Social Experiments and Instrumental Variables with Duration Outcomes

Jaap H. Abbring[1]
Gerard J. van den Berg[1,2]

[1] *Department of Economics, Vrije Universiteit Amsterdam, and Tinbergen Institute,*
[2] *IFAU-Uppsala, IZA, IFS London, and CEPR.*

Social Experiments and Instrumental Variables with Duration Outcomes

Jaap H. Abbring [*]
Gerard J. van den Berg [†]

May 3, 2005

Abstract

This paper examines the empirical analysis of treatment effects on duration outcomes from data that contain instrumental variation. We focus on social experiments in which an intention to treat is randomized and compliance may be imperfect. We distinguish between cases where the treatment starts at the moment of randomization and cases where it starts at a later point in time. We derive exclusion restrictions under various informational and behavioral assumptions and we analyze identifiability under these restrictions. It turns out that randomization (and by implication, instrumental variation) by itself is often insufficient for inference on interesting effects, and needs to be augmented by a semi-parametric structure. We develop corresponding non- and semi-parametric tests and estimation methods.

1 Introduction

Social experiments have become important tools for policy evaluation in general, and the evaluation of active labor market policies in particular (see Heckman, LaLonde and Smith, 1999, for a survey). Until recently, it was generally thought that policy evaluation using social experiments merely involved trivial comparisons of outcomes between the various treatment statuses assigned in the experiment. The studies of Ham and LaLonde (1996), Meyer (1996), and Eberwein, Ham and LaLonde (1997) show that complications arise if the outcome variable of interest is a duration variable (e.g. unemployment duration) or depends on the realization of a duration variable (e.g. the post-unemployment wage or the subsequent employment duration). The evaluation based on outcomes among those who survive up to a certain point is confounded by dynamic selection effects even if initial treatment assignment is randomized. Further problems arise if compliance to the treatment status assigned in the experiment (the intention to treat, or ITT) is imperfect and if actual program enrollment is a time-consuming process. In this paper, we explore the use of social experiments in duration and event-history analysis. We consider the randomized ITT as an instrumental variable (IV), and more generally consider the use of instrumental variables in duration analysis.

To motivate the paper and outline its contributions, it is useful to give (in the next paragraphs) a taxonomy of different cases that may arise in practice. This corresponds to the way the paper is organized. In all cases we allow the population under study to be heterogeneous. Throughout the paper, we use the terms "randomization" and "random assignment" to denote situations in which an ITT is assigned independently of agents' individual characteristics. An agent is said to "comply" with the ITT outcome if the actual treatment status coincides with it.[1] In every case, the policy setting determines what is meant with "treatment".

In Section 2, we start with the benchmark case where (*i*) randomization of ITT occurs at time 0 (which is usually the moment of inflow into the state of which the subsequent time spent in it is the outcome of interest), (*ii*) agents are immediately subject to the treatment, and (*iii*) there is perfect compliance. We briefly review the main insights from the literature and we discuss how to define and estimate average treatment effects.

Subsequently, as the second main case, we relax in Section 3 the perfect com-

---

[1]In the literature, compliance is often only used for agents who are assigned to be treated instead of being in the control group, but here we also use it for those who are assigned to the control group, unless stated otherwise.

pliance assumption (*iii*) but maintain assumptions (*i*) and (*ii*). This case arises e.g. if there is randomization of program offers (ITT) at time 0 and agents choose immediately whether to accept the offer and enroll. In such cases, ITT can be used as an IV for, but does not coincide with, actual treatment. In fact, researchers often restrict analyze the effect of ITT instead of the actual treatment. Methodologically, such an analysis fits our first main case. Moreover, under the appropriate IV conditions, the null of no treatment effects and the null of no ITT effects are equivalent (e.g. Robins and Tsiatis, 1991, and Abadie, 2002). Two advantages of an ITT-analysis are often cited (see e.g. Robins and Tsiatis, 1991). First, randomization can be ensured even if compliance is imperfect and selective. Second, ITT rather than actual treatment is the relevant public-policy instrument if the policy maker cannot control compliance any better than the analyst.[2] On the other hand, an IV analysis allows to some extent to disentangle the effects of compliance (participation) and the effects of actual treatment. If enough structure is imposed, the results of an IV analysis can be extrapolated beyond the scope of the experiment. Therefore, and because an ITT-analysis is already covered by our first case, we consider the effects of actual treatment using ITT as an IV. We first develop and analyze a non-parametric IV estimator that, unlike existing estimators, allows for censoring. We then provide novel semi-parametric identification results.

In the third main case, which is examined in Section 4, both assumptions (*ii*) and (*iii*) are violated, i.e. randomization of ITT occurs at time 0, the actual treatment occurs later, and compliance is imperfect. Again, ITT can be used as an instrument (see e.g. Eberwein, Ham and LaLonde, 1997). However, as argued by e.g. Abbring and Van den Berg (2003b), it is often hard to justify exclusion restrictions in a dynamic setting with forward-looking agents. In particular, if ITT affects the treatment, as it should if it is to be a valid instrument, then it is likely to affect the outcome hazard up to the moment of treatment enrollment as well. To proceed, it is useful to distinguish between such ex ante effects on the outcome hazard and the ex post effects on the hazard after treatment enrollment. From an economic and policy point of view it is often interesting to know the magnitude of both of these effects. We give some examples; one on training programs for unemployed workers, and one on vouchers for public housing tenants. We introduce a weaker-than-usual IV assumption that only requires "ex post exclusion", meaning exclusion of ITT from the outcome hazard after treatment enrollment. We discuss informational and behavioral assumptions under which

---

[2]For example, the threat of a punishment treatment may be considered as a treatment itself (Abbring, Van den Berg and Van Ours, 2005, and Black et al., 2003).

this weaker IV assumption holds, and discuss its implications for identifiability. We show that, even though ITT is randomly assigned, "ex ante" effects of ITT on treatment assignment and on outcome hazards before treatment enrollment cannot be identified, and can typically not even be signed. Moreover, the effects of actual treatment are poorly identified in this case. More constructively, we argue that the observable relation between ITT and the outcome hazard after treatment enrollment is informative on selection effects. After all, under ex post exclusion such a relation cannot be causal. A selectivity test based on this idea bears an analogy to tests on cohort effects in mortality rates (Vaupel and Yashin, 1985) and on unobserved heterogeneity in duration models with time-varying explanatory variables (Van den Berg and Van Ours, 1997).

All results also apply to cases where there is no deliberate randomization at time 0 but the data contain appropriate instrumental variation (e.g., because of a natural experiment). Also, our results carry over to more complex, dynamic experimental designs than considered here.[3] In all cases, we follow non-parametric as well as semi-parametric approaches. We do not restrict attention to effects on survival probabilities but, in line with duration analysis in general, we focus on the effects on the hazard rate of the outcome duration variable, because of the intimate link between the hazard rate and economic behavioral models (Van den Berg, 2001). Knowing the effect on individual behavior as reflected in the hazard rate and the way this changes over time enables one to learn something about the reasons for the effectiveness or ineffectiveness of a policy, and this allows one to extrapolate the experimental results to different policies and policies in different environments. For example, consider the third case above, where the empirical distinction between "ex ante" and "ex post" effects calls for a hazard rate analysis. We return to this below.

Various papers consider methods for IV analysis of distributional treatment effects that apply to some extent to the problems studied in this paper. The non-parametric approach for effects on survival probabilities in our second main case is closely related to Imbens and Rubin (1997) and Abadie (2002), who discuss identification, estimation, and testing of distributional treatment effects in a non-parametric setting.[4] Their methods, however, do not handle censoring. In

_____

[3]Section 5 provides some discussion. Further generalizations concern cases where treatment and instrument may change over time at the same discrete (deterministic or random) points of times. This avoids the substantive problems discussed in Abbring and Van den Berg (2003b), but is unlikely to have many applications. The methodology for such cases has been discussed in the context of transformation models by Bijwaard (2003). We do not consider such cases.

[4]Also, Abadie, Angrist and Imbens (2002) develop and apply a semi-parametric IV estimator of quantile treatment effects. This estimator does not allow for censoring. It is moreover based

3

contrast, our focus is on methods that allow for right-censoring. Duration-model IV estimators that allow for censoring have been developed by Robins and Tsiatis (1991), Bijwaard and Ridder (2005), Bijwaard (2003), and Chesher (2003). These estimators apply to our second and/or third main cases. They require (semi-)parametric structure and additional substantive assumptions like perfect compliance in the control group (Bijwaard and Ridder, 2005). With some exceptions, they do not focus on effects on the individual hazard rate. We discuss these estimators where appropriate.

Throughout the paper, we consider experiments at face value and we do not address their external validity. Specifically, we do not address endogenous selection into the experimental population, differences between the experiment and the permanent imposition of a policy, and equilibrium effects in general. See Heckman, LaLonde and Smith (1999) for a detailed discussion, Ferrall (2002) for a comprehensive dynamic economic framework, and Van den Berg and Van der Klaauw (2005) for an empirical illustration in a reduced-form duration-analysis framework.

The remainder of the paper is organized as follows. Sections 2, 3, and 4 discuss our three main cases. Section 5 concludes.

2   Randomization with perfect compliance

2.1   Potential outcomes, individual treatment effects, and available data

We consider the population of agents or individuals flowing into a state of interest, and the durations these individuals subsequently spend in that state. Upon inflow, an individual is assigned to a treatment s from a set $\mathcal{S}$. In this section, we assume that the individual complies with the assigned treatment, so that ITT and actual treatment coincide. We do not need to specify the length of the time span during which a treatment takes place. For example, the length can be zero (if the time span equals the singular point $\{0\}$) or it can be infinite. We are interested in the causal effect of the treatment on the duration spent in the state of interest (the "outcome").

