Nonparametric Instrumental Variable Methods for Treatment Evaluation among Survivors

Gerard J. van den Berg *
Petyo Bonev †
Enno Mammen ‡

Mannheim, February 2011

Abstract

Dynamic selection and endogeneous noncompliance hamper the evaluation of treatment effects when the outcome of interest is a duration variable. Existing methods either restrict their analysis to settings where only one of those two problems exists, or adopt parametric or semi-parametric structure. In this paper we develop two completely nonparametric Instrumental Variable approaches for duration data which enable us to identify treatment effects in the presence of both dynamic selection and endogeneous noncompliance. We suggest corresponding estimators. Our approaches are revealed to have as special cases numerous existing models. We suggest simple procedures to test for endogeneity.

---

*University of Mannheim, VU University Amsterdam, IFAU-Uppsala, IZA, CERP, and IFS. Adress: Department of Economics, University of Mannheim, L7,3-5, 68131 Mannheim, Germany.
†University of Mannheim.
‡University of Mannheim. Adress: Department of Economics, University of Mannheim, L7,3-5, 68131 Mannheim, Germany.
Jel codes: C14, C41.
Acknowledgements: We thank Bettina Drepper, Christoph Breunig, Andreas Landmann and Robert Poppe for useful comments.

1 Introduction

Duration models play an important role for the evaluation of treatment effects in a dynamic context (see e.g. [Van den Berg, 2001] and [Abbring, 2003] for an overview). A canonical example from labor economics is the assessment of the impact of job training on the duration of subsequent employment and unemployment spells (see [Heckman et al.,] for a study). Often the identification of treatment effects in dynamic settings is hampered by two major problems. First, as the studies of [Meyer, 1996], [Ham and LaLonde, 1996] and [Eberwein et al., 1997] reveal, dynamic selection contaminates the evaluation of outcomes of survivors at a point in time even if the treatment was initially randomized. Second, if the assigned treatment status does not have to be equal to the actual treatment received, then selection into treatment arms might additionally bias comparisons between groups with different actual treatment statuses. Both problems arise when unobservable heterogeneity has an impact on the distribution of the duration variable. A simple comparison of the outcomes in the various treatment groups is likely to capture not only the treatment effect but also these differences in the distributions of the unobserved heterogeneity. We will refer to these problems as to dynamic and static endogeneity respectively.

In this paper we develop two very general instrumental variable (IV) approaches for the identification and estimation of treatment effects in duration data which can handle both dynamic and static endogeneity and also allow for censoring. We do not adopt parametric or semi-parametric structure and do not impose independence of observed and unobserved characteristics.

In our first model we demonstrate identification, estimation and testing of average causal treatment effects on the conditional survival function in a regression discontinuity setting. A single comprehensive treatment is assigned to start on a specific day and noncompliance is possible. A well studied example is the welfare-to-work program New Deal for the Young People in the UK, which was released nationwide on April 1, 1998. Cohorts of individuals in the state of interest, here unemployment, receive the treatment at a different elapsed duration in this state. In this setting the distribution of the unobserved heterogeneity changes over time, which causes dynamic endogeneity. Moreover, the possibility to select out of treatment (even though perhaps costly) possibly induces static endogeneity. We use the time variation in the elapsed duration at the moment when treatment starts as an instrument (which can also be interpreted as an intention to treat (ITT)) to achieve identification. We compare the outcomes of all survived compliers who are assigned to receive the treatment at an elapsed duration of say t time units with the outcomes of the survived compliers from an older cohort at the same elapsed duration (and hence still without the treatment). We show that dynamic selection for both cohorts proceeds in the same way given randomization at the beginning, an idea first used in [Van den Berg et al., 2010]. Therefore the difference in observed outcomes should be due only to the treatment. A major difficulty in this setting, however, is that one cannot identify the compliers from the older cohort at t because they reveal their preferences at a later point in time, provided they survive that long. Thus

some compliers will never be identified. We solve this problem by using in an elegant way the information of the younger cohort on individuals who selected themselves out of the treatment. We show that with this information we can split the observed conditional distribution of the older cohort at t into the distributions of compliers and noncompliers. Thus, we are able in the presence of unobserved heterogeneity and without any (semi-)parametric assumptions to solve the problems of dynamic and static selection and identify an average treatment effect on the conditional distribution of the duration variable. This is a novel result. We show that if the set of all noncompliers is a null set then our model reduces to the model of [Van den Berg et al., 2010] as a special case. We also show that our model applies to a setting where the point in time of inflow is common for all individuals but the assignment of treatment is random over time and may differ across individuals. Similar results are obtained for treatment effects on the hazard of the duration variable.

In our second approach we abandon the regression discontinuity design and adopt a setting, in which randomization of the treatment assignment occurs at time 0, i. e. the moment of inflow. Agents are subject to treatment at some later common for everybody (random or deterministic) point in time. Unlike in our first model, we allow assignment to the control group as well, and noncompliance is possible in both treatment and control group. Thus we set a very general framework where both dynamic and static endogeneity arise. Using noncompliance information and the assigned treatment status as an instrument (ITT), we demonstrate identification of average treatment effects on the conditional survival function of the duration variable. This is an important contribution to the existing literature. If we restrict the actual treatment to occur at time 0, our model is revealed to reduce to the second model of [Abbring and van den Berg, 2005] as a special case. If we additionally rule out noncompliance, then we obtain as a special case the first model of [Abbring and van den Berg, 2005]. It can be easily shown that our results maintain their validity if we allow randomization of the assignment to treatment to occur simultaneously with the actual treatment at some later point in time.

Thus our two IV approaches are shown 1) to cover an extensive taxonomy of existing models in the literature which deal with **either** dynamic **or** static endogeneity in the context of duration data and 2) moreover, unlike existing methods, to handle **both** endogeneity types **simultaneously**. All results are attained through intuitive, nonrestrictive assumptions and without imposing any parametric or semi-parametric structure. Corresponding estimators have a natural interpretation and are closely related to the Wald-type statistics. Our IV methods can be therefore easily used in applied work.

An interesting application arises in both models. Comparing noncompliers from the younger cohort to a whole older cohort without the treatment in the first model, as well as comparing noncompliers in both treatment groups with observed outcomes in the second model, enables us to develop simple tests for static endogeneity. These tests have important applications, e.g. in the evaluation of pilot projects, where it might be of interest whether noncompliance is related to potential outcomes.

This paper is closely related to different branches of literature. First, literature on regression discontinuity design is connected to our first model. Some recent develop-

ments in this field were made by [Hahn et al., 2001], [Porter, 2003] and [Frlich, 2007]. Their approaches do not allow for censoring and are therefore of restricted applicability in modelling duration data. [Van den Berg et al., 2010] demonstrate nonparametric identification of treatment effects in a setting with policy discontinuity. Their approach allows for censoring and dynamic selection. As a major difference to our work, however, they assume that the actual treatment is exogeneous, which precludes static endogeneity. We allow for an endogeneous treatment. Their paper is revealed to be a special case of our first model.

A second branch of literature related to our paper is the literature on Instrumental Variable analysis. IV methods in nonparametric settings have been developed by [Imbens and Angrist, 1994], [Imbens and Rubin, 1997] and [Abadie, 2002]. They cannot handle censoring and do not deal with dynamic endogeneity. IV methods for duration data are considered in [Robins and Tsiatis, 1991], [Chesher, 2003], [Bijwaard and Ridder, 2005], [Bijwaard, 2006] and [Abbring and van den Berg, 2005]. They all can handle censoring. The methods of [Robins and Tsiatis, 1991], [Chesher, 2003], [Bijwaard and Ridder, 2005] and [Bijwaard, 2006], however, adopt semi-parametric or even parametric structure. In their first two settings, which are revealed to be special cases of our second model, [Abbring and van den Berg, 2005] do not postulate (semi-)parametric assumptions, but their analysis applies only to unconditional survival functions[1], thus precluding dynamic selection. In their last setting they show identification in the presence of dynamic endogeneity, however they adopt semi-parametric structure. We allow for dynamic endogeneity and do not impose (semi-)parametric structure on the distribution of the outcome variables.

The remainder of this paper is organized as follows. In section 2 we introduce our first model and show identification and estimation of treatment effects. In section 3 we discuss our second model. Section 4 concludes.

2 Instrumental variable approach in duration models with policy discontinuity

2.1 Notation and treatment effects

Suppose we observe a set of individuals (or agents, or simply objects) in a state of interest, for example hospital patients in a state of illness. Let T be a real nonnegative Borel-measurable function which represents the time of staying in the condition of interest. We call an object, whose elapsed time since inflow is k time units, a k-years-old object, and k itself is its age. Its inflow will be called birth, its outflow - dead. The point in time of inflow will be called its birthday, and of outflow simply day of death.

[1] Or to hazards at t when t converges to zero.

