the remainder of the Appendix we use the following notation

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

Proof of Theorem 1

Using the notation above (20) is equivalent to

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

$$\frac{\mathbb{E}(Y_2^{00}|Y_1^1=0,Y_1^0=0)p(0,0)+p^{00}(1|1,0)p(1,0)}{\Pr(Y_1=0|D_1=0)}$$

Observing that $\mathbb{E}(Y_2^{00}|Y_1^1=0,Y_1^0=0)=p^{00}(1|0,0)$ we find

$$p^{00}(1|0,0)=$$

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

From (18) $p(1,0)=\Pr(Y_1=0|D_1=0)-p(0,0)$ and upon substitution

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

$$\frac{p^{00}(1|1,0)\left[\Pr(Y_1=0|D_1=0)-p(0,0)\right]}{p(0,0)}.$$

The probability $p^{00}(1|0,0)$ depends on the unknown probabilities $p^{00}(1|1,0)$ and $p(0,0)$. Now note that, because $p(0,0)=\Pr(Y_1^0=0,Y_1^1=0)\leq\Pr(Y_1^0=0)=\Pr(Y_1=0|D_1=0)$, the function on the right hand side is decreasing in $p^{00}(1|1,0)$ for all $p(0,0)>0$. Therefore for all $p(0,0)>0$

$$\frac{\left[\Pr(Y_2=1|D_1=0,D_2=0,Y_1=0)-1\right]\Pr(Y_1=0|D_1=0)}{p(0,0)}+1 \quad \text{(A.2)}$$

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

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

where the lower bound applies if $p^{00}(1|1,0)=1$ and the upper bound if $p^{00}(1|1,0)=0$. 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 $\Pr(Y_1=0|D_1=1)+\Pr(Y_1=0|D_1=0)-1>0$ then the probability of survival for the treated and controls is relatively large. In that case the degree of dynamic selection is relatively small. The opposite is true is true if $\Pr(Y_1=0|D_1=1)+\Pr(Y_1=0|D_1=0))-1\leq 0$. Because in the latter case the lower bound on $p(0,0)$ is 0, $\mathbb{E}(Y_2^{00}|Y_1^1=0,Y_1^0=0)$ is not defined at this lower bound, because it is an average for a non-existent subpopulation. Therefore we only need to consider the case that $\Pr(Y_1=0|D_1=1)+\Pr(Y_1=0|D_1=0)-1>0$. Upon substitution of the lower bound on $p(0,0)$ in (A.2) we find

$$\max\left\{\frac{[\Pr(Y_2=1|D_1=0,D_2=0,Y_1=0)-1]\Pr(Y_1=0|D_1=0)}{\Pr(Y_1=0|D_1=1)+\Pr(Y_1=0|D_1=0))-1}+1,0\right\}\leq \tag{A.3}$$

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

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

Next we derive bounds on $\mathbb{E}(Y_2^{11}|Y_1^1=0,Y_1^0=0)=p^{11}(1|0,0)$. Equation (19) is equivalent to

$$\Pr(Y_2=1|D_1=1,Y_1=0)=\frac{\mathbb{E}[Y_2^{11}|Y_1^1=0,Y_1^0=0]p(0,0)+p^{11}(1|0,1)p(0,1)}{\Pr(Y_1=0|D_1=1)}$$

so that

$$p^{11}(1|0,0)=\frac{\Pr(Y_2=1|D_1=1,Y_1=0)\Pr(Y_1=0|D_1=1)-p^{11}(1|0,1)p(0,1)}{p(0,0)}.$$

By (17) $p(0,1)=\Pr(Y_1=0|D_1=1)-p(0,0)$ and upon substitution

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

$$\frac{p^{11}(1|0,1)\,[\Pr(Y_1=0|D_1=1)-p(0,0)]}{p(0,0)}.$$

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 right hand side is decreasing in $p^{11}(1|0,1)$ for all $p(0,0)>0$, so that we have the bounds

$$\frac{[\Pr(Y_2=1|D_1=1,Y_1=0)-1]\Pr(Y_1=0|D_1=1)}{p(0,0)}+1\leq \tag{A.5}$$

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

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

with the lower bound decreasing and the upper bound increasing in $p(0,0)$. We have the same lower bound on $p(0,0)$ and as before if this bound is strictly positive then $\mathbb{E}[Y_2^{11}|Y_1^1=0,Y_1^0=0]$ exists and is bounded by