We model this effect using the potential-outcome framework pioneered by Neyman (1923). To each treatment $s \in \mathcal{S}$ corresponds a random variable $T(s)$, the potential outcome duration that would prevail if we would intervene and

---

on quantile models that are designed for outcomes like earnings and that do not apply naturally to duration and event-history outcomes.

assign treatment s. We assume that $\{T(s)\} := \{T(s); s \in \mathcal{S}\}$ is a measurable stochastic process.[5] Moreover, for ease of exposition we assume that each $T(s)$ is continuous, and we denote the hazard rate of $T(s)$ by $\theta_{T(s)}$. We are interested in contrasts between the distributions of $T(s)$ and $T(s')$ corresponding to treatments $s, s' \in \mathcal{S}$. These contrasts are summarized in so-called treatment effects. We focus on two of these,

$$\frac{\theta_{T(s')}(t)}{\theta_{T(s)}(t)} \qquad \text{and} \qquad \Pr(T(s') > t) - \Pr(T(s) > t), \qquad (1)$$

which are, respectively, the relative effect on the hazard rate at t and the additive effect on the survival probability at t of replacing one treatment s by another treatment s', as functions of t. The former captures the effect on the most interesting feature of the duration distribution. The latter, as we shall see, can easily be related to the standard literature on treatment evaluation. For the time being we consider a single subject, or, equivalently, a homogeneous group of subjects. We may of course consider a whole range of alternative treatment effects, like $\Pr(T(s') > t)/\Pr(T(s) > t)$ and $\mathbb{E}[T(s')]/\mathbb{E}[T(s)]$. In fact, the relative effect on the survival probability may often be more interesting than the additive effect. However, as we shall see in the next subsection, heterogeneity across subjects causes related methodological difficulties for all average treatment effects except for the average additive effect on the survival probability and the effects derived from it. The two special effects in (1) therefore cover the whole range of treatment effects from a methodological point of view.

The treatment is assigned according to a $\mathcal{S}$-valued random variable S. Throughout this section, we assume that treatment assignment is randomized, i.e.[6]

Assumption 1 (Randomization). $S \perp\!\!\!\perp \{T(s)\}$.

The actual outcome duration is $T := T(S)$; all other potential outcomes are counterfactual.

We allow for random right-censoring.[7] To this end, define a random censoring time C that is independent of (T, S). Our data are derived from the

---

[5]This process, viewed as a random function $s \mapsto T_s$, can alternatively be interpreted as a non-parametric structural equation for the determination of the outcome by the treatment.

[6]More generally, randomization could be conditional on observed covariates, or even observed external covariate processes. Throughout much of this paper, we ignore observed covariates. If appropriate, it is implicitly understood that results hold conditional on covariates. In the case of discrete covariates, all empirical methods can be directly applied to strata defined by the covariates.

[7]The censoring mechanism specified here is usually referred to as "simple random right-censoring". Extensions to more general forms of independent censoring and filtering are straightforward (see Andersen et al., 1993, and Fleming and Harrington, 1991).

full-information random sample $((T_1, S_1, C_1), \ldots, (T_n, S_n, C_n))$ from the distribution of (T, S, C). Suppose that observation i is censored if $T_i > C_i$, and complete if $T_i \leq C_i$. Then our data are the limited-information random sample $((\widetilde{T}_1, S_1, D_1), \ldots, (\widetilde{T}_n, S_n, D_n))$, where D_i is defined as $D_i := I(T_i \leq C_i)$, so $D_i = 1$ indicates a complete observation, and

$$\widetilde{T} = \begin{cases} T_i & \text{if } T_i \leq C_i \\ C_i & \text{if } T_i > C_i \end{cases}$$

is the possibly censored outcome duration. In the sequel we do not consider treatment effects on moments of the outcome duration (like the mean) because typically the observation window is finite (i.e., observed durations are always right-censored if they exceed some finite time) and the moments are not non-parametrically identified from such data.

2.2 Heterogeneity, average treatment effects, and non-parametric inference

We only consider ex ante individual heterogeneity in outcomes that can be captured by time-invariant unobserved characteristics V, because this is sufficient to obtain the main insights. Randomization (Assumption 1) implies that $S \perp\!\!\!\perp V$. For ease of exposition, we take V to be a continuous random variable and S to be a binary indicator with realizations 1 ("treatment") and 0 ("control"), and we assume that the hazard rates $\theta_{T(s)}(t|V)$ satisfy the usual regularity conditions that guarantee existence of the expressions below.

Individual treatment effects are now defined in terms of the distributions of $T(0)|V$ and $T(1)|V$, whereas average treatment effects now concern averages over the relevant population, i.e. over the distribution of V in the relevant population. Thus, the individual treatment effects on the hazard rate and the survival probability at t (see (1)) are defined as $\theta_{T(1)}(t|V)/\theta_{T(0)}(t|V)$ and $\Pr(T(1) > t|V) - \Pr(T(0) > t|V)$, respectively.

The average additive treatment effect on the survival probability at t is naturally defined as $\mathbb{E}[\Pr(T(1) > t|V) - \Pr(T(0) > t|V)]$. This equals $\Pr(T(1) > t) - \Pr(T(0) > t)$. With randomization, as in Assumption 1, this in turn is equal to $\Pr(T > t|S = 1) - \Pr(T > t|S = 0)$. The two survivor functions in the latter can be estimated straightforwardly using non-parametric survival analysis taking account of right-censoring (like Kaplan-Meier estimation; see e.g. Andersen et al., 1993, and Fleming and Harrington, 1991). One can immediately derive uniform confidence bounds on the potential duration distributions and tests of hypotheses

like $\overline{F}_{T(1)} = \overline{F}_{T(0)}$ (see Andersen et al., 1993). One may also obtain point-wise results for isolated survival probabilities, e.g. to assess the effect of training on the probabilities of staying unemployed for 6 or 12 months.[8] These results are straightforwardly extended to general discrete s,[9] and to more general types of heterogeneity than considered here (e.g. V may vary over time).

Unfortunately, the above approach (of writing an average treatment effect as the difference of quantities that can be straightforwardly non-parametrically estimated) cannot be applied to any other treatment effect of interest. Consider the effect on the hazard rate. As indicated by Ham and LaLonde (1996) and Meyer (1996), if the treatment has a causal effect on the duration, then, typically, the distribution of V among the survivors at points in time $t > 0$ depends on the treatment, so $V \not\perp S | T > t$. In other words, there is no randomization at $t > 0$ despite the randomization at 0. To see this, let f, F, Θ and $\overline{F}$ be generic symbols for a density, a distribution function, an integrated hazard, and a survivor function, with subscripts denoting the corresponding random variable (note that $\overline{F} = 1 - F = e^{-\Theta}$). Further, denote the hazard and integrated hazard of T by θ_T and Θ_T. From e.g. Lancaster (1990),

$$f_V(v | T > t, S) = \frac{\overline{F}_T(t|S, V) f_V(v)}{\int_0^\infty \overline{F}_T(t|S, V) dF_V(v)}, \tag{2}$$

with $\overline{F}_T(t|S, V) = \exp\left[-\Theta_T(t|S, V)\right]$. It is not difficult to construct examples in which the distribution of V among the treated survivors at t is first-order stochastically dominated by the distribution of V among the non-treated survivors at t, in particular if there is a strong positive interaction between S and V in the hazard rate of T, and this hazard rate increases in V and S. Then the individual hazard rate at t is very large if both $S = 1$ and V is large, and as a result the survivors at t may contain relatively few treated individuals with a high V.

A first implication is that one has to carefully define the subpopulation over which an individual treatment effect on the hazard rate is averaged. Consider the

---

[8]Because of right-censoring, discrete-choice models like probit models are not a good alternative. Also, those models have difficulties handling time-varying explanatory variables. The dynamically assigned treatment of Section 4 of this paper provides an example of such an explanatory variable.

[9]The model is versatile in dealing with alternative timing patterns. In particular, the treatment space itself may include time dimensions. For example, a treatment in $\mathcal{S}$ may not only specify a particular program but also the time of enrolling into this program. Perfect compliance requires that agents comply with the full ITT, including the assigned timing of the program. We do not elaborate on this case. As noted in Section 1, perfect compliance is a strong assumption in this case.

average treatment effect ATE(t) defined by

$$\mathbb{E}[\theta_{T(1)}(t|V)/\theta_{T(0)}(t|V)]$$

This involves aggregation over V in the whole population. However, a hazard concerns a subpopulation of survivors at t, which is systematically different from the population, so instead one would like to take the average over V among survivors. But because the selectivity of survivors at $t > 0$ depends on the treatment status, one has to be specific about this status as well. We propose the following novel average treatment effects on the individual hazard rate,

$$\mathbb{E}\left[\frac{\theta_{T(1)}(t|V)}{\theta_{T(0)}(t|V)} \;\middle|\; T(1) \geq t\right], \qquad \mathbb{E}\left[\frac{\theta_{T(1)}(t|V)}{\theta_{T(0)}(t|V)} \;\middle|\; T(1) \geq t, T(0) \geq t\right]$$

which can be called the *Average Treatment effect on the Treated Survivors at t* (ATTS(t)) and the *Average Treatment effect on the Survivors at t* (ATS(t)). ATTS(t) averages over the distribution of V among the survivors at t if the agents are assigned to the treatment. Under randomization, this is equivalent to averaging over the distribution of V among the treated survivors at t (so with $T \geq t, S = 1$). ATS(t) averages over the distribution of V among individuals who are survivor at t under both possible treatments. It basically equals the *survivor average causal effect* of Rubin (2000) in case the latter would be applied to the duration outcome itself rather than to effects on non-duration outcomes. Note that in general both ATTS(t) and ATS(t) are properties of subpopulations whose composition depends on the treatment effect on $[0, t)$.