With τ^* we denote a calendar point in time, at which all objects in the state of interest are **supposed** to be exposed to a permanent change in some binary external conditions, and this change will be called treatment (a separate variable for these change or these conditions is not necessary, we will define its impact implicitly). We refer to τ^* as to (global) treatment day. A treatment is possible only on that day. Let the discrete random variable Z_i represent the elapsed time from point in time of birth of an object i till the treatment day τ^*. One object can be born on countable many points in time. All individuals who are born z years before the treatment day are called a z-cohort. Suppose objects can choose whether they get treated on τ^* or not. Let the random variable S_i stands for the actual time from birth until treatment of an individual i. $S_i = s$ represents a treatment after s time units, $S_i = \infty$ represents the case of no treatment. S_i will be also discrete since it depends on the point in time of inflow (for simplicity we refer to S as to th treatment itself). Let $s, z \in \mathbb{R}_+$. Following the standard potential outcome notation we denote $T(s,z)$ as the duration of "life" (= stay in the condition of interest) that would occur if an individual i would belong to the cohort z and get the treatment s. We will assume throughout this paper that an exclusion restriction holds: $T(s,z) = T(s)$, meaning that Z does not directly impact the potential duration. Let $S_i(z)$ represent the potential choice of time from birth to treatment, given that the individual belongs to the z-cohort. Stated later as an assumption, at each point in time t we will allow only two types of individuals according to their treatment preferences: 1) $S_i(t) = t$: compliers, who would accept the treatment they are given t, and 2) $S_i(t) = +\infty$: noncompliers, who would choose to be never treated, provided they were asked at t. This assumption about $S(t)$ is compatible with the case, where individuals know that they can be treated only once, at treatment day.
Further, let all factors that have an influence on the potential duration $T(s)$ be summarized in the random variables (or vectors) X and V, where X is observed and V not, and both are time-constant. One can think about V as about the error term in a linear regression model. We assume that for a concrete value of X and V $T(s)$ is a continuous random variable, reflecting some intrinsic randomness. Further, let C be the random censoring time, which we assume to be independent of T and S. We define $\widetilde{T}_i$ to be equal to C_i if $T_i > C_i$ and to T_i if $T_i \leq C_i$. Analogously, $\widetilde{S} = \min T, C, S$. We observe the sample consisting of the independent observations

$$(\tilde{T}_1, S_1, Z_1, X_1, \mathbb{I}_1), \ldots, (\tilde{T}_n, S_n, Z_n, X_n, \mathbb{I}_n),$$

drawn from the distribution of $(\tilde{T}, S, Z, X, \mathbb{I}_{T \leq C})$, where $\mathbb{I}$ indicates the order of T, C and S.
Our research question is: **what is the effect of the treatment on the duration variable when getting the treatment at age** t**,** $(t \in \mathbb{R}_+)$**?** Denote with $\theta_{T(s)}$ the hazard of $T(s)$. One possible way to define a treatment effect for this research question is

$$\theta_{T(s)}(t \mid X, V) - \theta_{T(s')}(t \mid X, V), \tag{2.1}$$

which is the individual additive treatment effect on the hazard of $T(s)$ at t, changing the treatment from s' to s. Two adjustments are necessary. First, we will be never able to

learn from the data about the effect of the treatment on the noncompliers. Therefore it makes sense to consider only compliers. With this first adjustment the treatment effect would be

$$\theta_{T(s)}(t \mid X, V, S(s) = s) - \theta_{T(s')}(t \mid X, V, S(s') = s'). \tag{2.2}$$

Second, since do not observe V, we would like to "average it out" and concentrate on an average treatment effect. The question arises over which population we have to average 2.2. [Abbring and van den Berg, 2005] argue, that if V influences $T(s)$, than its distribution could be different at different points in time. Since the hazard of $T(s)$ at t conditions on $T(s) \geq t$, we have to average over all complying survivors at t. Now we can define the Average Treatment Effect on the Treated Complying Survivors (ATETCS), on the Nontreated Complying Survivors (ATENTCS), and on the Treated Survivors (ATECS), respectively, as

$$\begin{aligned}
ATETCS(s,s',t) \coloneqq \quad & \mathbb{E}\Big[\theta_{T(s)}(t \mid X, V, S(t) = t) - \theta_{T(s')}(t \mid X, V, S(t) = t) \mid T(s) \geq t, X, S(t) = t\Big] \\
ATENTCS(s,s',t) \coloneqq \quad & \mathbb{E}\Big[\theta_{T(s)}(t \mid X, V, S(t) = t) - \theta_{T(s')}(t \mid X, V, S(t) = t) \mid T(s') \geq t, X, S(t) = t\Big] \\
ATECS(s,s',t) \coloneqq \quad & \mathbb{E}\Big[\theta_{T(s)}(t \mid X, V, S(t) = t) - \\
- \quad & \theta_{T(s')}(t \mid X, V, S(t) = t) \mid T(s) \geq t, T(s') \geq t, X, S(t) = t\Big].
\end{aligned}$$

More broadly, we define the General Average Treatment Effect on the Treated Complying Survivors (GATETCS) as

$$\begin{aligned}
GATETCS(s,s',t) \coloneqq \quad & \mathbb{E}\Big[P(T(s) \in B \mid T(s) \geq t, X, V, S(t) = t) - \\
- \quad & P(T(s') \in B \mid T(s') \geq t, X, V, S(t) = t) \mid T(s) \geq t, X, S(t) = t\Big],
\end{aligned}$$

where B is defined as $B \coloneqq [t, t+a)$ with $a \geq 0$ and $t + a < t'$. Thus, B is an interval of time, where the t cohort has received the treatment but the t'-cohort still not. Similarly we define $GATENTCS(s,s',t)$ and $GATECS(s,s',t)$. We will show that under some nonrestrictive and intuitive assumptions all of these treatment effects are identified at (t,t',t).

2.2 Identification of treatment effects without censoring

We will first concentrate on $GATETCS(t,t',t)$. We postulate the following assumptions:

A1 **(Single treatment knowledge)** : for any t and all i it holds either $S_i(t) = t$ or $S_i(t) = +\infty$.

A2 **(No anticipation)** : For each $t' \geq t \geq 0$ and each X, V holds

$$\Theta_{T(t')}(t \mid X, V, S(t) = t) = \Theta_{T(\infty)}(t \mid X, V, S(t) = t),$$

where $\Theta_{T(s)}$ is the integrated hazard of $T(s)$. We also assume no anticipation conditioning on the never-takers.

A3 **(Randomization)** : For the instrument Z it holds

$$i) \quad Z \perp\!\!\!\perp \{T(s), S(t)\} \mid X, V \qquad \text{and} \qquad ii) \quad Z \perp\!\!\!\perp V \mid X.$$

A4 : For all t and s $0 < P(S(t) = t \mid T(s) \geq t) < 1$

A5 **(Consistency restrictions)** For all $t, s \in \mathbb{R}_+ \bigcup \{+\infty\}$

i) $Z = t \Rightarrow S(t) = S$

ii) $S = s \Rightarrow T(s) = T$

Furthermore, in line with A4, we assume that all further expressions exist and are well defined. This will imply, among other things, that $0 < P(Z = t)$. We also restrict for now our attention on the case, where T is observed (= no censoring). The generalization is done in subsection 2.5. Before we proceed to our identification strategy, we first explain intuitively the assumptions.

A1 Assumption 1 means that at a concrete point in time an individual is either a complier or a nevertaker. It excludes the type of preferences $S(t) = t'$ for some $t' \neq t, t' < \infty$. This is compatible with the case, where individuals are informed that they can take the treatment only at treatment day, and hence choose between taking the treatment at treatment day or never. This information might be known before the treatment or it might be told to the individuals on the treatment day.

A2 No anticipation means that individuals would always act until treatment in a way, as if they wouldn't know at which age their treatment is assigned. This is in line with the case, at which agents come to know the assigned day of their treatment directly on the treatment day. This will imply, as shown later, that individuals from different cohorts but at the same age, who are still not treated, "behave" in the same way. Note that we only require this for compliers, since in a world, where individuals are free to take or not to take the treatment at treatment day, this would be an empty assumption for all individuals who choose not to take it.

A3 Randomization i) postulates that when controlled for X, V, the distributions of $T(s)$ and $S(z)$ are the same for all values of Z. ii) postulates the same for V. This would be compatible with the case, where individuals are randomly assigned to the cohorts.

A5 This assumption connects potential outcomes with observed variables (for a detailed discussion see [Pearl, 2000]). i) for example states that if the actual value of Z happens to be t, then $S(t)$ and S will have the same values.

A further preliminary note should help understand the expressions below. One can read $P(T(s) > k \mid X, S(t) = t, Z = z)$ as the probability for a complier ($S(t) = t$) from a cohort z ($Z = z$) with observed characteristics X to live longer than k years, had he have taken the treatment at the age of s. Thus we can interpret $GATETCS(t, t', t)$ as the treatment effect on the compliers if they would have taken the treatment at the age of t (if $t' > t$ lies outside B than taking the treatment at t' corresponds to the no treatment case).

We are going to use the time variation to treatment as an instrument to identify the treatment effect. To motivate our identification strategy, consider first the following expressions, which at first glance appear to be simple and intuitive candidates for a treatment effect:

$$P(T \in B \mid T \geq t, X, S = t, Z = t) - P(T \in B \mid T \geq t, X, S = \infty, Z = t) \tag{2.3}$$

$$P(T \in B \mid T \geq t, X, S = t, Z = t) - P(T \in B \mid T \geq t, X, S = t, Z = t') \tag{2.4}$$

$$P(T \in B \mid T \geq t, X, S = t, Z = t) - P(T \in B \mid T \geq t, X, S = t', Z = t'). \tag{2.5}$$