$$\max\left\{\frac{[\Pr(Y_2=1|D_1=1,Y_1=0)-1]\Pr(Y_1=0|D_1=1)}{\Pr(Y_1=0|D_1=1)+\Pr(Y_1=0|D_1=0))-1}+1,0\right\}\leq \quad \text{(A.6)}$$

$$\mathbb{E}[Y_2^{11}|Y_1^1=0,Y_1^0=0]\leq$$

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

Finally, we combine the previous bounds to obtain bounds on $ATES_2^{11,00}$. From (A.1) and (A.4) we obtain a lower and upper bound on $ATES_2^{11,00}$ that depend on $p^{11}(1|0,1),p^{00}(1|1,0),p(0,0)$. The lower bound is decreasing in $p^{11}(1|0,1)$ and increasing in $p^{00}(1|1,0)$ for all $p(0,0)>0$ and the upper bound is increasing in $p^{11}(1|0,1)$ and decreasing in $p^{00}(1|1,0)$ for all $p(0,0)>0$. Setting $p^{11}(1|0,1)=1$ and $p^{00}(1|1,0)=0$ for the lower bound and $p^{11}(1|0,1)=0$ and $p^{00}(1|1,0)=1$ for the upper bound we obtain a lower bound that decreases in $p(0,0)$ and an upper bound that increases in $p(0,0)$. Hence, we can set $p(0,0)$ at is lower bound if that is positive and otherwise $ATES_2^{11,00}$ does not exist.

We now prove that the bounds are best possible, i.e. for each (lower or upper) bound we find the 14 parameters of the joint distribution of the potential outcomes $p(y_1^1,y_1^0),p(y_2^{01},y_2^{00}|1,0),p(y_2^{11}|0,1),p(y_2^{11},y_2^{01},y_2^{00}|0,0)$ such that the bound is binding and satisfy (17)-(21) (5 restrictions). First, consider the upper bound on $ATES_2^{11,00}$. This upper bound is 1 if and only if $\Pr(Y_2^{11}=1|Y_1^1=0,Y_1^0=0)=1$ and $\Pr(Y_2^{00}=1|Y_1^1=0,Y_1^0=0)=0$. This is equivalent to the following restrictions on the parameters

$$0=\Pr(Y_2^{00}=1|Y_1^1=0,Y_1^0=0)=\sum_{y_2^{11}=0}^{1}\sum_{y_2^{01}=0}^{1}p(y_2^{11},y_2^{01},1|0,0)\Leftrightarrow \quad \text{(A.7)}$$

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

$$1=\Pr(Y_2^{11}=1|Y_1^1=0,Y_1^0=0)=\sum_{y_2^{00}=0}^{1}\sum_{y_2^{01}=0}^{1}p(1,y_2^{01},y_2^{00}|0,0)\Leftrightarrow \quad \text{(A.8)}$$

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

The restrictions on the remaining parameters are

$$\Pr(Y_2 = 1|D_1 = 1, Y_1 = 0) = \tag{A.9}$$

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

and

$$\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0) = \tag{A.10}$$

$$\frac{\sum_{y_2^{01}=0}^{1} p(y_2^{01}, 1|1,0)p(1,0)}{\Pr(Y_1 = 0|D_1 = 0)}$$

and

$$\Pr(Y_2 = 1|D_1 = 0, D_2 = 1, Y_1 = 0) = \tag{A.11}$$

$$\frac{p(1,1,0|0,0)p(0,0) + \sum_{y_2^{00}=0}^{1} p(1, y_2^{00}|1,0)p(1,0)}{\Pr(Y_1 = 0|D_1 = 0)}.$$

and

$$\Pr(Y_1 = 0|D_1 = 1) = \sum_{y_1^0=0}^{1} p(0, y_1^0) \tag{A.12}$$

and

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

We substitute (A.12) and (A.13) into (A.9)-(A.11) to obtain

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

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

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

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

$$\Pr(Y_2 = 1|D_1 = 0, D_2 = 1, Y_1 = 0) = \tag{A.16}$$

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

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.15) the choice of $p(0,0)$ implies