Non-parametric inference produces sample equivalents of $\theta_T(t|S = 1)/\theta_T(t|S = 0)$. There holds that

$$\frac{\theta_T(t|S = 1)}{\theta_T(t|S = 0)} = \frac{\mathbb{E}[\theta_T(t|S = 1, V) \mid T \geq t, S = 1]}{\mathbb{E}[\theta_T(t|S = 0, V) \mid T \geq t, S = 0]},$$

so this ratio reflects (i) the treatment effect and (ii) the selection effect that at $T = t$, among the survivors at t, the treated and controls have systematically different unobserved characteristics despite the randomization at $t = 0$. The non-parametric estimator therefore does not capture any meaningful treatment effect.

In fact, one can construct examples where

$$\theta_T(t|S = 1) < \theta_T(t|S = 0)$$

even if

$$\theta_{T(1)}(t|V) > \theta_{T(0)}(t|V) \quad \text{almost surely for all } t. \tag{3}$$

For example, let V have a discrete distribution with $\Pr(V = 0.2) = \Pr(V = 2.5) = 0.5$, and let $\Theta_{T(s)}(t|V) = \Theta^*_{T(s)}(t)V$ with $\Theta^*_{T(1)}(t) > \Theta^*_{T(0)}(t)$ for all $t > 0$ (note that this is weaker than inequality (3) for $\theta_{T(s)}(t|V)$ in this example). Then $\theta_T(t|S = 1) < \theta_T(t|S = 0)$ for values of $\Theta^*_{T(s)}(t)$ in an interval around 1. In such cases the dynamic selection effect on the observed hazard rate dominates the treatment effect, in certain time intervals. Obviously, this may lead to invalid non-parametric inference on the sign of the treatment effect.

Similar results can be derived for e.g. the use of the sample equivalent of $F_T(t + a|T > t, S = 1)/F_T(t + a|T > t, S = 0)$ for $a > 0$ (e.g. from a probit analysis of whether $T \in (t, t + a]$ given $T > t, S$) to estimate the corresponding average effect on the individual conditional survival probability.[10]

Ham and LaLonde (1996) consider inference of treatment effects on post-spell outcomes if randomization occurs at $t = 0$. Consider an outcome W that is realized immediately after T. Let $W(s)$ be the random outcome that would prevail if we hypothetically assign the treatment s to the agent. In general, $\{W(s)\}$ may depend on V. The observation window is denoted by $[0, \mathcal{T})$ with $\mathcal{T} < \infty$. We observe W iff $T < \mathcal{T}$. The distribution of $V|S = s$ among the agents for whom W is observed equals the distribution of $V|T < \mathcal{T}, S = s$. It follows that, among the observed W, the distribution of V among the treated in general differs from the distribution of V among the controls. So, despite randomization at $t = 0$, we cannot simply compare the observed mean outcomes of W among treated and controls in order to uncover e.g. $\mathbb{E}[W(1) - W(0)]$. We conclude that the problem with causal inference on post-duration outcomes arises because of a bounded observational window (i.e., right-censoring on a bounded interval). Without the latter, the former does not arise.

2.3 Semi-parametric approaches

To obtain more constructive results, we need to impose some semi-parametric structure on the distribution of $T(s)|V$. Two approaches can be taken, depending

---

[10]Average non-additive treatment effects on quantities that are defined for the whole population, such as the average of the relative effect on the survival probability defined in Sub-section 2.1, can also not be inferred non-parametrically, but this is only because of the non-additivity. Note that if $\Theta_{T(1)}(t|V) > \Theta_{T(0)}(t|V)$ for all $t > 0$ and at all V, then the observable quantity $\overline{F}_T(t|S = 1)/\overline{F}_T(t|S = 0) - 1$ is always negative (this follows from $\overline{F}_T(t|S = s) = \mathbb{E}[\exp(-\Theta_{T(s)}(t|V))])$ and vice versa, so then this observable quantity always has the same sign as the average treatment effect $\mathbb{E}[\exp(-\Theta_{T(1)}(t|V))/\exp(-\Theta_{T(0)}(t|V))] - 1$. A similar result holds for $\theta_T(0|S = 0) - \theta_T(0|S = 1)$. All these quantities are based on samples drawn at $t = 0$, at which there has been no dynamic selection yet.

on the amount of structure imposed. First, one may abandon point identification and estimation, and focus on bounding the parameters of interest (e.g. Manski, 1997). Second, one may impose a structure that is sufficiently specific to enable point identification and estimation of relevant treatment effects.

We start by assuming

Assumption 2 (Multiplicative unobserved heterogeneity).

$$\theta_{T(s)}(t|V) = \theta^*_{T(s)}(t)V \tag{4}$$

for all t, s, with $\Theta^*_{T(s)}(t)$ defined analogously. This imposes that the individual characteristics V affect the counterfactual hazard rates in the same proportional way. It implies that the individual treatment effect on the hazard at t equals $\theta^*_{T(1)}(t)/\theta^*_{T(0)}(t)$, so it is homogeneous across individuals (but not necessarily over time). Indeed, the individual effect at t equals ATE(t), ATTS(t) and ATS(t).

Moreover, let the treatment effect be monotone in the following way: the individual-level potential-outcome distributions can be ranked in terms of first-order stochastic dominance, i.e. $\Theta^*_{T(1)}(t) > \Theta^*_{T(0)}(t)$ for all $t > 0$, or $\Theta^*_{T(1)}(t) = \Theta^*_{T(0)}(t)$ for all $t > 0$, or $\Theta^*_{T(1)}(t) < \Theta^*_{T(0)}(t)$ for all $t > 0$.

Along the lines of Meyer (1996) it can be shown directly that

$$\frac{\theta_T(t|S=1)}{\theta_T(t|S=0)} < \text{ATE}(t) = \text{ATTS}(t) = \text{ATS}(t). \tag{5}$$

if $\Theta^*_{T(1)}(t) > \Theta^*_{T(0)}(t)$, for all $t > 0$. The results for the other two rankings follow straightforwardly. The ranking of the individual-level potential-outcome distributions can be inferred from the ranking of $F_{T;S=1}(t)$ and $F_{T;S=0}(t)$ near $t = 0$, so the average treatment effects on the hazard rate can be bounded by the observable left-hand side of equation (5).

Imposition of (4) is insufficient for point identification. This can be achieved by the additional restriction that the hazard rate of $T(s)|V$ is proportional in t and s as well,

$$\theta_{T(s)}(t) = \gamma^s \, \lambda(t) \, V$$

for some scalar treatment effect parameter $\gamma > 0$ and some "baseline hazard" function λ. This results in the so-called MPH model for $T|S, V$ with $\theta_T(t|S, V) = \lambda(t) \, \gamma^S \, V$. Under the additional assumption that $\mathbb{E}[V] < \infty$ the parameter γ is identified from (see Elbers and Ridder, 1982, and Kortram et al., 1995),[11]

---

[11]Intuitively, the non-parametric hazard rates at zero are not yet affected by dynamic selection on survival, and semi-parametric identification involves an extrapolation of the treatment effect on the hazard at zero to positive durations. This does of course not mean that semi-parametric estimation results are completely driven by extremely short durations.

$$\gamma = \lim_{t \downarrow 0} \frac{F_T(t|S=1)}{F_T(t|S=0)}. \tag{6}$$

In applications of MPH models, the "random effects" assumption that observed and unobserved explanatory variables are independent is often controversial. However, here this assumption follows from the randomization of S.

Other semi-parametric structures of the hazard rate of $T(s)|V$ may also lead to point identification and estimation of treatment effects (see Heckman and Taber, 1994, and Van den Berg, 2001, for surveys). Concerning treatment effects of post-duration outcomes in case of a finite observational window, a solution is to estimate a model that takes unobserved heterogeneity into account.

3 Randomization and immediate treatment with imperfect compliance

3.1 Model and data

We now relax Section 2's assumption of perfect compliance, but retain the assumption that treatment is immediate. Instead of being randomly assigned to a treatment as in Assumption 1, agents are randomly assigned a label $Z \in \mathcal{Z}$ at time 0. In the typical experimental setup, Z takes values in the set $\mathcal{Z} = \mathcal{S}$ of possible treatments and is interpreted as the ITT with the treatment Z. More generally, we allow Z to be an IV with support $\mathcal{Z} \neq \mathcal{S}$. We will nevertheless refer to Z as the ITT throughout.

Let $S(z)$ be the random actual treatment that would prevail if we hypothetically assign the agent the label $z \in \mathcal{Z}$. The process $\{S(z)\}$, which is assumed measurable, can again be given a structural interpretation. The actual treatment is $S = S(Z)$. In the ITT case with perfect compliance, we would have that $S(z) = z$ for all $z \in \mathcal{Z} = \mathcal{S}$, and $S = S(Z) = Z$. In the present section, however, we allow for imperfect compliance (so $S = S(Z) \neq Z$) by allowing $S(z)$ to be non-degenerate. If Z is discrete and an effective ITT, it should be the case that $\Pr(S(z) = z) > \Pr(S(z) = z')$ for all $z, z' \in \mathcal{S}$ such that $z' \neq z$.

Throughout this section we maintain

Assumption 3 (IV). *(i)* $z \mapsto S(z)$ is non-trivial (in a way that will be further specified in special cases below), and *(ii)* $Z \perp\!\!\!\perp (\{T(s)\}, \{S(z)\})$.

Underlying Assumption 3(ii) is the notion that (a) the ITT does not causally affect outcomes directly, so that potential outcomes $T(s)$ need not be indexed by z, (b) outcomes do not causally affect treatment, so that $S(z)$ need not be indexed by possible values t of T, and (c) the ITT is not causally affected by either outcomes or treatment (see Abbring, 2003, for discussion).