2.3 compares outcomes of the compliers and never-takers from the same cohort. Writing 2.3 in the form

$$\begin{aligned} & \mathbb{E}\Big[P(T \in B \mid T \geq t, X, S = t, Z = t, V) \mid T \geq t, X, S = t, Z = t\Big] - \\ - \quad & \mathbb{E}\Big[P(T \in B \mid T \geq t, X, S = \infty, Z = t, V) \mid T \geq t, X, S = \infty, Z = t\Big] \end{aligned}$$

reveals that we compare averages over two different populations: the survived compliers from cohort t, $\{i : T_i \geq t, X_i, S_i = t, Z_i = t\}$, and survived never-takers from cohort t, $\{i : T_i \geq t, X_i, S_i = \infty, Z_i = t\}$. These two populations will probably have different distributions of V. The reason for this is that the treatment variable S is a choice variableand it will most likely depend on some unobserved characteristics and hence on V. As a consequence it would hold $V \not\perp\!\!\!\perp S \mid T \geq t, X, Z = t$ (this we will refer to as static endogeneity, or static selection). Therefore part of the difference captured by 2.3 will be due to this difference in the distributions of V and not entirely to the treatment effect. Hence we should not use the never-takers as a control group. The question arises where to take a control group from. One approach would be to compare the conditional distributions of T at t on the survived compliers from two different cohorts, as 2.5 and 2.4 do. 2.5 is not defined since $P(S = t, Z = t') = 0$. In 2.4 we would encounter the following problem: for a cohort $Z = t'$ we can identify the compliers not until age of t', on the treatment day, where they reveal their preferences. For reasons mentioned in the notation chapter untreated survivor populations of age t and t' will most likely have different distributions of V (this we refer to as dynamic selection). Therefore again 2.4 will capture some of the difference of the distributions of V and hence not reflect the correct treatment effect. Previous studies in the literature of regression design either preclude static endogeneity,

see e.g. [Van den Berg et al., 2010], or cannot handle censoring, as in [Hahn et al., 2001] or [Frlich, 2007]. Our identification strategy has to account for both static and dynamic selection and is thus of a more complex nature. A simple interpretation of it is that we use Z as an instrument for the endogeneous S. As a treatment group we use the treated compliers of the t-cohort. To define a correct control group we show in a first step that if we take the t-years-old compliers from an older cohort t' (note that we cannot observe who are these compliers), they would have the same distribution of V as the t-years-old compliers from the t-cohort. In other words, the distribution of V among nontreated compliers develops over time in each cohort in the same way. This idea was first used in [Van den Berg et al., 2010] in the case of an exogeneous treatment and we have to adopt it for an endogeneous one. This is done in the following

Proposition 2.1. *Let F be a cdf. Under Assumptions A2 to A5 it holds for all $\infty \geq t' \geq t \geq 0$*

$$F_{V|T(t)\geq t,X,S(t)=t} = F_{V|T(t')\geq t,X,S(t)=t} = F_{V|T\geq t,X,S=t,Z=t}.$$

The proof is provided in the appendix. Its first consequence is the following result:

$$\begin{aligned}
& GATETCS(t,t',t) = \mathbb{E}\Big[P(T(t)\in B \mid T(t)\geq t, X, V, S(t)=t) - \\
- \quad & P(T(t')\in B \mid T(t')\geq t, X, V, S(t)=t) \mid T(t)\geq t, X, S(t)=t\Big] = \\
= \quad & \mathbb{E}\Big[P(T(t)\in B \mid T(t)\geq t, X, V, S(t)=t) \mid T(t)\geq t, X, S(t)=t\Big] - \\
- \quad & \mathbb{E}\Big[P(T(t')\in B \mid T(t')\geq t, X, V, S(t)=t) \mid T(t)\geq t, X, S(t)=t\Big] = \\
\overset{\text{Proposition 2.1}}{=} \quad & \mathbb{E}\Big[P(T(t)\in B \mid T(t)\geq t, X, V, S(t)=t) \mid T(t)\geq t, X, S(t)=t\Big] - \\
- \quad & \mathbb{E}\Big[P(T(t')\in B \mid T(t')\geq t, X, V, S(t)=t) \mid T(t')\geq t, X, S(t)=t\Big] = \\
= \quad & P(T(t)\in B \mid T(t)\geq t, X, S(t)=t) - P(T(t')\in B \mid T(t')\geq t, X, S(t)=t) = \\
=: \quad & p_1^* - p_2^*.
\end{aligned} \tag{2.6}$$

As we show in proposition 2.2, it is easy to show that $p_1^* = P(T \in B \mid T \geq t, X, S = t, Z = t)$. The identification of p_2^* is more complex, since we do not observe at age t who is a complier from the older t'-cohort. But we can estimate the ("average") conditional probability distribution of T at t of the t-years-old noncompliers of the t-cohort (notice that they are not treated at treatment day τ^*). We show that this distribution is equal to the distribution of the t-years-old noncompliers of the t'-cohort (which are also untreated). We also show that the proportions of compliers and noncompliers at age t are the same in both cohorts. Using this and also the fact, that at age t we can estimate the ("average") conditional distribution of T for the **whole** t'-cohort (compliers and noncompliers together), we can then split it using the known proportions into compliers and noncompliers. To explain this in a better and more formal way, consider the following equalities

:

$$\begin{aligned} & P(T \in B \mid T \geq t, X, Z = t') = P(T \in B, \text{Complier} \mid T \geq t, X, Z = t') + \\ + \quad & P(T \in B, \text{Noncomplier} \mid T \geq t, X, Z = t') \\ = \quad & P(T \in B \mid T \geq t, X, Z = t', \text{Complier}) P(\text{Complier} \mid T \geq t, X, Z = t') + \\ + \quad & P(T \in B \mid T \geq t, X, Z = t', \text{Noncomplier}) P(\text{Noncomplier} \mid T \geq t, X, Z = t'). \end{aligned}$$

We observe $P(T \in B \mid T \geq t, X, Z = t')$ and we show that

$$\begin{aligned} P(\text{Complier} \mid T \geq t, X, Z = t') &= P(\text{Complier} \mid T \geq t, X, Z = t) \\ P(\text{Noncomplier} \mid T \geq t, X, Z = t') &= P(\text{Noncomplier} \mid T \geq t, X, Z = t), \end{aligned}$$

where the right-hand side of these two equalities can be identified: we show that it holds

$$\begin{aligned} P(\text{Complier} \mid T \geq t, X, Z = t) &= P(S = t \mid T \geq t, X, Z = t) \\ P(\text{Noncomplier} \mid T \geq t, X, Z = t) &= P(S = \infty \mid T \geq t, X, Z = t). \end{aligned}$$

With these equalities it is possible to identify $P(T \in B, \text{Complier} \mid T \geq t, X, Z = t')$ and it turns out that this probability is equal to p_2^*.

In the following proposition we state the main result of our study:

Proposition 2.2. *Under Assumptions A1-A5 $GATETCS(t, t', t)$ is nonparametrically identified and it holds*

(2.7)

$$GATETCS(t, t', t) = \frac{P(T \in B \mid T \geq t, X, Z = t) - P(T \in B \mid T \geq t, X, Z = t')}{P(S = t \mid T \geq t, X, Z = t)}$$

The proof is provided in the appendix. This treatment effect has a nice interpretation. It adjusts the difference between the observed outcomes in the groups of compliers and noncompliers by the probability to be a complier. Thus, this estimator accounts for unobserved heterogeneity, captured by this difference. The identification strategy naturally applies to hazards. This result is similar to the result of [Imbens and Angrist, 1994], which turns to be its static special case.

Another interesting setting case arises if we restrict the model to full compliance. Then $P(S = t \mid T \geq t, X, Z = t)$ is equal to 1 and we obtain

$$GATETCS(t, t', t) = P(T \in B \mid T \geq t, X, Z = t) - P(T \in B \mid T \geq t, X, Z = t').$$

This is the result from [Van den Berg et al., 2010], where an exogeneous treatment is assumed. Thus, their model is revealed to be a special case of our model.

One additional advantage of our approach is that the setting can be defined and interpreted in an alternative way. All results apply to a case, where all individuals have the same point in time of inflow but the assigned treatment time vary in a random way.

2.3 Estimation of treatment effects and testing

To motivate our estimation strategy, we write first

$$P(T \in B \mid T \geq t, X = x, Z = t) = \mathbb{E}[\mathbb{I}_{\{T \in B\}} \mid T \geq t, X = x, Z = t], \tag{2.8}$$

where $\mathbb{I}$ is an indicator function (we assume here for expositional reasons that X is discrete and one-dimensional, this is however not a restriction for our further analysis). We are using a local linear estimator to estimate this expression. In particular, define $g(t, x, t)$ as

$$g(t, x, t) = \int_t^{+\infty} \frac{1}{nh} \sum_{i=1}^{n} (\mathbb{I}_{\{T \in B\},i} - \beta_0 - \beta_1(T_i - y) - \beta_2(X_i - x) - \beta_3(Z_i - t))^2 K\Big(\frac{T_i - y}{h(n)}\Big) \mathbb{I}(X_i = x, Z_i = t) dy.$$

K is a kernel function and h is a bandwidth (technical details to come), and the expression is minimized with respect to all β-s. We assume that all regularity conditions are fulfilled. Analogously, we define $g(t, x, t')$ and $k(t, x, t)$ to be the corresponding local linear estimators of $P(T \in B \mid T \geq t, X = x, Z = t')$ and $P(S = t \mid T \geq t, X = x, Z = t)$, respectively. Then, we define our estimator of $GATETCS(t, t', t)$, $\hat{GATETCS}(t, t', t)$, as

$$\hat{GATETCS}(t, t', t) = \frac{g(t, x, t) - g(t, x, t')}{k(t, x, t)}.$$

Proposition 2.3 (Consistency)**.** *Under standard regularity conditions,*

$$\hat{GATETCS}(t, t', t) \xrightarrow{p} GATETCS(t, t', t).$$

Proof. Follows directly from the properties of the local linear estimator and the fact, that the function $f(a, b, c) = \frac{a-b}{c}$ is a continuous function at all points where it is defined.

□

This estimator is a Wald-type statistics. Its limiting distribution is to be provided soon.

2.4 A test for static endogeneity

In this subsection we suggest a simple procedure to test for static endogeneity, i.e. to test whether noncompliance is related to potential outcomes. The idea is that if it is not, then noncompliers and compliers, and hence noncompliers and a whole cohort, should have the same distributions of the potential outcomes. In particular, they should have the same distributions of the potential outcomes without the treatment. This argument naturally leads to a comparison between the observed outcomes of the noncompliers

from a t-cohort at elapsed duration t (and hence at the moment where they can be identified as noncompliers) with the observed outcomes of a whole older cohort, say the t'-cohort, at the same elapsed duration t.