$$p(0,1|1,0)+p(1,1|1,0) = \frac{\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0)\Pr(Y_1 = 0|D_1 = 0)}{1 - \Pr(Y_1 = 0|D_1 = 1)} \tag{A.17}$$

with the right hand side less than or equal to 1 if and only if the lower bound in (A.3) is 0. Next, by (A.14)

$$p(1|0,1) = \tag{A.18}$$

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

with the right hand side greater than or equal to 0 if and only if the upper bound in (A.6) is 1. Finally, (A.16) holds if

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

If we set

$$p(1,1|1,0) = \tag{A.20}$$

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

we have obtained a set of parameters that satisfies all the restrictions and gives an $ATES_2^{11,00}$ equal to 1 with for the remaining parameters

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

and if e.g.

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

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

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

and

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

The case that

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

is dealt with analogously.

The next step is to consider the case that the upper bound on $\Pr(Y_2^{11}=1|Y_1^1=0,Y_1^0=0)$ is less than 1 and the lower bound on $\Pr(Y_2^{00}=1|Y_1^1=0,Y_1^0=0)$ is greater than 0. After substitution we have the restrictions

$$\Pr(Y_2=1|D_1=1,Y_1=0)= \tag{A.21}$$
$$\frac{\Pr(Y_2^{11}=1|Y_1^1=0,Y_1^0=0)p(0,0)+p(1|0,1)\left(\Pr(Y_1=1|D_1=1)-p(0,0)\right)}{\Pr(Y_1=0|D_1=1)}$$

and

$$\Pr(Y_2=1|D_1=0,D_2=0,Y_1=0)= \tag{A.22}$$
$$\frac{\Pr(Y_2^{00}=1|Y_1^1=0,Y_1^0=0)p(0,0)}{\Pr(Y_1=0|D_1=0)}+$$
$$\frac{\sum_{y_2^{01}=0}^{1}p(y_2^{01},1|1,0)(\Pr(Y_1=0|D_1=0)-p(0,0))}{\Pr(Y_1=0|D_1=0)}$$

and

$$\Pr(Y_2=1|D_1=0,D_2=1,Y_1=0)= \tag{A.23}$$
$$\frac{\sum_{y_2^{11}=0}^{1}\sum_{y_2^{00}=0}^{1}p(y_2^{11},1,y_2^{00}|0,0)p(0,0)}{\Pr(Y_1=0|D_1=0)}+$$
$$\frac{\sum_{y_2^{00}=0}^{1}p(1,y_2^{00}|1,0)\left(\Pr(Y_1=0|D_1=0)-p(0,0)\right)}{\Pr(Y_1=0|D_1=0)}.$$

After substitution of the upper bound on $\Pr(Y_2^{11} = 1|Y_1^1 = 0, Y_1^0 = 0)$ and the lower bound on $\Pr(Y_2^{00} = 1|Y_1^1 = 0, Y_1^0 = 0)$ and rearranging we have the system

$$\Pr(Y_2 = 1|D_1 = 1, Y_1 = 0)\Pr(Y_1 = 0|D_1 = 1) = \tag{A.24}$$

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

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

$$\Pr(Y_2 = 1|D_1 = 0, D_2 = 0, Y_1 = 0)\Pr(Y_1 = 0|D_1 = 0) = \tag{A.25}$$

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

$$\sum_{y_2^{01}=0}^{1} p(y_2^{01}, 1|1,0)\left(\Pr(Y_1 = 0|D_1 = 0) - p(0,0)\right)$$

$$\Pr(Y_2 = 1|D_1 = 0, D_2 = 1, Y_1 = 0)\Pr(Y_1 = 0|D_1 = 0) = \tag{A.26}$$

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

$$\sum_{y_2^{00}=0}^{1} p(1, y_2^{00}|1,0)\left(\Pr(Y_1 = 0|D_1 = 0) - p(0,0)\right).$$