With imperfect compliance, the actual treatment $S = S(Z)$ and the potential outcomes $\{T(s)\}$ are typically dependent, because agents non-experimentally self-select or are selected in actual treatment. This sets the present analysis apart from that of Section 2. Note, however, that a reduced-form analysis of the effect of ITT on outcomes fits Section 2's framework. Formally, the outcome equation can be reduced to $\{T(S(z))\}$. Under Assumption 3, $Z \perp\!\!\!\perp \{T(S(z))\}$, so that Assumption 1 holds for the reduced-form model with Z replacing S.

The data of Subsection 2.1 are accordingly enriched with instrumental variation. We now have a random sample $((\widetilde{T}_1, S_1, D_1, Z_1), \ldots, (\widetilde{T}_n, S_n, D_n, Z_n))$ from the joint distribution of $(\widetilde{T}, S, D, Z)$.

3.2 Non-parametric estimation and testing with IV

For now, consider the binary treatment-binary instrument case that $\mathcal{S} = \{0, 1\}$ and $\mathcal{Z} = \{0, 1\}$. Again, it is convenient to think of the treatment statuses as "treatment" ($s = 1$) and "control" ($s = 0$) in this case. Define $p(z) := \Pr(S(z) = 1)$. Note that $p(z) = \Pr(S = 1 | Z = z)$ under Assumption 3. Assume that Imbens and Angrist's (1994) monotonicity property holds, that is $S(0) \leq S(1)$ or $S(0) \geq S(1)$. Without further loss of generality, we take $S(0) \leq S(1)$. Then, the subpopulation that switches treatment status between propensities $p(0)$ and $p(1)$ all switch from treatment 0 to treatment 1. This subpopulation, called "compliers" by Imbens and Rubin (1997), is therefore $Q := \{S(0) = 0, S(1) = 1\}$, and has probability mass $\Pr(Q) = p(1) - p(0)$. We formalize Subsection 3.1's Assumption 3(i) by assuming that $\Pr(Q) > 0$.

In this subsection we only consider the average additive treatment effect on the survival probabilities, since from Section 2 we know that non-parametric inference of other effects is not feasible. Existing methods in the literature would focus on various average contrasts between $T(1)$ and $T(0)$ on Q, and the marginal distributions $F_{0;Q}$ and $F_{1;Q}$ of, respectively, $T(0)$ and $T(1)$ on Q. However, we have to adapt them to allow for censoring.

Identification of the marginal potential-outcome distributions $F_{0;Q}$ and $F_{1;Q}$

on Q is straightforward (Imbens and Rubin, 1997). In particular,

$$\overline{F}_{0;Q}(t) = \frac{\Pr\left(T > t, S = 0|Z = 0\right) - \Pr\left(T > t, S = 0|Z = 1\right)}{p(1) - p(0)} \quad \text{and}$$

$$\overline{F}_{1;Q}(t) = \frac{\Pr\left(T > t, S = 1|Z = 1\right) - \Pr\left(T > t, S = 1|Z = 0\right)}{p(1) - p(0)}.$$

By implication, the mean survival probability contrast on Q

$$\Delta_Q(t) := \overline{F}_{1;Q}(t) - \overline{F}_{0;Q}(t),$$

is identified (Imbens and Angrist, 1994). This is a local average treatment effect on survival for at least t periods.

Next, for testing, we can use that

$$\Delta_Q = \frac{\overline{F}_{T;Z=1} - \overline{F}_{T;Z=0}}{p(1) - p(0)},$$

where $\overline{F}_{T;Z=z}(t) := \Pr(T > t|Z = z)$. This equality implies that the causal null that $\overline{F}_0 = \overline{F}_1$ is equivalent to the reduced-form null that $\overline{F}_{T;Z=0} = \overline{F}_{T;Z=1}$ (Robins and Tsiatis, 1991, and Abadie, 2002). Thus, under the IV assumptions, we can test for distributional treatment effects using any of the non-parametric tests of Section 2.

Next, consider estimating $\overline{F}_{0;Q}$, $\overline{F}_{1;Q}$, and Δ_Q. First note that $\overline{F}_{T;Z=z}(t)$ and $\overline{F}_{T;S=s,Z=z}(t) := \Pr(T > t|S = s, Z = z)$ can be estimated by the Kaplan-Meier estimator. Also, $p(z)$ can be estimated by

$$\widehat{p}(z) = \frac{\sum_{i=1}^{n} S_i I(Z_i = z)}{\sum_{i=1}^{n} I(Z_i = z)}.$$

Thus, $\overline{F}_0$ and $\overline{F}_1$ can be estimated by

$$\widehat{\overline{F}}_{0;Q} = \frac{[1 - \widehat{p}(0)]\widehat{\overline{F}}_{T;S=0,Z=0} - [1 - \widehat{p}(1)]\widehat{\overline{F}}_{T;S=0,Z=1}}{\widehat{p}(1) - \widehat{p}(0)} \quad \text{and} \tag{7}$$

$$\widehat{\overline{F}}_{1;Q} = \frac{\widehat{p}(1)\widehat{\overline{F}}_{T;S=1,Z=1} - \widehat{p}(0)\widehat{\overline{F}}_{T;S=1,Z=0}}{\widehat{p}(1) - \widehat{p}(0)}, \tag{8}$$

respectively, and Δ_Q can be estimated by

$$\widehat{\Delta}_Q = \widehat{\overline{F}}_{1;Q} - \widehat{\overline{F}}_{0;Q} = \frac{\widehat{\overline{F}}_{T;Z=1} - \widehat{\overline{F}}_{T;Z=0}}{\widehat{p}(1) - \widehat{p}(0)}.$$

Asymptotic behavior of these estimators follows from standard results for the Kaplan-Meier estimator. Let $q_z := \Pr(Z = z)$. We have

Proposition 1. *Under Assumption 3,* $\sqrt{n}\left(\widehat{\overline{F}}_{1;Q} - \overline{F}_{1;Q}, \widehat{\overline{F}}_{0;Q} - \overline{F}_{0;Q}\right) \xrightarrow{\mathcal{D}}$

$$\frac{1}{p(1) - p(0)}\bigg(\left(\overline{F}_{T;S=1,Z=1} - \overline{F}_{1;Q}\right)\mathcal{N}_1 - \left(\overline{F}_{T;S=1,Z=0} - \overline{F}_{1;Q}\right)\mathcal{N}_0$$

$$- \sqrt{\frac{p(1)}{q_1}}\overline{F}_{T;S=1,Z=1}\mathbb{G}_{11} + \sqrt{\frac{p(0)}{q_0}}\overline{F}_{T;S=1,Z=0}\mathbb{G}_{10},$$

$$\left(\overline{F}_{T;S=0,Z=1} - \overline{F}_{0;Q}\right)\mathcal{N}_1 - \left(\overline{F}_{T;S=0,Z=0} - \overline{F}_{0;Q}\right)\mathcal{N}_0$$

$$+ \sqrt{\frac{1-p(1)}{q_1}}\overline{F}_{T;S=0,Z=1}\mathbb{G}_{01} - \sqrt{\frac{1-p(0)}{q_0}}\overline{F}_{T;S=0,Z=0}\mathbb{G}_{00}\bigg),$$

with $\mathcal{N}_z$ $(z = 0, 1)$ *a normal random variable with zero mean and variance* $p(z)\left[1 - p(z)\right]/q_z$, $\mathbb{G}_{sz}$ $(s, z = 0, 1)$ *a Gaussian martingale such that* $G_{sz}(0) = 0$ *and, for* $t < t'$,

$$\text{cov}\left(\mathbb{G}_{sz}(t), \mathbb{G}_{sz}(t')\right) = \int_0^t \frac{\theta_{T;S=s,Z=z}\left(\tau\right)}{\overline{F}_{T;S=s,Z=z}\left(\tau\right)\overline{F}_C\left(\tau^-\right)}d\tau =: \sigma_{sz}^2\left(t\right),$$

and $\mathcal{N}_1$, $\mathcal{N}_0$, $\mathbb{G}_{11}$, $\mathbb{G}_{01}$, $\mathbb{G}_{10}$, *and* $\mathbb{G}_{00}$ *mutually independent.*

Proof. See Appendix 1. $\qquad\square$

Proposition 1 can be used to compute asymptotic standard errors of $\widehat{F}_{0,Q}(t)$, $\widehat{F}_{1;Q}(t)$, and $\widehat{\Delta}_Q(t)$, and (point-wise and uniform) confidence bounds on $F_{0,Q}$, $F_{1;Q}$, and Δ_Q. In particular, for $\widehat{\Delta}_Q$ we have

Corollary 1. *For* $t \leq t'$, *the asymptotic covariance of* $\widehat{\Delta}_Q(t)$ *and* $\widehat{\Delta}_Q(t')$ *equals*

$$\frac{1}{[p(1) - p(0)]^2}\bigg\{ \frac{p(1)\left[1 - p(1)\right]}{nq_1}\left[\overline{F}_{T;S=1,Z=1}(t) - \overline{F}_{T;S=0,Z=1}(t) - \Delta_Q(t)\right]$$

$$\times \left[\overline{F}_{T;S=1,Z=1}(t') - \overline{F}_{T;S=0,Z=1}(t') - \Delta_Q(t')\right]$$

$$+ \frac{p(0)\left[1 - p(0)\right]}{nq_0}\left[\overline{F}_{T;S=1,Z=0}(t) - \overline{F}_{T;S=0,Z=0}(t) - \Delta_Q(t)\right]$$

$$\times \left[\overline{F}_{T;S=1,Z=0}(t') - \overline{F}_{T;S=0,Z=0}(t') - \Delta_Q(t')\right]$$

$$+ \frac{p(1)\overline{F}_{T;S=1,Z=1}(t)\overline{F}_{T;S=1,Z=1}(t')\sigma_{11}^2(t)}{nq_1}$$

$$+ \frac{\left[1 - p(1)\right]\overline{F}_{T;S=0,Z=1}(t)\overline{F}_{T;S=0,Z=1}(t')\sigma_{01}^2(t)}{nq_1}$$

$$+ \frac{p(0)\overline{F}_{T;S=1,Z=0}(t)\overline{F}_{T;S=1,Z=0}(t')\sigma_{10}^2(t)}{nq_0}$$

$$+ \frac{\left[1 - p(0)\right]\overline{F}_{T;S=0,Z=0}(t)\overline{F}_{T;S=0,Z=0}(t')\sigma_{00}^2(t)}{nq_0}\bigg\}.$$