To develop this idea in a formal framework, we adopt the following assumption:
A6 (No static endogeneity): $S(t) \perp\!\!\!\perp \{T(s)\} \mid X$,

A direct implication of A6 is the following equality:

(2.9)
$$P(T(\infty) \in B \mid T(\infty) \geq t, X, S(t) = t) = P(T(\infty) \in B \mid T(\infty) \geq t, X, S(t) = \infty)$$

We state the following

Proposition 2.4. *Under assumptions A1-A6*

$$P(T \in B \mid T \geq t, X, Z = t') = P(T(\infty) \in B \mid T(\infty) \geq t, X, S(t) = \infty).$$

for all $t' \geq t$

Verbally stated, the average conditional distribution of the observed outcome at elapsed duration t from a t'-cohort, $t' \geq t$, is equal to the average conditional distribution of the potential duration without the treatment on the set of all noncompliers at age t.

Proof. See appendix.

We already showed in a previous subsection that

$$P(T(\infty) \in B \mid T(\infty) \geq t, X, S(t) = \infty) = P(T \in B \mid T \geq t, X, S = \infty, Z = t),$$

so under no static endogeneity it must hold

$$P(T \in B \mid T \geq t, X, Z = t') = P(T \in B \mid T \geq t, X, S = \infty, Z = t).$$

Both probabilities can be estimated from the data. This naturally leads to the following Null hypothesis:
H_0 **(no static endogeneity)**:

$$P(T \in B \mid T \geq t, X, Z = t') - P(T \in B \mid T \geq t, X, S = \infty, Z = t) = 0.$$

The derivation of a test statistic with a local linear technique, as well as its distribution is still work in progress.

2.5 Identification of treatment effects: the case with right censoring

So far we have assumed we can observe the whole lenght of spells in the state of interest, T. A typical feature of duration data is that some observations are censored. In this section we consider right censoring (it is straightforward to extend all results to left censoring). In labor market studies right censoring typically arizes when at the end of the study the individuals are still unemployed, so the unemployment spell has an unknown length. The unemployed might also simply stop attending the training and drop out of the study (sample attrition), or their job search might be interrupted by a transition into the out of the labor force state because of maternity, invalidity or other reasons. In biomedical studies, and particulary in clinical trials, spells might be right-censored because patients withdraw from treatment or die from another cause (competing risks). In telemetry studies the radio transmitter attached to the animal might break down or get lost.
Right censoring can be introduced formally in the following way: let C be a real Borel measurable fuction with nonnegative values. We observe $(\min\{T, C\}, \mathbb{I})$ and not directly (T, C), where $\mathbb{I}$ indicates whether $\min\{T, C\} = T$ or $\min\{T, C\} = C$ or both. Unfortunately, it is not possible to recover the joint distribution of T and C from the distribution of the observables $(\min\{T, C\}, \mathbb{I})$ without imposing additional structure. To each pair of dependend latent variables (T_d, C_d) there exists an independent pair (T_i, C_i), which is observationally equivalent, a result which goes back to [Cox, 1962] and [Tsiatis, 1975]. To achieve identification some structure has to be imposed. We prove identification of treatment effects with two different types of right censoring.

2.5.1 Censoring with independence of T and $\mathbb{I}_{\{C>T\}}$

Suppose the following assumption holds:

A6)
$$\mathbb{I}_{\{C>T\}} \perp\!\!\!\perp (T, S) \mid X, Z$$
where $\mathbb{I}$ is an indicator function.

It implies, that censored and uncensored observations have the same distributions conditional on X and Z: $F_{T|X,Z,\mathbb{I}_{\{C>T\}}=1} = F_{T|X,Z,\mathbb{I}_{\{C>T\}}=0}$. This is a strong assumption but is often made in empirical studies (ZITAT?), since it allows easy implementation and intutitive interpretation. To demonstrate these, set $\widetilde{T} = \min\{T, C\}$ and $\widetilde{S} = \min\{T, C, S\}$. It holds the following

Proposition 2.5. *Under assumptions **A1 - A6** GATETCS is identified and equal to*

$$\frac{P(\widetilde{T} \in B \mid \widetilde{T} \geq t, \mathbb{I}_{\{C>T\}} = 1, X, Z = t) - P(\widetilde{T} \in B \mid \widetilde{T} \geq t, \mathbb{I}_{\{C>T\}} = 1, X, Z = t')}{P(\widetilde{S} = t \mid \widetilde{T} \geq t, \mathbb{I}_{\{C>T\}} = 1, X, Z = t)}. \tag{2.10}$$

Proof. It is straightforward to show that

$$P(\widetilde{T} \in B \mid \widetilde{T} \geq t, \mathbb{I}_{\{C>T\}} = 1, X, Z = t) = P(T \in B \mid T \geq t, X, Z = t)$$
$$P(\widetilde{T} \in B \mid \widetilde{T} \geq t, \mathbb{I}_{\{C>T\}} = 1, X, Z = t') = P(T \in B \mid T \geq t, X, Z = t') \qquad \text{and}$$
$$P(\widetilde{S} = t \mid \widetilde{T} \geq t, \mathbb{I}_{\{C>T\}} = 1, X, Z = t) = P(S = t \mid T \geq t, X, Z = t).$$

The lefthand sides of these equations contain only observables.

So to obtain GATETCS, it is enough to take the uncensored observations.

2.5.2 Random censoring

In this subsection we adopt the following assumption:

A7) (Random censoring)
$$C \perp\!\!\!\perp (T, S) \mid X, Z.$$

A8) The random vector $(T, C, S) \mid X, Z$ has an absolutely continuous distribution.

A9) $T \mid X, Z$, $C \mid X, Z$ and $S \mid X, Z$ have distributions with nonvanishing right tails, $F_{T|X,Z}(t) < 1$ for $x < \infty$.

We can prove the following propostion:

Proposition 2.6. *Under assumptions **A1 - A5** and **A7- A9** GATETCS is identified.*

Proof. Let F be a simbol for p.d.f and f for a density. Following [Nadas, 1970], it holds for the joint density of T and C given X, Z, $f_{T,C|X,Z=k}$,

$$P(T > t, C > T \mid X, Z = k) = \int_t^\infty \int_y^\infty f_{T,C|X,Z=t}(y, s) ds dy. \tag{2.11}$$

Since $C \perp\!\!\!\perp (T, S) \mid X, Z$, we have

$$\int_t^\infty \int_y^\infty f_{T,C|X,Z=t}(y, s) ds dy = \int_t^\infty \int_y^\infty f_{T|X,Z=t}(y) f_{C|X,Z=t}(s) ds dy$$
$$= \int_t^\infty f_{T|X,Z=t}(y) \Big(\int_y^\infty f_{C|X,Z=t}(s) ds \Big) dy = \int_t^\infty f_{T|X,Z=t}(y) P(C > y \mid X, Z = k) dy =$$
$$= \int_t^\infty f_{T|X,Z=t}(y)(1 - P(C < y \mid X, Z = k)) dy.$$

On the other hand, we have

$$P(T > t, C > T \mid X, Z = k) = P(C > T \mid X, Z = k) P(T > t \mid C > T, X, Z = k)$$

so we obtain, after taking the derivatives w.r.t. t on both sides in 2.11,

$$P(C > T \mid X, Z = k) f_{T|C>T,X,Z=t}(t) = f_{T|X,Z=t}(t)(1 - P(C < t \mid X, Z = k)) \qquad \text{and hence}$$

$$P(C > T \mid X, Z = k) f_{T|C>T,X,Z=t}(t) =$$

$$= \frac{f_{T|X,Z=t}}{1 - P(T < t \mid X, Z = k)}(t)(1 - P(C < t \mid X, Z = k))(1 - P(T < t \mid X, Z = k)) \qquad \text{and therefore}$$

$$P(C > T \mid X, Z = k) f_{T|C>T,X,Z=t}(t) = -\frac{\partial \ln(1 - F_{T|X,Z=k}(t))}{\partial t}(t) P(\widetilde{T} > t \mid X, Z = k) \qquad \text{so finally}$$

$$-\frac{\partial \ln(1 - F_{T|X,Z=k}(t))}{\partial t} = \frac{P(C > T \mid X, Z = k)}{P(\widetilde{T} > t \mid X, Z = k)} f_{T|C>T,X,Z=t}(t).$$

The right hand side can be estimated from the data. □

3 Instrumental variable approach for a duration model with noncompliance in the treatment and control groups

3.1 Notation and treatment effects

To motivate our second model, recall that in our first approach we used the different timing of the assigned treatment to identify the treatment effect. We either postulated a common for all individuals point in time for the assigned treatment but different points in time of inflow, or a common point in time of inflow but a treatment assignment at different elapsed durations across individuals. We allowed noncompliance as a selection out of the treatment (see section 4 for a discussion.

In our second approach we abandon the regression discontinuity setting. Here we are not going to use different timing as an instrument. We adopt rather a setting, where all individuals are assigned in time 0, point in time of inflow, to a treatment or control group. The actual treatment is obtained at some later (random or deterministic) point in time $Z = z$, $z \in \mathbb{R}_+$, which is common for everybody. We assume that Z can take countable many values $t_1, t_2, \ldots$[2]. Let the random variable W_i have possible values 0 and 1 and denote the assignment to the treatment or control group, respectively. We assume that this assignment is randomized across individuals at time 0 conditional on observable and unobservable characteristics, see next section for a formal statement of the assumptions. At point in time Z the individuals have to choose the actual treatment and then immediately obtain their choice. Following the standard counterfactual notation, let $S_i(w, z)$ denote the potential actual treatment chosen by individual i if he was assigned to treatment w at day z, $w = 0, 1$. The observed actual treatment is denoted as before with

[2] If Z is a random variable we assume that $P(Z = t_n) > 0$ for all $n \in \mathbb{N}$.