We find a solution with

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

Substitution in (A.24) gives $p(1|0,1) = 0$ and in (A.25) gives $p(0,1|1,0) + p(1,1|1,0) = 1$ so that $p(1,0|1,0) = p(0,0|1,0) = 0$. In the final equation (A.26) this gives the solution

$$p(1,1|1,0) = \sum_{y_2^{11}=0}^{1}\sum_{y_2^{00}=0}^{1} p(y_2^{11}, 1, y_2^{00}|0,0) = \Pr(Y_2 = 1|D_1 = 0, D_2 = 1, Y_1 = 0) \tag{A.27}$$

so that

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

This is a solution if also

$$p(0,0,1|0,0)+p(0,1,1|0,0)+p(1,0,1|0,0)+p(1,1,1|0,0) = \tag{A.28}$$
$$\frac{[\Pr(Y_2=1|D_1=0,D_2=0,Y_1=0)-1]\Pr(Y_1=0|D_1=0)}{\Pr(Y_1=0|D_1=1)+\Pr(Y_1=0|D_1=0)-1}+1 \geq 0$$

and

$$p(1,0,0|0,0)+p(1,1,0|0,0)+p(1,0,1|0,0)+p(1,1,1|0,0) = \tag{A.29}$$
$$\frac{\Pr(Y_2=1|D_1=1,Y_1=0)\Pr(Y_1=0|D_1=1)}{\Pr(Y_1=0|D_1=1)+\Pr(Y_1=0|D_1=0))1} \leq 1$$

A solution to (A.27)-(A.29) is

$$p(1,1,1|0,0) = \min\Bigg\{\Pr(Y_2=1|D_1=0,D_2=1,Y_1=0),$$
$$\frac{[\Pr(Y_2=1|D_1=0,D_2=0,Y_1=0)-1]\Pr(Y_1=0|D_1=0)}{\Pr(Y_1=0|D_1=1)+\Pr(Y_1=0|D_1=0)-1}+1,$$
$$\frac{\Pr(Y_2=1|D_1=1,Y_1=0)\Pr(Y_1=0|D_1=1)}{\Pr(Y_1=0|D_1=1)+\Pr(Y_1=0|D_1=0)-1}\Bigg\} \equiv C$$
$$p(0,1,1|0,0) =$$
$$\min\Bigg\{\frac{[\Pr(Y_2=1|D_1=0,D_2=0,Y_1=0)-1]\Pr(Y_1=0|D_1=0)}{\Pr(Y_1=0|D_1=1)+\Pr(Y_1=0|D_1=0)-1}+1,$$
$$\Pr(Y_2=1|D_1=0,D_2=1,Y_1=0)\Bigg\}-C$$
$$p(1,1,0|0,0) = \min\Bigg\{\frac{\Pr(Y_2=1|D_1=1,Y_1=0)\Pr(Y_1=0|D_1=1)}{\Pr(Y_1=0|D_1=1)+\Pr(Y_1=0|D_1=0)-1},$$
$$\Pr(Y_2=1|D_1=0,D_2=1,Y_1=0)\Bigg\}-C.$$

For values of the remaining parameters consider e.g. the case that

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

so that

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

The remaining parameters satisfy the equations

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

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

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

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

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

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

so that

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

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

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

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

If e.g.

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

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

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

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

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

Finally,

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

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

Besides these two cases we have to find the joint distributions of the potential outcomes for the cases that $\mathbb{E}(Y_2^{11}|Y_1^1 = 0, Y_1^0 = 0) < 1, \mathbb{E}(Y_2^{00}|Y_1^1 = 0, Y_1^0 = 0) = 0$ and $\mathbb{E}(Y_2^{11}|Y_1^1 = 0, Y_1^0 = 0) = 1, \mathbb{E}(Y_2^{00}|Y_1^1 = 0, Y_1^0 = 0) > 0$. 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.

Proof of Theorem 2

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

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

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

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

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

we have

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

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

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 $p_{t-1}(0,0)$. We have by the Bonferroni inequality

$$p_{t-1}(0,0) \geq$$
$$\max\left\{\Pr(Y^1_{t-1}=0,\ldots,Y^1_1=0)+\Pr(Y^0_{t-1}=0,\ldots,Y^0_1=0)-1,0\right\}.$$

Also with $Y_0 \equiv 0$

$$\Pr(Y^1_{t-1}=0,\ldots,Y^1_1=0)=\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=1)$$

and

$$\Pr(Y^0_{t-1}=0,\ldots,Y^0_1=0)=\prod_{s=1}^{t-1}\Pr(Y_s=0|\overline{Y}_{s-1}=0,\overline{D}_s=0)$$

so that

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

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 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 $\Pr(Y^0_t=1|Y^1_{t-1},\ldots,Y^1_1=0,Y^0_{t-1}=0,\ldots,Y^0_1=0)$. The bounds on $ATES^{1,0}_t$ follow directly.