14

$$(9)$$

A consistent estimator of this asymptotic covariance follows by plugging the estimators $\widehat{p}(z)$ of $p(z)$, $M_z := \sum_{i=1}^{n} I(Z_i = z)$ of nq_z, $\widehat{\overline{F}}_{T;S=s,Z=z}$ of $\overline{F}_{T;S=s,Z=z}$, $\widehat{\Delta}_{s;Q}$ of $\Delta_{s;Q}$, and consistent estimators

$$\widehat{\sigma}_{sz}^2(t) = \sum_{\{j:D_j=1,\widetilde{T}_j \leq t, S_j=s, Z_j=z\}} \left[R_{sz}(\widetilde{T}_j)\right]^{-2}$$

of $\sigma_{sz}(t)$ into equation (9). Note that Corollary 1 provides the asymptotic variance of $\widehat{\Delta}_Q(t)$, and therefore its asymptotic standard error, for $t = t'$.

One special case deserves some attention. In the case that there is no censoring, $\widehat{\Delta}_Q(t)$ reduces to Imbens and Angrist's (1994) IV estimator of the local average treatment effect on the binary outcome $I(T > t)$, the Wald estimator[12]

$$\widehat{\Delta}_Q(t) = \frac{M_1^{-1}\sum_{i=1}^{n} Z_i I(T_i > t) - M_0^{-1}\sum_{i=1}^{n}(1 - Z_i)I(T_i > t)}{M_1^{-1}\sum_{i=1}^{n} Z_i S_i - M_0^{-1}\sum_{i=1}^{n}(1 - Z_i)S_i}.$$

In this case, Proposition 1 implies

Corollary 2. *If $\overline{F}_C = 1$ (no censoring), then, for $t \leq t'$, the asymptotic covariance of $\widehat{\Delta}_Q(t)$ and $\widehat{\Delta}_Q(t')$ equals*

$$\frac{1}{[p(1) - p(0)]^2}\Bigg\{ \frac{p(1)\left[1 - p(1)\right]}{nq_1} \left[\overline{F}_{T;S=1,Z=1}(t) - \overline{F}_{T;S=0,Z=1}(t) - \Delta_Q(t)\right]$$
$$\times \left[\overline{F}_{T;S=1,Z=1}(t') - \overline{F}_{T;S=0,Z=1}(t') - \Delta_Q(t')\right]$$
$$+ \frac{p(0)\left[1 - p(0)\right]}{nq_0} \left[\overline{F}_{T;S=1,Z=0}(t) - \overline{F}_{T;S=0,Z=0}(t) - \Delta_Q(t)\right]$$
$$\times \left[\overline{F}_{T;S=1,Z=0}(t') - \overline{F}_{T;S=0,Z=0}(t') - \Delta_Q(t')\right]$$
$$+ \frac{p(1)F_{T;S=1,Z=1}(t)\overline{F}_{T;S=1,Z=1}(t')}{nq_1}$$
$$+ \frac{\left[1 - p(1)\right]F_{T;S=0,Z=1}(t)\overline{F}_{T;S=0,Z=1}(t')}{nq_1}$$
$$+ \frac{p(0)F_{T;S=1,Z=0}(t)\overline{F}_{T;S=1,Z=0}(t')}{nq_0}$$
$$+ \frac{\left[1 - p(0)\right]F_{T;S=0,Z=0}(t)\overline{F}_{T;S=0,Z=0}(t')}{nq_0}\Bigg\}.$$

---

[12]See e.g. Angrist and Krueger (1999) and Heckman, LaLonde and Smith (1999) for general discussions of the Wald estimator in the treatment evaluation context.

For $t = t'$, this reduces to the asymptotic variance given by Imbens and Angrist (1994),

$$\frac{\mathbb{E}\left[\left(I(T > t) - \overline{F}_T(t) - \Delta_Q(t)\left(S - \mathbb{E}[S]\right)\right)^2 \left(Z - \mathbb{E}[Z]\right)^2\right]}{\left[\mathrm{cov}\left(S, Z\right)\right]^2}.$$

In general, the functions $\overline{F}_{0;Q}$, $\overline{F}_{1;Q}$, and Δ_Q inherit the disadvantages of the local average treatment effect parameter in Imbens and Angrist's (1994) work. Unless $p(0) = 0$ and $p(1) = 1$ (in which case $S = Z$ a.s.) the set Q, and therefore these functions, are instrument-dependent for given propensity scores $p(0)$ and $p(1)$ (see Heckman, 1997, Heckman, LaLonde and Smith, 1999, and Abbring, 2003, for discussion). Policy-evaluation problems may require information on "parameters of interest" other than the identified local average treatment effects. Moreover, one may be more interested in average treatment effects on individual hazard rates than on survivor functions. This requires imposition of some semi-parametric structure.

3.3 Semi-parametric IV in a proportional hazards framework

In this subsection we adopt structures for the hazard rate of $T(s)$ that are related to the familiar mixed proportional hazards (MPH) model, allowing for unobserved heterogeneity V across individuals. As in Subsection 2.2, this enables us to focus on individual treatment effects and achieve point identification. We only allow $\{T(s)\}$ and S to be dependent by way of a common dependence on the individual V, so $\{T(s)\} \perp\!\!\!\perp S | V$. This means that in the case where S differs from the randomized assignment due to selective compliance, this selection mechanism is captured by V.

We start by adopting the multiplicative structure for $\theta_{T(s)}(t|V)$ from Subsection 2.3 resulting in the MPH model for $T|S, V$,

$$\theta_T(t|S, V) = \lambda(t)\, \gamma^S\, V. \tag{10}$$

We again assume that $\mathbb{E}\left[V\right] < \infty$, but we now replace the assumption from Subsection 2.3 that $S \perp\!\!\!\perp V$ by the assumption that there is an instrument Z that satisfies Assumption 3. We again take Z to be binary (e.g. an ITT indicator, or an IV capturing that otherwise identical regions supply different labor market programs). Among other things, Assumption 3(ii) in this case implies that $Z \perp\!\!\!\perp V$. We now formalize Assumption 3(i) by assuming that $p(1)\mathbb{E}[V|S = 1, Z = 1] \neq$

$p(0)\mathbb{E}[V|S=1,Z=0]$. Note that $p(0) \neq p(1)$ is necessary for this condition to hold. Here, as elsewhere in the paper, the semi-parametric model framework is not well-suited to incorporate endogenous drop-out from ongoing treatments. One would like to model this as self-selection, and the identification of its effect on hazard rates would require additional information (see Heckman, Smith and Taber, 1998, for a discussion).

In the MPH model it is useful to focus on limits as $t \downarrow 0$, because at 0 the dynamic selection that we examined in Subsection 2.2 has not yet taken place. There holds that

$$\gamma = \lim_{t\downarrow 0} \frac{p(1)F_{T;S=1,Z=1}(t) - p(0)F_{T;S=1,Z=0}(t)}{[1-p(0)]F_{T;S=0,Z=0}(t) - [1-p(1)]F_{T;S=0,Z=1}(t)} \tag{11}$$

(see Appendix 1). The right-hand side of this only depends on observable quantities. Thus, γ is identified. We summarize this result in

Proposition 2. *With a valid binary instrument (i.e. that satisfies Assumption 3), and under the assumption that $\mathbb{E}[V] < \infty$, the treatment effect parameter γ in an MPH model with an endogenous binary treatment is identified.*

Note that we do not require exogenous explanatory variables. We also do not require parametric assumptions (like a parametric latent-variable selection equation) on the treatment selection process $\{S(z)\}$. If, in violation of Assumption 3, Z is not informative on S (i.e., if $p(1) = p(0)$) then equation (11) does not have a solution for γ.

We conjecture that, by analogy to Lenstra and Van Rooij (1998), γ can be consistently estimated by the sample equivalent of the right-hand side of (11). This IV estimator can be seen as a version for our non-linear model of the Wald IV estimator of a treatment effect in the linear regression model.[13] To see this, note that at $t \downarrow 0$, the specification (10) resembles a non-linear regression model with an endogenous regressor and a constant treatment effect parameter, and that equation (11) can be re-expressed as follows,

$$\gamma - 1 = \lim_{t\downarrow 0} \frac{\theta_T(t|Z=1) - \theta_T(t|Z=0)}{(1-p(0))\theta_T(t|S=0,Z=0) - (1-p(1))\theta_T(t|S=0,Z=1)} \tag{12}$$

The practical use of estimators that require $t \downarrow 0$ is limited because very short durations are often ill-recorded.

---

[13]See e.g. Angrist and Krueger (1999) and Heckman, LaLonde and Smith (1999) for discussions in a regression model context.

Although the proposition does not concern identification of other model parameters, this may be achieved by way of including observed exogenous explanatory variables X. Our results may be seen as a justification of empirical models that consist of two equations: (i) an MPH model equation for $T|S, X, V$, say $\theta_T(t|S, X, V) = \lambda(t)\gamma^S \exp(\alpha X)V$, and (ii) a latent-variable selection equation, say $S^* = \beta_0 + \beta_1 Z + \beta_2 X + \varepsilon$, with $S := \mathrm{I}(S^* > 0)$, and where ε and V are possibly dependent. The corresponding estimates are not fully driven by functional-form assumptions.