S. Thus, we adopt a setting, where 1) all individuals obtain an assignment (to a treatment or control group) at the same elapsed duration and 2) a "real" randomization occurs, in the sense that there is some intermediate instance or layer which conducts the assignment W. Our purpose is to introduce a less mechanical model than our first approach. A typical example from labor economics is a social experiment, where a case worker assigns at random individuals upon inflow into unemployment to receive ($W = 1$) or not receive ($W = 0$) a job training after some, common for everybody unemployment duration. We allow for noncompliance in both treatment and control group, an assumption which describes well examples from social experiments or pilot projects. Let $T_i(s, w, z)$ be the potential random duration, which the individual i would spend in the state of interest if he was assigned to treatment w and received a treatment s at a point in time z. Throughout this paper we adopt the standard exclusion restriction, $T(s, w, z) = T(s, z)$. This implies that the treatment assignment has no direct causal impact on the duration variable. As before we denote with T the actual duration. The meaning of $\tilde{T}$,X and V stays unchanged as well, with additional independence of C and W. To each realization of Z we observe the sample consisting of the independent observations

$$(\tilde{T}_1, S_1, W_1, X_1, \mathbb{I}_{T_1 \leq C_1}), \ldots, (\tilde{T}_n, S_n, W_n, X_n, \mathbb{I}_{T_n \leq C_n}),$$

drawn from the distribution of $(\tilde{T}, S, W, X, \mathbb{I}_{T \leq C})$. Define B as in our first model, $B := [t, t + a]$, but here we are not going need restrictions on a. We define our treatment effects in a similar way as in approach 1, here a definition only for *GATETCS*, here denote simply with *TE*:

$$\begin{aligned} TE(t, z) \;:=\; & \mathbb{E}\Big[P(T(1, z) \in B \mid T(1, z) \geq t, X, S(0, z) = 0, S(1, z) = 1, Z = z, V) \\ - \;& P(T(0, z) \in B \mid T(0, z) \geq t, X, S(0, z) = 0, S(1, z) = 1, Z = z, V) \mid \\ & T(1, z) \geq t, X, S(0, z) = 0, S(1, z) = 1, Z = z\Big]. \end{aligned}$$

3.2 Identification of treatment effects

In this subsection we are going to prove identification of $TE(t, t)$. For expositional reasons we are againg considering the case of no censoring. We impose assumptions that are quite similar to these from our first model:

B0 **(Exclusion restriction)**: $T(s, w, z) = T(s, z)$

B1 **(Global monotonicity)**: for any $z \in \mathbb{R}_+$ and all $i \in \{1, 2, \ldots, n\}$, $w \in \{0, 1\}$ it holds either $S_i(w, z) = w$ or $S_i(w, z) = 1$ or $S_i(w, z) = 0$. Following [Imbens and Rubin, 1997], we refer to an individual in these three cases as to a complier, always-taker or never-taker, respectively.

B2 **(No anticipation)**: For all $0 \leq t \leq z < +\infty$ and each X, V it holds

$$\begin{aligned} & \Theta_{T(1,z)}(t \mid X, V, S(1, z) = 1, S(0, z) = 0, Z = z) = \\ = \;& \Theta_{T(0,z)}(t \mid X, V, S(1, z) = 1, S(0, z) = 0, Z = z). \end{aligned}$$

This assumption is already discussed in section 1. Here we condition on the set of all compliers. We assume further, that the above equality holds also for the sets of always-takers and never-takers (these are in fact empty assumptions if we allow for actual noncompliance. We adopt them for technical reasons).

B3 **(Randomization)**: For the instrument W it holds for all $z \in \mathbb{R}_+$

$$i) \quad W \perp\!\!\!\perp \{T(s,z), S(w,z)\} \mid X, V, Z \qquad \text{and} \qquad ii) \quad W \perp\!\!\!\perp V \mid X, Z.$$

B4 **(Consistency)**: For all $z \in \mathbb{R}_+$, $w \in \{0,1\}$ and $s \in \{0,1\}$

i) $W = w, Z = z \Rightarrow S(w,z) = S$

ii) $S = s, Z = z \Rightarrow T(s,z) = T$.

Also we further assume, that all expressions below are well-defined, in particular that the sets of all compliers, never-takers and always-takers at $Z = z$ have a positive measure.

In line with our first model we state the following

Proposition 3.1. *Let F be a cdf. Under assumptions B0-B4 it holds for all $0 \le t \le z < +\infty$*

$$F_{V|T(1,z)\ge t,X,S(1,z)=1,S(0,z)=0,Z=z} = F_{V|T(0,z)\ge t,X,S(1,z)=1,S(0,z)=0,Z=z}.$$

Proof. See appendix.

An immediate consequence of this proposition, following the same steps as in model 1, is that we can rewrite $TE(t,z)$ in the following way:

(3.1)

$$\begin{aligned} TE(t,z) &= P(T(1,z) \in B \mid T(1,z) \ge t, X, S(0,z) = 0, S(1,z) = 1, Z = z) - \\ &- P(T(0,z) \in B \mid T(0,z) \ge t, X, S(0,z) = 0, S(1,z) = 1, Z = z). \end{aligned}$$

To prove identification of 3.1, we use the strategy of [Imbens and Rubin, 1997] adapted to our dynamic case. We are going to 1) split observed distributions into their potential components and 2) then use the information on noncompliance to identify these components. We now elaborate on these two points.

1) Consider first individuals with observed $S = 1, W = 1$. They can be either compliers or always-takers. Similarly, individuals with $S =, W = 0$ are either compliers or never-takers. The expression $P(T \in B \mid T \ge t, X, S = 1, W = 1, Z = t)$ contains only observables. Using consistency and randomization, we have

$$\begin{aligned} & P(T \in B \mid X, S = 1, W = 1, Z = t) = P(T(1,t) \in B \mid X, S = 1, W = 1, Z = t) = \\ = \; & \frac{P(T(1,t) \in B, S = 1 \mid X, W = 1, Z = t)}{P(S = 1 \mid X, W = 1, Z = t)} = \frac{P(T(1,t) \in B, S(1,t) = 1 \mid X, W = 1, Z = t)}{P(S(1,t) = 1 \mid X, W = 1, Z = t)} = \\ = \; & \frac{P(T(1,t) \in B, S(1,t) = 1 \mid X, Z = t)}{P(S(1,t) = 1 \mid X, Z = t)} = P(T(1,t) \in B \mid X, S(1,t) = 1, Z = t), \end{aligned}$$

so finally we obtain

(3.2)
$$P(T \in B \mid T \geq t, X, S = 1, W = 1, Z = t) = P(T(1,t) \in B \mid T(1,t) \geq t, X, S(1,t) = 1, Z = t).$$

An intuitive interpretation is that the conditional distribution of the observed duration on the group with observed $S = 1, W = 1$ is equal to its potential counterpart, given the individual is either a complier or always-taker and has received the treatment. Denote for simplicity

$$Q := \{S(0,z) = 0, S(1,z) = 1\},$$
$$A := \{S(0,z) = 1, S(1,z) = 1\},$$
$$N := \{S(0,z) = 0, S(1,z) = 0\}.$$

Using simple rules for probabilities, we write

$$\begin{aligned} & P(T(1,t) \in B \mid T(1,t) \geq t, X, S(1,t) = 1, Z = t) \\ = \; & P(T(1,t) \in B \mid T(1,t) \geq t, X, Q, Z = t) \frac{P(Q \mid T(1,t) \geq t, X, Z = t)}{P(Q \cup A \mid T(1,t) \geq t, X, Z = t)} + \\ + \; & P(T(1,t) \in B \mid T(1,t) \geq t, X, A, Z = t) \frac{P(A \mid T(1,t) \geq t, X, Z = t)}{P(Q \cup A \mid T(1,t) \geq t, X, Z = t)}, \end{aligned}$$

which leads to the result

(3.3)
$$\begin{aligned} & P(T \in B \mid T \geq t, X, S = 1, W = 1, Z = t) = \\ = \; & P(T(1,t) \in B \mid T(1,t) \geq t, X, Q, Z = t) \frac{P(Q \mid T(1,t) \geq t, X, Z = t)}{P(Q \cup A \mid T(1,t) \geq t, X, Z = t)} + \\ + \; & P(T(1,t) \in B \mid T(1,t) \geq t, X, A, Z = t) \frac{P(A \mid T(1,t) \geq t, X, Z = t)}{P(Q \cup A \mid T(1,t) \geq t, X, Z = t)}. \end{aligned}$$

In other words, we splitted the observed distribution into the sum of the potential outcomes of compliers and always-takers, weighted by their proportions. In order to obtain the term $P(T(1,t) \in B \mid T(1,t) \geq t, X, Q, Z = t)$ we will identify all other components on the right side.