Next, we prove that the bounds are sharp. We find the parameters of the joint distribution of the potential outcomes, such that a (lower or upper) bound is binding and such that the joint distribution of the potential outcomes satisfies the restrictions implied by the observed outcomes. The joint distribution of the potential outcomes up to period t is determined by the probabilities

$$p_s(y^1_s,y^0_s|0,0)\quad,p_{s-1}(0,0),\ s=1,\ldots,t$$
$$p^1_s(y^1_s|0,k)\ ,\quad p_{s-1}(0,k)\ ,\quad k=1,\ldots,s-1,\ s=1,\ldots,t$$

$$p_s^0(y_s^0|k,0)\ ,\quad p_{s-1}(k,0)\ ,\quad k=1,\ldots,s-1,\ s=1,\ldots,t$$

where $p_t^0(1|k,0) = \Pr(Y_t^0 = 1|Y_k^1 = 1,\ldots,Y_1^1 = 0, Y_{t-1}^0 = 0,\ldots,Y_1^0 = 0)$ and $p_t(k,0) = \Pr(Y_k^1 = 1,\ldots,Y_1^1 = 0, Y_t^0 = 0,\ldots,Y_1^0 = 0)$. The restrictions imposed by the observed outcomes are for $s = 1,\ldots,t$

$$\Pr(Y_s = 1|\overline{Y}_{s-1} = 0, \overline{D}_s = 1)\Pr(\overline{Y}_{s-1} = 0|\overline{D}_{s-1} = 1) = \tag{A.31}$$

$$\sum_{y_s^0=0}^{1} p_s(1, y_s^0|0,0)p_{s-1}(0,0) + \sum_{k=1}^{s-1} p_s^1(1|0,k)p_{s-1}(0,k)$$

$$\Pr(Y_s = 1|\overline{Y}_{s-1} = 0, \overline{D}_s = 0)\Pr(\overline{Y}_{s-1} = 0|\overline{D}_{s-1} = 0) = \tag{A.32}$$

$$\sum_{y_s^1=0}^{1} p_s(y_s^1, 1|0,0)p_{s-1}(0,0) + \sum_{k=1}^{s-1} p_s^0(1|k,0)p_{s-1}(k,0)$$

$$\Pr(\overline{Y}_{s-1} = 0|\overline{D}_{s-1} = 1) = p_{s-1}(0,0) + \sum_{k=1}^{s-1} p_{s-1}(0,k) \tag{A.33}$$

$$\Pr(\overline{Y}_{s-1} = 0|\overline{D}_{s-1} = 0) = p_{s-1}(0,0) + \sum_{k=1}^{s-1} p_{s-1}(k,0) \tag{A.34}$$

Note that we simplify the problem by only considering individuals who are never treated and individuals who are always treated. The bounds only depend on the outcomes for these subpopulations and the proof of sharpness for $t = 2$ shows that the restrictions derived from the individuals who enter treatment after period 1 do not bound parameters that enter into the definition of ATES. The common parameter in the restrictions is $p(0,0)$ that is set at its lower Bonferroni bound that only involves outcomes for the always and never treated.