It is clear that a continuous instrument Z enables identification of more general models. Chesher (2003) considers an MPH-type model with an endogenous continuous treatment indicator S as well as exogenous variables X, a continuous instrument Z, and a latent variable equation relating S and Z. He demonstrates local identification of ratios of the derivatives of the individual hazard rate with respect to S and X.

We now proceed to the case where Z represents an ITT and non-compliance is asymmetric in the sense that agents always comply if assigned to the control group ($z = 0$), i.e. $S(0) = 0$ and $p(0) = 0$. Under this restriction, Bijwaard and Ridder (2005) develop an estimator of a treatment effect in an MPH model with a parametric baseline hazard. They exploit that, because of randomization, the subpopulation of agents with $Z = 0$ is representative for the population. This ensures that all parameters except the treatment effect are identified from the data on this subpopulation. The treatment effect is subsequently identified from the outcomes of the agents with $Z = 1$. Here we follow the same approach. This requires an MPH model that is fully identified in the absence of treatments.[14] This is usually achieved by including exogenous X variables, so we augment equation (10) with such variables. In addition, we now allow the treatment effect γ to depend on the elapsed time t since treatment and on X, leading to

$$\theta_T(t|S, V, X) \;=\; \lambda(t)\,\gamma(t, X)^S\,\phi(X)\,V, \tag{13}$$

and we make standard assumptions that ensure identification of λ, ϕ, and the distribution of V in the population in the absence of treatments (notably, this requires $X \perp\!\!\!\perp V$). We obtain

Proposition 3. *Consider a standard MPH model that is augmented by an endogenous binary treatment and that is identified in the absence of this treatment. Assume perfect compliance among the controls. With a valid binary instrument (i.e. that satisfies Assumption 3), the treatment effect as a function of the elapsed*

---

[14]See Heckman and Taber (1994) and Van den Berg (2001) for surveys.

duration and observed covariates is identified, as are all other functions in the model.

Proof. See Appendix 1. □

From a policy point of view it is obviously important to be able to identify the way in which the individual treatment effect changes over time and across individuals.[15]

4 Randomization with later treatment and imperfect compliance

4.1 Model and data

We further extend the model of Subsection 3.1 by not only allowing for imperfect compliance, but also for positive amounts of time between treatment assignment and treatment enrollment. Again, a $\mathcal{Z}$-valued label Z is randomly assigned at time 0 (see Subsection 4.3 for examples). Then, the agents engage in a time-consuming process of enrolling in a program. We focus on the case of a binary program, in which the agents either enroll at some time in $[0, \infty)$ or not enroll at all. Following Abbring and Van den Berg (2003b), we can formalize this by taking $\mathcal{S} = [0, \infty]$. Then, S simply denotes the random time at which an agent enrolls in the program. The point ∞ corresponds to never enrolling at all.

As before, denote the model for the treatment as a function of the instrument by $\{S(z)\}$. In a social experiment, perfect compliance would again arise if $\mathcal{Z} = \mathcal{S}$ and $S(z) = z$. An interpretation is that a full treatment plan Z, stipulating the timing of future program participation, is randomly assigned at time 0 and is adhered to in all states of the world. If Z is observed by the agent, which we typically assume it is, this is the perfect-foresight case briefly mentioned in Subsection 2.2. In this section, we allow for the more relevant case of imperfect compliance, i.e. non-degenerate $\{S(z)\}$.

For expositional convenience, we take the instrument (ITT) to be binary ($\mathcal{Z} = \{0, 1\}$). Because $\mathcal{S}$ is larger than $\mathcal{Z}$, $\mathcal{Z}$ cannot contain a treatment plan for each possible treatment. Therefore, $S(0)$ and $S(1)$ cannot both be non-degenerate, except in the trivial case that S has binary support. Thus, there is imperfect compliance.

---

[15]One might want to consider inference in models where the individual treatment effect γ is allowed to depend in a general way on V, but this seems too ambitious.

In the next subsection we argue that the ITT is likely to affect outcomes directly in this case. To accommodate this, we augment the outcomes model by indexing potential outcomes not only by treatments in $\mathcal{S}$, but also by the labels in $\mathcal{Z}$. Thus, the potential-outcomes process is now $\{T(s,z)\} := \{T(s,z); (s,z) \in \mathcal{S} \times \mathcal{Z}\}$. We again assume that each $T(s,z)$ is continuously distributed, with hazard rate $\theta_{T(s,z)}$ and integrated hazard $\Theta_{T(s,z)}$. For expositional convenience, we restrict the joint distributions of $\{T(s,z)\}$ for fixed z as in

Assumption 4. For all $z \in \mathcal{Z}$, there exists a unit exponential random variable E_z such that $T(s,z) = \Theta^{-1}_{T(s,z)}(E_z)$ for all $s \in [0,\infty]$.

Because we never observe two potential outcomes jointly, Assumption 4 is empirically innocuous.

Following Abbring and Van den Berg (2003b), we assume that there is no anticipation of future treatment. This means that current potential integrated hazards do not depend on future treatment enrollment, i.e.

Assumption 5. For all $s \in [0,\infty)$, $z \in \mathcal{Z}$, $\Theta_{T(s,z)}(t) = \Theta_{T(\infty,z)}(t)$ for all $t \leq s$.

Recall that $\Theta_{T(\infty,z)}$ is the integrated hazard of the potential duration corresponding to never enrolling in treatment.[16]

We assume that treatments are only observed if enrollment has taken place before the outcome spell is completed. This is natural in combination with Assumption 5, and natural in many applications.[17] Thus, we now have a random sample $y((\widetilde{T}_1, S_1 I(S_1 < \widetilde{T}_1), D_1, I(S_1 < \widetilde{T}_1), Z_1), \ldots, (\widetilde{T}_n, S_n I(S_n < \widetilde{T}_n), D_n, I(S_n < \widetilde{T}_n), Z_n))$ from the joint distribution of $(\widetilde{T}, SI(S < \widetilde{T}), D, I(S < \widetilde{T}), Z)$.

4.2   Exclusion restrictions and causal effects

An equivalent of Assumption 3 for this section would require that (i) $z \mapsto S(z)$ is non-trivial, and (ii) $Z \perp\!\!\!\perp (\{T(s,Z)\}, \{S(z)\})$. A sufficient condition for (ii) is that ITT does not causally affect outcomes directly, i.e. $T(s,z) = T(s,z')$ for all $s \in \mathcal{S}$, $z, z' \in \mathcal{Z}$. With imperfect compliance and dynamic enrollment in

---

[16]Note that Assumptions 4 and 5 imply that $T(s,z) = T(\infty,z)$ on $\{T(\infty,z) \leq s\}$.

[17]The analysis can be straightforwardly extended to the case that treatments are always observed. A natural symmetric extension of the present model allows outcomes to affect future treatment and imposes that neither future outcomes nor future treatments are anticipated. Alternatively, under the assumption that treatments are not causally affected by outcomes at all, we can allow for for anticipation of future treatment. See Abbring and Van den Berg (2003b) and Abbring (2003) for details.

treatment, randomization is unlikely to ensure such exclusion. To see this, consider a social experiment in which the treatment is a public training program for the unemployed, and the outcome is the unemployment duration. Suppose that agents operate in a continuous-time dynamic environment in which they may affect treatment and outcomes by (i) investing in some (human) capital, and (ii) searching for job and training opportunities. Agents are informed about their ITT status and possible some other predetermined variables (called V below). Otherwise, information accumulates in the obvious way. In this framework, the ITT status may affect the present value of unemployment, and this affects the optimal strategy concerning exit to work. Thus, at the level of an individual agent, ITT not only causally affects the moment of treatment, but it may also causally affect outcomes. In that case there is no exclusion restriction and we may not apply the methods of Case II. We will now make these claims more precise by distinguishing between the periods before and after treatment.

The causal effect on the outcome before the actual treatment is called the ex ante causal ITT effect on the outcome. In line with this, we say that if this effect is absent then the ex ante exclusion restriction is satisfied. Violation of the ex ante exclusion restriction is often likely, because ITT has to causally affect actual treatment in order to be a useful instrument in the first place (as in Assumption $3(i)$), and anything observable that influences the distribution of the moment of a relevant event in the future has an effect on current behavior. To put it simple: if you like training and you are likely to get it, then you stop searching until the training; if you do not like training, and you are likely to get it, then you make a maximum effort to find a job before the training starts; if you are unlikely to get it then you search with moderate effort (see Van den Berg, 2004, for a decision-theoretical analysis of IV in general in a dynamic setting with forward-looking agents).

The causal effect of ITT on the outcome after actual enrollment in the treatment is called the ex post causal ITT effect on the outcome, and we refer to the absence of such effects as ex post exclusion. The ex post exclusion restriction is not per se inconsistent with ITT being an instrument and may be reasonable in some applications. In the training example, this restriction may be violated if ITT affects investments in financial or human capital.

A formal statement of such an assumption requires notation that allows us to explicitly control for dynamic selection. To this end, again suppose that all ex ante heterogeneity is captured by a random variable V and that $\{T(s,z)\} \perp\!\!\!\perp \{S(z)\}|V$. Then, a weak version of Assumption 3 that only imposes ex post exclusion is

Assumption 6 (IV with ex post exclusion). (i) $z \mapsto S(z)$ is non-trivial, (ii)

$Z \perp\!\!\!\perp (\{T(s,z)\}, \{S(z)\}, V)$, and (iii) for all $s \in [0, \infty)$, and $z, z' \in \mathcal{Z}$,

$$\theta_{T(s,z)}(t|V) = \theta_{T(s,z')}(t|V) \quad \text{almost surely,} \quad \text{for all } t > s. \tag{14}$$

Informally, (iii) states that $\theta_{T(s,z)}(t|V)$ does not vary with z after s. Note that without Assumption 6 there would be no need to introduce V at this stage. Also note that ex ante exclusion would require equation (14) to hold for all $t \leq s$.