2) This we are going to achieve using the information on noncompliance. We can identify all individuals with observed $S = 1, W = 0$ as always-takers and all individuals with $S = 0, W = 1$ as never-takers. We use this information (together with no anticipation) to identify the potential distributions of $T(1,t) \mid T(1,t) \geq t$ and $T(0,t) \mid T(0,t) \geq t$ on the sets of the always-takers and never-takers, respectively. Following the steps of [Imbens and Rubin, 1997], it holds

(3.4)
$$\begin{aligned} & P(T(1,t) \in B \mid T(1,t) \geq t, X, S(0,z) = 1, S(1,z) = 1, Z = t) = \\ = \; & P(T \in B \mid T \geq t, X, S = 1, W = 0, Z = t), \end{aligned}$$

(3.5)
$$\begin{aligned} & P(T(0,t) \in B \mid T(0,t) \geq t, X, S(0,z) = 0, S(1,z) = 0, Z = t) = \\ = \; & P(T \in B \mid T \geq t, X, S = 0, W = 1, Z = t), \end{aligned}$$

and for the proportions

(3.6)
$$\begin{aligned} & P(S(0,z) = 1, S(1,z) = 1 \in B \mid T(1,t) \geq t, X, Z = t) = \\ = \; & P(S = 1 \mid T \geq t, X, W = 0, Z = t), \end{aligned}$$

and

(3.7)
$$\begin{aligned} & P(S(0,z) = 0, S(1,z) = 0 \in B \mid T(1,t) \geq t, X, Z = t) = \\ = \; & P(S = 0 \mid T \geq t, X, W = 1, Z = t). \end{aligned}$$

Inserting 3.4, 3.5, 3.6, 3.7 and

$$P(Q \in B \mid T(1,t) \geq t, X, Z = t) = 1 - P(A \in B \mid T(1,t) \geq t, X, Z = t) - P(N \in B \mid T(1,t) \geq t, X, Z = t)$$

into equation 3.3, we finally obtain

(3.8)
$$\begin{aligned} & P(T(1,t) \in B \mid T(1,t) \geq t, X, S(0,z) = 0, S(1,z) = 1, Z = t) = \\ = \; & \Big(P(T \in B, S = 1 \mid T \geq t, X, W = 1, Z = t) - P(T \in B, S = 1 \mid T \geq t, X, W = 0, Z = t)\Big). \\ . \; & \Big(P(S = 1 \mid T \geq t, X, W = 1, Z = t) - P(S = 1 \mid T \geq t, X, W = 0, Z = t)\Big)^{-1} \\ =: \; & F_1. \end{aligned}$$

Analogously,

(3.9)
$$\begin{aligned} & P(T(1,0) \in B \mid T(1,0) \geq t, X, S(0,z) = 0, S(1,z) = 1, Z = t) = \\ = \; & \Big(P(T \in B, S = 0 \mid T \geq t, X, W = 0, Z = t) - P(T \in B, S = 0 \mid T \geq t, X, W = 1, Z = t)\Big). \\ . \; & \Big(P(S = 0 \mid T \geq t, X, W = 0, Z = t) - P(S = 0 \mid T \geq t, X, W = 1, Z = t)\Big)^{-1} \\ =: \; & F_0. \end{aligned}$$

With these steps we proved the following:

Proposition 3.2. *Under assumptions B0-B4 $TE(t,t)$ is identified and equal to $F_1 - F_0$.*

This is the central result of this model. Again, an interesting case arises when $t = 0$, namely $TE(0,0) = \Delta_Q(a)$, where is the treatment effect defined (and identified) in [Abbring and van den Berg, 2005], and $t + a$ is the right boundary of $B = [t, t + a]$ (here $t = 0$).

The estimation and testing part of this model is still work in process.

3.3 A test for static endogeneity

Results 3.4 and 3.5 give rise to an interesting application. We can use them to test for endogeneous selection. The idea is that if noncompliance is not related to potential outcomes, then the potential conditional distributions of the duration variable with and without the treatment of the always-takers and never-takers, respectively, should be the same as those of the compliers and hence equal to the observed distributions in both treatment arms. This leads to the natural Null hypothesis

$$
\begin{aligned}
&H_0 \qquad \text{(exogeneous selection)}: \\
&P(T \in B \mid T \geq t, X, S = 1, W = 0, Z = t) = P(T \in B \mid T \geq t, X, S = 1, Z = t), \\
&P(T \in B \mid T \geq t, X, S = 0, W = 1, Z = t) = P(T \in B \mid T \geq t, X, S = 0, Z = t).
\end{aligned}
$$

The deriving of a test-statistic and its distribution is to come.

4 Conclusion

In this paper we developed two very general IV approaches for duration data. We proved identification without imposing parametric or semi-parametric structure and in the presence of dynamic and static endogeneity. Our assumptions are not restrictive and can be even further relaxed. As an example, in our first model always-takers can be introduced without hampering the main identification strategy. All results apply also for hazards. As a future work one could consider conditioning on the propensity score instead on covariates. There is a lot of work to do in this fruitful research field.

5 Appendix

5.1 Proof of Proposition 2.1

1. First we show that from the no anticipation assumption the following result holds:

   $$P(T(t) \geq t \mid X, S(t) = t) = P(T(t') \geq t \mid X, S(t) = t). \tag{5.1}$$

   This is so because

   $$
   \begin{aligned}
   & P(T(t) \geq t \mid X, S(t) = t, V) = \exp(-\Theta_{T(t)}(t \mid X, S(t) = t, V)) = \\
   \overset{\text{No anticipation}}{=} \quad & \exp(-\Theta_{T(t')}(t \mid X, S(t) = t, V)) = P(T(t') \geq t \mid X, S(t) = t, V)
   \end{aligned}
   $$

so that we obtain

$$
\begin{aligned}
& P(T(t) \geq t \mid X, S(t) = t) = \mathbb{E}\left[I_{\{T(t)\geq t\}} \mid X, S(t) = t\right] = \\
= \; & \mathbb{E}\left[\mathbb{E}\left[I_{\{T(t)\geq t\}} \mid X, S(t) = t, V\right] \mid X, S(t) = t\right] = \\
= \; & \mathbb{E}\left[P(T(t) \geq t \mid X, S(t) = t, V) \mid X, S(t) = t\right] = \\
= \; & \mathbb{E}\left[P(T(t') \geq t \mid X, S(t) = t, V) \mid X, S(t) = t\right] = \\
= \; & \mathbb{E}\left[\mathbb{E}\left[I_{\{T(t')\geq t\}} \mid X, S(t) = t, V\right] \mid X, S(t) = t\right] = P(T(t') \geq t \mid X, S(t) = t)
\end{aligned}
$$

where $I_{\{T(s)\in B\}}$ is an indicator function equal to 1 when $T(s) \in B$ (of course from these steps we also see that $P(T(t) \geq t \mid X, S(t) = t, V) = P(T(t') \geq t \mid X, S(t) = t, V)$).

2. Next, using this result, we show $F_{V|T(t)\geq t,X,S(t)=t} = F_{V|T(t')\geq t,X,S(t)=t}$. Let B be an element from the standard Borel sigma-algebra $\mathcal{B}$. It holds

$$
\begin{aligned}
& P(V \in B \mid T(t') \geq t, X, S(t) = t) = \frac{P(V \in B, T(t') \geq t \mid X, S(t) = t)}{P(T(t') \geq t \mid X, S(t) = t)} = \\
\overset{\text{result 5.1}}{=} \; & \frac{P(V \in B, T(t') \geq t \mid X, S(t) = t)}{P(T(t) \geq t \mid X, S(t) = t)} = \\
= \; & \frac{P(V \in B \mid X, S(t) = t) P(T(t') \geq t \mid X, S(t) = t, V \in B)}{P(T(t) \geq t \mid X, S(t) = t)} = \\
\overset{\text{result 5.1}}{=} \; & \frac{P(V \in B \mid X, S(t) = t) P(T(t) \geq t \mid X, S(t) = t, V \in B)}{P(T(t) \geq t \mid X, S(t) = t)} = \\
= \; & \frac{P(V \in B, T(t) \geq t \mid X, S(t) = t)}{P(T(t) \geq t \mid X, S(t) = t)} = P(V \in B \mid T(t) \geq t, X, S(t) = t).
\end{aligned}
$$

3. Now we show $F_{V|T(t)\geq t,X,S(t)=t} = F_{V|T\geq t,X,S=t,Z=t}$. First we observe that $Z \perp\!\!\!\perp \{T(s), S(z)\} \mid X, V$ and $Z \perp\!\!\!\perp V \mid X$ together imply $Z \perp\!\!\!\perp \{T(s), S(z)\} \mid X$ (ZITAT WEAK UNION; PEARL 2000). Then, we have

$$
P(V \in B \mid T(t) \geq t, X, S(t) = t) = \frac{P(V \in B \mid X, S(t) = t) P(T(t) \geq t \mid X, S(t) = t, V \in B)}{P(T(t) \geq t \mid X, S(t) = t)},
$$

so now we study the separate components of this expression.

(a) It holds

$$
\begin{aligned}
P(V \in B \mid X, S(t) = t) &= \frac{P(V \in B, S(t) = t \mid X)}{P(S(t) = t \mid X)} \\
&\overset{Z \perp\!\!\!\perp S(z) \mid X}{=} \frac{P(V \in B, S(t) = t \mid X)}{P(S(t) = t \mid X, Z = t)} = \\
&\overset{\text{Consistency}}{=} \frac{P(V \in B, S(t) = t \mid X)}{P(S = t \mid X, Z = t)} = \frac{P(V \in B \mid X) P(S(t) = t \mid X, V \in B)}{P(S = t \mid X, Z = t)} = \\
&\overset{Z \perp\!\!\!\perp V \mid X,\ Z \perp\!\!\!\perp S(z) \mid X, V}{=} \frac{P(V \in B \mid X, Z = t) P(S(t) = t \mid X, V \in B, Z = t)}{P(S = t \mid X, Z = t)} = \\
&\overset{\text{Consistency}}{=} \frac{P(V \in B \mid X, Z = t) P(S = t \mid X, V \in B, Z = t)}{P(S = t \mid X, Z = t)} = \\
&= P(V \in B \mid X, S = t, Z = t).
\end{aligned}
$$

(b) Further,

$$
\begin{aligned}
&P(T(t) \geq t \mid X, S(t) = t, V \in B) = \\
&= \frac{P(T(t) \geq t, S(t) = t \mid X, V \in B)}{P(S(t) = t \mid X, V \in B)} = \\
&\overset{Z \perp\!\!\!\perp \{T(s), S(z)\} \mid X, V}{=} \frac{P(T(t) \geq t, S(t) = t \mid X, V \in B, Z = t)}{P(S(t) = t \mid X, V \in B, Z = t)} \\
&\overset{\text{Consistency}}{=} \frac{P(T(t) \geq t, S = t \mid X, V \in B, Z = t)}{P(S = t \mid X, V \in B, Z = t)} = \\
&= P(T(t) \geq t \mid X, S = t, V \in B, Z = t) = \\
&\overset{\text{Consistency}}{=} P(T \geq t \mid X, S = t, V \in B, Z = t).
\end{aligned}
$$