By definition

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

$$\sum_{y_t^0=0}^{1} p_t(1, y_t^0|0,0) - \sum_{y_t^1=0}^{1} p_t(y_t^1, 1|0,0) = p_t^1(1|0,0) - p_t^0(1|0,0)$$

We need to consider the system

$$\Pr(Y_t = 1|\overline{Y}_{t-1} = 0, \overline{D}_t = 1)\Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1) = \tag{A.35}$$

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

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

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

$$\Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 1) = p_{t-1}(0,0) + \sum_{k=1}^{t-1} p_{t-1}(0,k) \tag{A.37}$$

$$\Pr(\overline{Y}_{t-1} = 0|\overline{D}_{t-1} = 0) = p_{t-1}(0,0) + \sum_{k=1}^{t-1} p_{t-1}(k,0) \tag{A.38}$$

We find the parameters consistent with the upper bound on $ATES_t^{1,0}$. The parameters are

$$p_t^1(1|0,k) = 0 \qquad p_t^0(1|k,0) = 1 \qquad k = 1, \dots, t-1$$

and

$$p_{t-1}(0,0) =$$

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

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

Substituting these values in (A.35) and (A.36) gives the upper bound on $p_t^1(1|0,0)$ and the lower bound on $p_t^0(1|0,0)$. If the upper bound is larger than 1 it is set at 1 and if the lower bound is less than 0 it is set at 0. The remaining parameters of the joint distribution of the counterfactuals satisfy

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

and

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

and

$$p_t^1(1|0,0) = \sum_{y_t^0=0}^{1} p_t(1, y_t^0|0,0) \quad p_t^0(1|0,0) = \sum_{y_t^1=0}^{1} p_t(y_t^1, 1|0,0)$$

and

$$\sum_{y_t^0=0}^{1} p_t(1, y_t^0|0,0) = 1.$$

These restrictions are easily satisfied for values between 0 and 1.

Parameter values consistent with the lower bound on $ATES_t^{1,0}$ can be found in the same way. Working back in time we obtain the joint distribution of the counterfactuals over time consistent with $ATES_t^{1,0}$ being at the upper or lower bound for all t. $\square$

Proof of Theorem 3

From equations (A.1) and (A.4) we obtain

$$ATES_2^{11,00} = \frac{\Pr(Y_2 = 1|D_1 = 1, Y_1 = 0)\Pr(Y_1 = 0|D_1 = 1)}{p(0,0)} -$$

$$\frac{p^{11}(1|0,1)\left[\Pr(Y_1 = 0|D_1 = 1) - p(0,0)\right]}{p(0,0)} -$$

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

$$\frac{p^{00}(1|1,0)\left[\Pr(Y_1 = 0|D_1 = 0) - p(0,0)\right]}{p(0,0)}.$$

The right-hand side is decreasing in $p^{11}(1|0,1)$ and increasing in $p^{00}(1|1,0)$, so that

$$\max\left\{\frac{\Pr(Y_2 = 1|D_1 = 1, Y_1 = 0)\Pr(Y_1 = 0|D_1 = 1)}{p(0,0)} - \frac{\Pr(Y_1 = 0|D_1 = 1) - p(0,0)}{p(0,0)}, 0\right\} \tag{A.39}$$

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

$$\leq ATES_2^{11,00} \leq$$

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

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

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

The lower bound is increasing and the upper bound decreasing in $p(0,0)$.

If

$$\Pr(Y_1 = 0|D_1 = 1) = \Pr(Y_1^1 = 0) \geq \Pr(Y_1^0 = 0) = \Pr(Y_1 = 0|D_1 = 0)$$

then by Assumption 4 for all i

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

and by Assumption 5

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

so that

$$\Pr(Y_1^1 \leq Y_1^0) = 1.$$

Therefore

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

Substitution gives the bounds. If

$$\Pr(Y_1 = 0|D_1 = 1) = \Pr(Y_1^1 = 0) \leq \Pr(Y_1^0 = 0) = \Pr(Y_1 = 0|D_1 = 0)$$

then by Assumptions 4 and 5

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

and this gives the other bounds.

Proof of Theorem 4

By the proof of Theorem 2 we have

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

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

$$ATES_t^{1,0}\leq$$

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

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

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

$$p_{t-1}(0,0)=\Pr(Y^1_{t-1}=0,\dots,Y^1_1=0,Y^0_{t-1}=0,\dots,Y^0_1=0)=$$

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

By Assumption 4 either

$$\Pr\left(Y^1_{i,t-1}=0\middle|\,S_{i,t-2}\right)\leq\Pr\left(Y^0_{i,t-1}=0\middle|\,S_{i,t-2}\right)\tag{A.40}$$

or

$$\Pr\left(Y^1_{i,t-1}=0\middle|\,S_{i,t-2}\right)>\Pr\left(Y^0_{i,t-1}=0\middle|\,S_{i,t-2}\right)\tag{A.41}$$

for all i. Assume that (A.40) holds. By Assumption 5 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 4 and 5 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 4 holds with (A.41), then