With ex post exclusion, we are left with three causal effects or causal parameters. In terms of effects on individual hazard rates, they can be expressed as follows. First and foremost, the ex post causal treatment effect of s on the outcome hazard rate after s ("the" treatment effect). Second, the ex ante causal effect of z on the outcome hazard rate before s. Third, the ex ante causal effect of z on the treatment rate $\theta_{S(z)}(t|V)$ before s. In the experimental training program example, these are, respectively, the effect of training on the exit rate to work, the effect of being randomized in, on the exit rate to work before the training, and the effect of being randomized in, on the rate of getting into training. Note that Assumption 6(i) states that the third causal effect exists.

The ex ante causal effect on the outcome is represented by the change in the distribution of $T(\infty, z)$ when z changes from 0 to 1. As discussed above, we expect this effect to operate because of the effect of z on the random treatment assignment rule $S(z)$. Under the assumption that this is the only channel through which z affects outcomes, the ex ante causal effect on the outcome reflects the behavioral effect of moving between the assignment rules $S(0)$ and $S(1)$. In general, both $S(0)$ and $S(1)$ may have at least some support in $(0, \infty)$, so under both rules, agents may face a positive probability of eventually enrolling in the treatment. We must therefore contrast the outcomes under these actual treatment rules $S(0)$ and $S(1)$ to an agent's outcomes in the hypothetical case the agent faces a rule that never assigns treatment (i.e. is degenerate at ∞). Unless either $S(0)$ or $S(1)$ is such a rule, this effect can in general not be identified. Still, this behavioral effect is likely to be non-zero (especially if the ex ante causal effect on the outcome hazard is non-zero), and in that case its presence affects the interpretation of the identified causal effects (see in particular Subsection 4.5.2).

4.3 Reduced-form ITT analysis

As with our Case II, one may carry out a reduced-form analysis of the over-all effect of ITT on the outcome duration. In particular, one may non-parametrically study the average additive effect of z on the survival probability of $T(S(z), z)$. This is a net effect of the three causal effects defined above. The analysis again fits under our Case I.

The usefulness of such an analysis is limited, for two reasons. The first also applies to reduced-form ITT analysis in Case II and was already mentioned in the introduction of the paper, namely that the analysis does not enable one to learn the size of the average causal effect of the actual treatment. Below we provide two examples in which in the context of our Case III it is useful to know this treatment effect. The second limitation only applies to Case III. Recall that in Case II the null hypothesis of no over-all ITT effects is equivalent to the null hypothesis of no average causal treatment effect (at least, if one is interested in the average additive effect on the outcome survival probability). As is clear from the previous subsection, this equivalence breaks down in the present Case III.

Our first example concerns the effect of a training program for unemployment insurance recipients on their unemployment duration. In a social experiment, a random subsample of the inflow into unemployment is chosen to become participant in the program. Actual training occurs at a later point in time, depending on the behavior of the case worker and other (potential) applicants. This moment is random from the point of view of the individual. Individuals may choose to withdraw from the program, in which case they face the distribution of the moment of training that is also faced by those who are initially randomized out. This description is not dissimilar to the actual enrollment into the experimental JTPA program (see e.g. Heckman, LaLonde and Smith, 1999). In a non-experimental situation, the randomization upon inflow may be replaced by different regional active labor market policies in otherwise identical regions, as an instrumental variable.

One may want to know the causal effect of having been trained on the exit rate to work, as well as the ex ante causal effect of being eligible for training on the exit rate to work before actual training. For example, if the former is positive and the latter is negative then this may suggest that the program should be offered more frequently or earlier in the spell of unemployment. In this case a reduced-form ITT analysis may just indicate that the net effect is zero. As another illustration, Richardson and Van den Berg (2001) show that the effect of an expensive training program only works up to three months after finishing the training. This suggests that the merits of the program are in job search assistance rather than human capital accumulation, in which case the program can be replaced by a much cheaper job search assistance program.

Our second example is the Moving to Opportunity for Fair Housing Demonstration (MTO). This is an experimental residential mobility program that has been carried out by the U.S. Department of Housing and Urban Development in five major U.S. cities since 1994. It is targeted to families living in public housing

or project-based subsidized private housing in poor neighborhoods.[18] Participating families are randomized into three groups, the MTO experimental group, the Section 8 Comparison group, and the Control group (Feins, Holin and Phipps, 1996). Families in the MTO experimental group receive vouchers that can only be used to subsidize private housing in low-poverty areas. Intensive counseling and assistance are provided to support these families' search for such housing. The Section 8 Comparison group receives geographically unrestricted Section 8 vouchers and ordinary briefings and assistance.[19] Finally, the Control group does not receive any vouchers, but continues to qualify for public housing or project-based support.

A wide variety of outcome measures, such as housing conditions, employment, earnings, and welfare dependence, are recorded in the MTO demonstration. A comparison of these outcomes across the three program groups is of clear policy interest. Because of the randomization of group assignment, such ITT analyses are informative on the effects of the MTO and the existing Section 8 voucher programs on the target group's socio-economic outcomes. However, the MTO demonstration explicitly states the ambition to move beyond ITT analyses and to separately explore the effects of the program on housing mobility and the subsequent effects of housing mobility on other outcomes (Feins, Holin and Phipps, 1996; Orr et al., 2003). There is a direct connection to the empirical literature on neighborhood effects. Katz, Kling and Liebman (2001), for example, study the MTO program in Boston and propose to use the initial program status as an instrument for housing mobility.

4.4   Non-parametric identification

Let $\Pr(S = T) = 0$. Following Abbring and Van den Berg (2003b), a large data set would provide[20]

$$
\begin{aligned}
Q_{S;Z=z}(t,s) &:= \Pr(T > t, S > s, T > S | Z = z) \quad \text{and} \\
Q_{T;Z=z}(t) &:= \Pr(T > t, T < S | Z = z)
\end{aligned}
\tag{15}
$$

for all $(t,s) \in \mathbb{R}_+^2$ and $z = 0, 1$. These are the sub-survivor-functions of (T, S) and T on $Z = z$ for the subpopulations with respectively $T > S$ and $T < S$.

---

[18]This includes only housing subsidized under the project-based Section 8 program. In this program, rent subsidies are connected to dwellings: if tenants move from project-based Section 8 housing they lose their project-based Section 8 subsidies.

[19]The Section 8 vouchers are housing subsidies connected to tenants and should not be confused with the project-based Section 8 assistance that some families originally receive.

[20]Simple random censoring does not matter for identification, provided that some straightforward support conditions hold.

As before, non-parametric identification is concerned with average additive effects on survival probabilities. We first consider the identifiability of the average additive ex ante causal effect of z on $\overline{F}_{T(\infty,z)}(t|V)$, and we simultaneously consider the likewise effect of z on $\overline{F}_{S(z)}(t|V)$ (note that the focus on $\overline{F}_{T(\infty,z)}$ here is justified by Assumption 5). Thus, we examine the differences $\overline{F}_{T(\infty,1)}(t) - \overline{F}_{T(\infty,0)}(t)$ and $\overline{F}_{S(1)}(t) - \overline{F}_{S(0)}(t)$.

Any effects of z on the treatment process are only observable (and only relevant) on $[0,T]$. Moreover, by Assumption 6, z only possibly causally affects the outcome process on $[0,S]$. Intuitively, we should therefore learn about these effects from data on the "identified minimum" of (T,S), i.e. the smallest of T and S joint with the identity of this smallest duration, for $Z = 0$ and $Z = 1$. The distribution of this identified minimum given $Z = z$ is fully characterized by $(Q_{S;Z=z}^0, Q_{T;Z=z})$, with $Q_{S;Z=z}^0(\cdot) := Q_{S;Z=z}(-\infty, \cdot)$ (Tsiatis, 1975). We can think of such data as being generated by a competing risks model that is embedded in our full model, and in which one risk is enrollment in the treatment (which terminates the ex ante causal effect on the outcome) and the other risk is realized by the outcome transition (which terminates observability of the ex ante causal effect on the treatment process).

In the particular case that $\{S(z)\} \perp\!\!\!\perp \{T(s,z)\}$ (randomized treatment), there holds that $(Q_{S;Z=z}^0, Q_{T;Z=z})$ is consistent with an independent competing risks model in which the outcome risk has survivor function $\overline{F}_{T(\infty,z)}$ and the treatment risk has survivor function $\overline{F}_{S(z)}$. If it is known that the risks are independent, this means that $\overline{F}_{T(\infty,0)}$ and $\overline{F}_{S(0)}$ are identified from $(Q_{S;Z=0}^0, Q_{T;Z=0})$, and $\overline{F}_{T(\infty,1)}$ and $\overline{F}_{S(1)}$ are identified from $(Q_{S;Z=1}^0, Q_{T;Z=1})$, all in this particular case of randomized treatment.