(c) Using $Z \perp\!\!\!\perp \{T(s), S(z)\} \mid X$ instead of $Z \perp\!\!\!\perp \{T(s), S(z)\} \mid X, V$ with else completely identical steps as in the last point, we obtain

$$P(T(t) \geq t \mid X, S(t) = t) = P(T \geq t \mid X, S = t, Z = t)$$

So finally we get the equality

$$
\begin{aligned}
&P(V \in B \mid T(t) \geq t, X, S(t) = t) = \\
&= \frac{P(V \in B \mid X, S = t, Z = t) P(T \geq t \mid X, S = t, V \in B, Z = t)}{P(T \geq t \mid X, S = t, Z = t)} = \\
&= P(V \in B \mid T \geq t, X, S = t, Z = t)
\end{aligned}
$$

□

5.2 Proof of proposition 2.2

First, consider the conditional distribution of the duration variable for the treatment group, $F_{T(t)|T(t)\geq t,X,S(t)=t}$. Let $\mathcal{B}_+$ be the spur (?correct in english?) sigma field $\mathbb{R}_+ \cap \mathcal{B}$, where $\mathcal{B}$ is the standard Borel sigma algebra. Set $B_t := B \cap [t, +\infty)$ for $B \in \mathcal{B}_+$. Then, for each $B \in \mathcal{B}_+$ it holds

$$P(T(t) \in B \mid T(t) \geq t, X, S(t) = t) = \frac{P(T(t) \in B_t \mid X, S(t) = t)}{P(T(t) \geq t \mid X, S(t) = t)}$$

and for $P(T(t) \in B_t \mid X, S(t) = t)$

$$P(T(t) \in B_t \mid X, S(t) = t) = \frac{P(T(t) \in B_t, S(t) = t \mid X)}{P(S(t) = t \mid X)} =$$

$$\overset{A3,\text{Randomization}}{=} \frac{P(T(t) \in B_t, S(t) = t \mid X, Z = t)}{P(S(t) = t \mid X, Z = t)} \overset{A5,\text{Consistency}}{=} \frac{P(T(t) \in B_t, S = t \mid X, Z = t)}{P(S = t \mid X, Z = t)} =$$

$$= P(T(t) \in B_t \mid X, S = t, Z = t) \overset{\text{Consistency}}{=} P(T \in B_t \mid X, S = t, Z = t)$$

Set $B = \mathbb{R}_+$ and one obtains

$$P(T(t) \geq t \mid X, , S(t) = t) = P(T \geq t \mid X, V, S = t, Z = t),$$

so finally

$$P(T(t) \in B \mid T(t) \geq t, X, S(t) = t) = \frac{P(T(t) \in B_t \mid X, S(t) = t)}{P(T(t) \geq t \mid X, S(t) = t)} =$$

$$= \frac{P(T \in B_t \mid X, S = t, Z = t)}{P(T \geq t \mid X, S = t, Z = t)} = P(T \in B \mid T \geq t, X, S = t, Z = t),$$

where the last expression consists only of observables. To procede in the same way for the control group, there is one obstacle. To see this, let $t' > t$. Then in a dynamic case both ∞ and t' mean at t: no treatment (unlike the binary case, where 0 is more concrete). So the complyers of the control group (the $Z = t'$) cannot be identified at t. Another way to state the problem is that $P(T \in B \mid T \geq t, S = t, Z = t')$ is not defined. A starting point to overcome this problem is the equality

$$P(T \in B \mid T \geq t, X, Z = t') = P(T \in B \mid T \geq t, X, Z = t', S(t) = t) P(S(t) = t \mid T \geq t, X, Z = t') +$$
$$+ \quad P(T \in B \mid T \geq t, X, Z = t', S(t) = \infty) P(S(t) = \infty \mid T \geq t, X, Z = t'),$$

where $P(T \in B \mid T \geq t, X, Z = t')$ contains only observables. Our identification proof contains the following steps:

1. Show that $P(T \in B \mid T \geq t, X, Z = t', S(t) = t)$ is equal to $P(T(t') \in B \mid T(t') \geq t, X, S(t) = t)$ and is therefore the expression we want to identify.

2. Show that

$$P(T \in B \mid T \geq t, X, Z = t', S(t) = \infty) = P(T \in B \mid T \geq t, X, Z = t, S(t) = \infty),$$

i.e. the noncompliers of the two cohorts have identical potential duration distributions at t.

3. Show that $P(S(t) = t \mid T \geq t, X, Z = t')$ and $P(S(t) = \infty \mid T \geq t, X, Z = t')$ are identified (these are the proportions of compliers and noncompliers at t).

Let's proof these 3 points:

1. It holds

$$\begin{aligned}
& P(T \in B \mid X, Z = t', S(t) = t) = \\
= \quad & P(T \in B \mid X, Z = t', S(t) = t, S = t')P(S = t' \mid X, Z = t', S(t) = t) + \\
+ \quad & P(T \in B \mid X, Z = t', S(t) = t, S = \infty)P(S = \infty \mid X, Z = t', S(t) = t) = \\
\overset{\text{Consistency}}{=} \quad & P(T(t') \in B \mid X, Z = t', S(t) = t, S = t')P(S = t' \mid X, Z = t', S(t) = t) + \\
+ \quad & P(T(\infty) \in B \mid X, Z = t', S(t) = t, S = \infty)P(S = \infty \mid X, Z = t', S(t) = t) = \\
\overset{\text{Consistency}}{=} \quad & P(T(t') \in B \mid X, Z = t', S(t) = t, S(t') = t')P(S(t') = t' \mid X, Z = t', S(t) = t) + \\
+ \quad & P(T(\infty) \in B \mid X, Z = t', S(t) = t, S(t') = \infty)P(S(t') = \infty \mid X, Z = t', S(t) = t) = \\
\overset{\text{No anticipation}}{=} \quad & P(T(t') \in B \mid X, Z = t', S(t) = t, S(t') = t')P(S(t') = t' \mid X, Z = t', S(t) = t) + \\
+ \quad & P(T(t') \in B \mid X, Z = t', S(t) = t, S(t') = \infty)P(S(t') = \infty \mid X, Z = t', S(t) = t) = \\
= \quad & P(T(t') \in B \mid X, Z = t', S(t) = t) \overset{\text{Randomization}}{=} P(T(t') \in B \mid X, S(t) = t).
\end{aligned}$$

   Hence, we proved

$$P(T \in B \mid T \geq t, X, Z = t', S(t) = t) = P(T(t') \in B \mid T(t') \geq t, X, S(t) = t) \tag{5.2}$$

2. Using exactly the same steps as in the previous point, we can show

$$\begin{aligned}
P(T \in B \mid X, S(t) = \infty, Z = t') &= P(T(\infty) \in B \mid X, S(t) = \infty) \\
P(T \in B \mid X, S(t) = \infty, Z = t') &= P(T(\infty) \in B \mid X, S(t) = \infty).
\end{aligned}$$

   Consequently, we showed that

$$\begin{aligned}
& P(T \in B \mid T \geq t, X, Z = t', S(t) = \infty) = P(T \in B \mid T \geq t, X, Z = t, S(t) = \infty) \\
= \quad & P(T \in B \mid T \geq t, X, Z = t, S = \infty)
\end{aligned} \tag{5.3}$$

   and the last probability can be estimated nonparametrically.

3. First we prove the following implication from the no anticipation assumption: $P(T \geq t \mid X, Z = t') = P(T \geq t \mid X, Z = t)$ for all $t' \geq t$. This we will refer to as *empirical*

no anticipation relation. Its validity can be proved as follows:

$$
\begin{aligned}
& P(T \geq t \mid X, Z = t') = P(T \geq t \mid X, Z = t', S = t')P(S = t' \mid X, Z = t') + \\
+ \quad & P(T \geq t \mid X, Z = t', S = \infty)P(S = \infty \mid X, Z = t') = \\
\overset{\text{Consistency}}{=} \quad & P(T(t') \geq t \mid X, Z = t', S(t') = t')P(S(t') = t' \mid X, Z = t') + \\
+ \quad & P(T(\infty) \geq t \mid X, Z = t', S(t') = \infty)P(S(t') = \infty \mid X, Z = t') = \\
\overset{\text{Randomization}}{=} \quad & P(T(t') \geq t \mid X, S(t') = t')P(S(t') = t' \mid X) + \\
+ \quad & P(T(\infty) \geq t \mid X, S(t') = \infty)P(S(t') = \infty \mid X,) = \\
\overset{\text{No anticipation}}{=} \quad & P(T(\infty) \geq t \mid X, S(t') = t')P(S(t') = t' \mid X) + \\
+ \quad & P(T(\infty) \geq t \mid X, S(t') = \infty)P(S(t') = \infty \mid X,) = \\
= \quad & P(T(\infty) \geq t \mid X)
\end{aligned}
$$

If we set $t' = t$ and follow exactly the same lines we get

$$P(T \geq t \mid X, Z = t) = P(T(\infty) \geq t \mid X)$$

which means $P(T \geq t \mid X, Z = t') = P(T \geq t \mid X, Z = t)$. Having shown this relation, we now want to prove the following equality:

$$P(S(t) = \infty \mid T \geq t, X, Z = t') = P(S(t) = \infty \mid T \geq t, X, Z = t).$$

We have

$$P(S(t) = \infty \mid T \geq t, X, Z = t') = \frac{P(S(t) = \infty \mid X, Z = t')P(T \geq t \mid S(t) = \infty, X, Z = t')}{P(T \geq t \mid X, Z = t')}.$$

Because of randomization it holds

$$P(S(t) = \infty \mid X, Z = t') = P(S(t) = \infty \mid X, Z = t).$$

Also, shown in point 2 of this proof, we have

$$P(T \geq t \mid S(t) = \infty, X, Z = t') = P(T \geq t \mid S(t) = \infty, X, Z = t).$$