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

We conclude that

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

and substitution gives the bounds. $\Box$

Proof of Theorem 5

First, (A.1) expresses $p^{00}(1|0,0)$ as a decreasing function of $p^{00}(1|1,0)$. Therefore the lower bound on $p^{00}(1|0,0)$ is obtained if $p^{00}(1|1,0)=1$ and by Assumption 6 the upper bound is obtained if $p^{00}(1|0,1)=p^{00}(1|0,0)$. For the upper bound

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

and this is equivalent to

$$p^{00}(1|0,0)\leq\Pr(Y_2=1|D_1=0,D_2=0,Y_1=0).\tag{A.42}$$

An analogous argument gives

$$p^{11}(1|0,0) \leq \Pr(Y_2 = 1|D_1 = 1, Y_1 = 0). \tag{A.43}$$

Combining (A.42) and (A.43) with the lower bounds in Theorem 1 we obtain the bounds on the ATES $\square$.

Proof of Theorem 6

By (A.30) and Assumption 6 we obtain the following inequality

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

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

and upon rearranging this is equivalent to

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

An analogous argument gives

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

Assumption 6 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

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

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

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 $1\le s\le t-1$ 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 $s=1,\dots,t-1$

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

then

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

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

Proof of Theorem 7

By the proof of Theorem 6 we have under Assumption 6

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

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

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

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

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. By the proof of Theorem 4 we have under Assumptions 4 and 5

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

Substitution gives the bounds.

Since

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

we have

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

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

and analogously

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

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

so that the lower bound on $p_{t-1}(0,0)$ under MTR, CS and PCO is is not smaller than that under PCO, so that bounds under all three assumptions are narrower than under just PCO $\Box$.

Publication series published by the Institute for Labour Market Policy Evaluation (IFAU) – latest issues

Rapporter/Reports

2011:1 Hall Caroline and Linus Liljeberg ”En jobbgaranti för ungdomar? Om Arbetsförmedlingens ungdomsinsatser”

2011:2 Angelov Nikolay, Per Johansson, Erika Lindahl and Elly-Ann Lindström ”Kvinnors och mäns sjukskrivningar”

2011:3 Eliason Marcus ”Undersköterskor och sjukvårdsbiträden i kristider: inkomst- och sysselsättningseffekter av friställningar inom den offentliga sektorn under 1990-talet”

2011:4 Brandén Maria and Sara Ström ”För vems skull flyttar par? Kön, karriärmöjligheter och pars regionala rörlighet i Sverige”

2011:5 Sjögren Anna and Helena Svaleryd ”Nitlott i barndomen – familjebakgrund, hälsa, utbildning och socialbidragstagande bland unga vuxna”

Working papers

2011:1 Eliason Marcus “Assistant and auxiliary nurses in crisis times: earnings and employment following public sector job loss in the 1990s”

2011:2 Forslund Anders, Peter Fredriksson and Johan Vikström “What active labor market policy works in a recession?”

2011:3 Brandén Maria and Sara Ström “For whose sake do couples relocate?: Gender, career opportunities and couples’ internal migration in Sweden”

2011:4 Bergemann Annette, Marco Caliendo, Gerard J. van den Berg and Klaus F. Zimmermann “The threat effect of participation in active labor market programs on job search behaviour of migrants in Germany”

2011:5 van den Berg Gerard J., Petter Lundborg, Paul Nystedt and Dan-Olof Rooth “Critical periods during childhood and adolescence: a study of adult height among immigrant silblings”

2011:6 Ridder Geert and Johan Vikström “Bounds on treatment effects on transitions”

2011:7 Vikström Johan, Michael Rosholm and Michael Svarer “The relative efficiency of active labour market policies: evidence from a social experiment and non-parametric methods”

Dissertation series

2010:1 Johansson Elly-Ann “Essays on schooling, gender, and parental leave”

2010:2 Hall Caroline “Empirical essays on education and social insurance policies”