And last, stated above as empirical no anticipation, $P(T \geq t \mid X, Z = t') = P(T \geq t \mid X, Z = t)$, so we get

$$
\begin{aligned}
& P(S(t) = \infty \mid T \geq t, X, Z = t') = \frac{P(S(t) = \infty \mid X, Z = t')P(T \geq t \mid S(t) = \infty, X, Z = t')}{P(T \geq t \mid X, Z = t')} = \\
= \quad & \frac{P(S(t) = \infty \mid X, Z = t)P(T \geq t \mid S(t) = \infty, X, Z = t)}{P(T \geq t \mid X, Z = t)} = P(S(t) = \infty \mid T \geq t, X, Z = t)
\end{aligned}
$$

which we wanted to prove. Further, using consistency,

$$P(S(t) = \infty \mid T \geq t, X, Z = t) = P(S = \infty \mid T \geq t, X, Z = t),$$

where the last expression can be estimated nonparametrically. Taking into account that

$$P(S(t) = t \mid T \geq t, X, Z = t') = 1 - P(S(t) = \infty \mid T \geq t, X, Z = t')$$

we finally obtain the equalities

$$P(S(t) = \infty \mid T \geq t, X, Z = t') = P(S = \infty \mid T \geq t, X, Z = t), \tag{5.4}$$

$$P(S(t) = t \mid T \geq t, X, Z = t') = P(S = t \mid T \geq t, X, Z = t). \tag{5.5}$$

Let's turn now back to the relation we stated previous to the 3 points,

$$P(T \in B \mid T \geq t, X, Z = t') = P(T \in B \mid T \geq t, X, Z = t', S(t) = t)P(S(t) = t \mid T \geq t, X, Z = t') +$$
$$+ \quad P(T \in B \mid T \geq t, X, Z = t', S(t) = \infty)P(S(t) = \infty \mid T \geq t, X, Z = t').$$

Taking into account 5.2, 5.3, 5.4, and 5.5, we can rewrite it as

$$P(T \in B \mid T \geq t, X, Z = t') = P(T(t') \in B \mid T(t') \geq t, X, S(t) = t)P(S = t \mid T \geq t, X, Z = t) +$$
$$+ \quad P(T \in B \mid T \geq t, X, Z = t, S = \infty)P(S = \infty \mid T \geq t, X, Z = t),$$

or finally

$P(T(t') \in B \mid T(t') \geq t, X, S(t) = t) = \frac{P(T \in B \mid T \geq t, X, Z = t') - P(T \in B \mid T \geq t, X, Z = t, S = \infty)P(S = \infty \mid T \geq t, X, Z = t)}{P(S = t \mid T \geq t, X, Z = t)}$.

The right side contains only observables and can be estimated directly via standard nonparametric techniques. The treatment effect $GATETCS(t, t', t)$, which was shown to be equal to $p_1^* - p_2^*$, can be easily simplified to

$$\frac{P(T \in B, S = t \mid T \geq t, X, Z = t) + P(T \in B, S = \infty \mid T \geq t, X, Z = t) - P(T \in B \mid T \geq t, X, Z = t')}{P(S = t \mid T \geq t, X, Z = t)},$$

which leads finally to

$$GATETCS(t, t', t) = \frac{P(T \in B \mid T \geq t, X, Z = t) - P(T \in B \mid T \geq t, X, Z = t')}{P(S = t \mid T \geq t, X, Z = t)}.$$

□

5.3 Proof of Proposition 2.4

Let $t' \geq t$. From no anticipation it follows

$$P(T(t') \geq t \mid X) = P(T(\infty) \geq t \mid X),$$

and since B has the form $B = [t, t + a)$, it follows

$$P(T(t') \in B \mid X) = P(T(\infty) \in B \mid X)$$

and hence

$$P(T(t') \in B \mid T(t') \geq t, X) = P(T(\infty) \in B \mid T(\infty) \geq t, X).$$

Further, we have

$$\begin{aligned}
& P(T \in B \mid T \geq t, X, Z = t') = \\
= \quad & P(T \in B \mid T \geq t, X, Z = t', S(t) = t) P(S(t) = t \mid T \geq t, X, Z = t') + \\
+ \quad & P(T \in B \mid T \geq t, X, Z = t', S(t) = +\infty) P(S(t) = +\infty \mid T \geq t, X, Z = t') = \\
\overset{\text{consistency}}{=} \quad & P(T(t') \in B \mid T(t') \geq t, X, Z = t', S(t) = t) P(S(t) = t \mid T \geq t, X, Z = t') + \\
+ \quad & P(T(\infty) \in B \mid T(\infty) \geq t, X, Z = t', S(t) = +\infty) P(S(t) = +\infty \mid T \geq t, X, Z = t') = \\
\overset{\text{no anticipation}}{=} \quad & P(T(\infty) \in B \mid T(\infty) \geq t, X, Z = t', S(t) = t) P(S(t) = t \mid T \geq t, X, Z = t') + \\
+ \quad & P(T(\infty) \in B \mid T(\infty) \geq t, X, Z = t', S(t) = +\infty) P(S(t) = +\infty \mid T \geq t, X, Z = t') = \\
\overset{2.9}{=} \quad & P(T(\infty) \in B \mid T(\infty) \geq t, X, Z = t', S(t) = +\infty) P(S(t) = t \mid T \geq t, X, Z = t') + \\
+ \quad & P(T(\infty) \in B \mid T(\infty) \geq t, X, Z = t', S(t) = +\infty) P(S(t) = +\infty \mid T \geq t, X, Z = t') = \\
= \quad & P(T(\infty) \in B \mid T(\infty) \geq t, X, Z = t', S(t) = +\infty).
\end{aligned}$$

□

5.4 Proof of proposition 3.1

Let A be some borel set. Denote $Q := \{S(0,z) = 0, S(1,z) = 1\}$. Then it holds

$$\begin{aligned}
& P(V \in A \mid T(1,z) \geq t, X, S(0,z) = 0, S(1,z) = 1, Z = z) = \\
= \quad & P(V \in A \mid T(1,z) \geq t, X, Q, Z = z) = \\
= \quad & \frac{P(T(1,z) \geq t \mid V \in A, X, Q, Z = z) P(V \in A \mid X, Q, Z = z)}{P(T(1,z) \geq t \mid X, Q, Z = z)} = \\
\overset{\text{no anticipation}}{=} \quad & \frac{P(T(0,z) \geq t \mid V \in A, X, Q, Z = z) P(V \in A \mid X, Q, Z = z)}{P(T(0,z) \geq t \mid X, Q, Z = z)} = \\
= \quad & P(V \in A \mid T(0,z) \geq t, X, S(0,z) = 0, S(1,z) = 1, Z = z).
\end{aligned}$$

□

References

[Abadie, 2002] Abadie, A. (2002). Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American Statistical Association*.

[Abbring, 2003] Abbring, J. H. (2003). Dynamic econometric program evaluation. *IZA Discussion Paper No. 804*.

[Abbring and van den Berg, 2005] Abbring, J. H. and van den Berg, G. J. (2005). Social experiments and instrumental variables with duration outcomes. *Working Paper*.

[Bijwaard, 2006] Bijwaard, G. (2006). Instrumental variable estimation for duration data. *Tinbergen Institute Discussion Paper*.

[Bijwaard and Ridder, 2005] Bijwaard, G. and Ridder, G. (2005). Correcting for selective compliance in a re-employment bonus experiment. *Journal of Econometrics*.

[Chesher, 2003] Chesher, A. (2003). Semiparametric identification in duration models. *Working paper, UCL, London*.

[Cox, 1962] Cox, D. (1962). *Renewal Theory*. Methuen, London.

[Eberwein et al., 1997] Eberwein, C., Ham, J. C., and LaLonde, R. J. (1997). The impact of being offered and receiving classroom training on the employment histories of disadvantaged women: evidence from experimental data. *Review of Economic Studies*.

[Frlich, 2007] Frlich, M. (2007). Regression discontinuity design with covariates. *Working Paper, IZA, Bonn*.

[Hahn et al., 2001] Hahn, J., Todd, P., and van der Klaauw, W. (2001). Regression-discontinuity design. *Econometrica*.

[Ham and LaLonde, 1996] Ham, J. C. and LaLonde, R. J. (1996). The effect of sample selection and initial conditions in duration models: Evidence from experimental data on training. *Econometrica*.

[Heckman et al.,] Heckman, J. J., LaLonde, R. J., and Smith, J. A. *Handbook of Labor Economics, Volume III*.

[Imbens and Angrist, 1994] Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. *Econometrica*.

[Imbens and Rubin, 1997] Imbens, G. W. and Rubin, D. B. (1997). Estimating outcome distributions for compliers in instrumental variables models. *Review of Economic Studies*.

[Meyer, 1996] Meyer, B. D. (1996). What have we learned from the illinois reemployment bonus experiment? *Journal of Labour Economics*.

[Nadas, 1970] Nadas, A. (1970). On estimating the distribution of a random vector when only the smallest coordinate is observable. *Technometrics, Vol 12, No 4.*

[Pearl, 2000] Pearl, J. (2000). *Causality. Models, reasoning, and inference*. Cambridge University Press.

[Porter, 2003] Porter, J. (2003). Estimation in the regression discontinuity model. *Working Paper, Harvard*.

[Robins and Tsiatis, 1991] Robins, J. M. and Tsiatis, A. A. (1991). Correcting for selective compliance in a re-employment bonus experiment. *Communications in Statistics - Theory and Methods*.

[Tsiatis, 1975] Tsiatis, G. (1975). A nonidentifiability aspect of the problem of competing risks. *Proc. Nat. Acad. Sci. 72*.

[Van den Berg, 2001] Van den Berg, G. J. (2001). Duration models: specification, identification, and multiple durations. *Handbook of Econometrics, Volume V*.

[Van den Berg et al., 2010] Van den Berg, G. J., Bozio, A., and Dias, M. C. (2010). Policy discontinuity and duration outcomes. *Working Paper*.