However, we allow for general dependence of $\{S(z)\}$ and $\{T(s,z)\}$ through a common dependence on the unobservable V. The data $(Q_{S;Z=z}^0, Q_{T;Z=z})$ are then consistent with a particular dependent risks model in which the outcome risk has marginal survivor function $\overline{F}_{T(\infty,z)}$ and the treatment risk has marginal survivor function $\overline{F}_{S(z)}$. Dependent competing risks models are not non-parametrically identified (Cox, 1959, 1962; Tsiatis, 1975). In general there is an observationally equivalent independent competing risks model, with marginal survivor functions that do not equal $\overline{F}_{T(\infty,z)}$ and $\overline{F}_{S(z)}$. In sum, the two ex ante causal effects of z are not non-parametrically identified from $(Q_{S;Z=z}^0, Q_{T;Z=z}; z = 0, 1)$.[21]

Application of Peterson's (1976) bounds for marginal survivor functions in

---

[21]See Abbring and Van den Berg (2003a) for some intuition on (non-)identification of competing risks models. Note that the difference between the "crude" ex ante outcome hazards $\theta_T(t|S \geq t, Z = z)$ at $z = 1$ and $z = 0$ does not only reflect the ex ante causal effect of z on the

dependent competing risks models gives that

Proposition 4. *For given data* $(Q^0_{S;Z=z}, Q_{T;Z=z})$, $\overline{F}_{T(\infty,z)}$ *and* $\overline{F}_{S(z)}$ *satisfy*

$$Q^0_{S;Z=z} + Q_{T;Z=z} \leq \overline{F}_{T(\infty,z)} \leq Q_{T;Z=z} + Q^0_{S;Z=z}(-\infty) \quad and$$
$$Q^0_{S;Z=z} + Q_{T;Z=z} \leq \overline{F}_{S(z)} \leq Q^0_{S;Z=z} + Q_{T;Z=z}(-\infty),$$

for $z = 0, 1$. *The bounds are sharp.*

The bounds in Proposition 4 are typically wide and may overlap across groups with different z, whether z affects $T(\infty, z)$ and $S(z)$ or not. With overlap, the ex ante causal effects of z cannot even be signed. This would imply that, contrary to typical IV analyses, one cannot infer empirically whether the ITT variable z has a causal effect on S. Moreover, a priori information on one of the risks is not informative on the marginal distributions of the other risk: the bounds on either marginal distribution can be attained even if we arbitrarily fix the other marginal distribution.[22]

As an example, suppose that $T(\infty, z)$ is exponential with parameter μ_z, $S(z)$ is exponential with parameter ν_z, and that $\{S(z)\} \perp\!\!\!\perp \{T(s, z)\}$. To shape thoughts, one may think of T, S, and Z as the unemployment duration, the unforeseeable duration at which unemployment benefits are reduced, and the randomized assignment of different rates of benefits reduction, respectively. Such a setting arises in the case of an unemployment-insurance system with imperfect monitoring and punitive benefits reductions (Abbring, Van den Berg and Van Ours, 2005). If the researcher does not know or assume that $\{S(z)\} \perp\!\!\!\perp \{T(s, z)\}$, then he has to settle for Proposition 4's bounds. If we replace the directly estimable expressions for the bounds in Proposition 4 by the corresponding model expressions based on

---

outcome, but also differences in unobservable characteristics between the subpopulations

$$\{S(0) \geq t, T(\infty, 0) \geq t, Z = 0\} \quad and \quad \{S(1) \geq t, T(\infty, 1) \geq t, Z = 1\}.$$

These differences in particular exist because $z \mapsto S(z)$ is non-trivial by Assumption 6, so that $S(0) \geq t$ and $S(1) \geq t$ in general select different subpopulations. These selection effects disappear as $t \downarrow 0$, but in the non-parametric case this cannot be exploited.

[22]Peterson's bounds on, for example, $\overline{F}_{T(\infty,z)}$ follow from $\overline{F}_{T(\infty,z)} = Q_{T;Z=z} + Q^*_{T;Z=z}$, where $Q^*_{T;Z=z}(t) := \Pr(T > t, S < T | Z = z)$. We know $Q_{T;Z=z}$. Non-tight bounds on $\overline{F}_{T(\infty,z)}$ arise because we only know that $Q^0_{S;Z=z} \leq Q^*_{T;Z=z} \leq Q^0_{S;Z=z}(\infty)$. Now, suppose that we do not only know $Q^0_{S;Z=z}$ and $Q_{T;Z=z}$, but also $\overline{F}_{S(z)}$. This is equivalent to also knowing the marginal sub-survivor-functions $Q^*_{S;Z=z}$ of S on $\{S > T\}$. Given $Q^0_{S;Z=z}$, $Q^*_{S;Z=z}$ is clearly not informative on the sub-survivor-functions $Q^*_{T;Z=z}$ of T on $\{S < T\}$. Therefore, this additional information cannot be used to tighten the bounds on $\overline{F}_{T(\infty,z)}$.

the exponential distributions then we obtain

$$\exp\left[-(\mu_z + \nu_z)t\right] \leq \overline{F}_{T(\infty,z)}(t) \leq \frac{\mu_z}{\mu_z + \nu_z}\exp\left[-(\mu_z + \nu_z)t\right] + \frac{\nu_z}{\mu_z + \nu_z} \quad \text{and}$$

$$\exp\left[-(\mu_z + \nu_z)t\right] \leq \overline{F}_{S(z)}(t) \leq \frac{\nu_z}{\mu_z + \nu_z}\exp\left[-(\mu_z + \nu_z)t\right] + \frac{\mu_z}{\mu_z + \nu_z},$$

for $z = 0, 1$.

Now let $\mu_0 = \nu_0 = 1$ and $\mu_1 = \nu_1 = 2$. (Agents in the group with the higher rate of benefits reduction respond to this by increasing their unemployment exit rate.) The bounds then reduce to

$$l_0(t) := \exp\left(-2t\right) \leq \overline{F}_{T(\infty,0)}(t), \overline{F}_{S(0)}(t) \leq \frac{1}{2}\exp\left(-2t\right) + \frac{1}{2} =: u_0(t) \quad \text{and}$$

$$l_1(t) := \exp\left(-4t\right) \leq \overline{F}_{T(\infty,1)}(t), \overline{F}_{S(1)}(t) \leq \frac{1}{2}\exp\left(-4t\right) + \frac{1}{2} =: u_1(t).$$

Now note that $l_0(0) = u_0(0) = l_1(0) = u_1(0) = 1$ and, more importantly, that $l_1 < l_0 < u_1 < u_0$ on $(0, \infty)$. Thus, even though the ex ante causal effect of z on the outcome is substantial, the bounds for $F_{T(\infty,z)}$ overlap for $z = 0, 1$, meaning that the ex ante causal effect of z on the outcome cannot be signed. The same applies to the ex ante causal effect of z on the treatment.

Clearly, without identification of the two ex ante causal effects of z on the survivor functions $\overline{F}_{T(\infty,z)}$ and $\overline{F}_{S(z)}$, the scope for non-parametric identification of the ex post causal effect of treatment enrollment on outcomes is very limited. However, the exclusion restriction embedded in Assumption 6(*iii*) does allow for an assessment of unobserved heterogeneity in the hazard rate $\theta_{T(s,z)}$ of $T(s,z)$ after treatment enrollment. To see this, note that absence of a direct causal effect of z on $\theta_{T(s,z)}(t|V)$ at $t > s$ implies that any observed relation between z and the hazard rate $\theta_T(t|S = s, Z = z)$ must be due to a selection effect of z on the distribution of V among survivors at s. Such a selection effect can arise for two reasons. First, z may have an ex ante causal effect on the individual hazard rate $\theta_{T(s,z)}(t|V)$ before treatment, and this affects the distribution of $[V \mid T \geq t, S = s, Z = z]$. Second, $z \mapsto S(z)$ is non-trivial by Assumption 6, and $\{S(z)\}$ and $\{T(s,z)\}$ may be dependent through joint dependence on V, in which case the distribution of $[V \mid T \geq t, S = s, Z = z]$ again varies with z.

This can be used for a non-parametric test on unobserved heterogeneity in the outcome hazard rate after treatment enrollment.[23] If the null hypothesis of no unobserved heterogeneity is accepted then one may proceed by estimating

---

[23]The underlying idea can be related to tests for unobserved heterogeneity and dynamic selection effects in the event-history literature. First, consider the test by Van den Berg and Van Ours (1997) on (dynamic selection effects due to) unobserved heterogeneity in duration

models that impose this hypothesis. Recall that the test requires that z has a causal effect on S, which is a non-testable assumption in this case.

4.5 Semi-parametric approaches

4.5.1 Bounds for the two ex ante causal effects

Bond and Shaw (2003) provide bounds on functions that are informative on covariate effects on marginal distributions in a general class of competing risks models. In this subsection we apply these to sign the average additive ex ante causal effects of z on the survival probabilities of $T(\infty, z)$ and $S(z)$.

   With the non-parametric bivariate competing risks model for $\{T(\infty, z), S(z)\}$ as a starting point, we make the assumption that the dependence structure is invariant to the values of the covariates (i.e., in our case, z), so that the covariates only affect the marginal distributions. Specifically,

Assumption 7 (Dependency-invariance to z). There exist increasing functions ξ_S and ξ_T such that $(S(0), T(\infty, 0))$ equals $(\xi_S(S(1)), \xi_T(T(\infty, 1)))$ in distribution.[24]

This implies that the copula (and therefore Kendall's τ) associated with the joint distribution is invariant to the value of z. The copula describes the dependence structure while being distribution-free with respect to the marginals.

   Assumption 7 is satisfied if the two hazard rates associated with $T(\infty, z)|V$ and $S(z)|V$ are multiplicative in V (this is the bivariate extension of Assumption 2 in Section 2). To see this, note first that we may allow V to be two-dimensional, which leads to

$$\Pr\left(S(z) > s, T(\infty, z) > t\right) = \mathcal{L}\left(\Lambda_{S(z)}(s), \Lambda_{T(\infty, z)}(t)\right) \quad (z = 0, 1),$$

with $\mathcal{L}$ the bivariate Laplace transform of the mixing distribution. This satisfies Assumption 7 with $\xi_S = \Lambda_{S(0)}^{-1} \circ \Lambda_{S(1)}$ and $\xi_T = \Lambda_{T(\infty, 0)}^{-1} \circ \Lambda_{T(\infty, 1)}$. Assumption 7 is also satisfied by the MPH competing risks model studied by e.g. Abbring

---

analysis. If the value of an exogenous time-varying explanatory variable in the first period is related to the observed hazard rate in the second period then this indicates such selection effects. Second, in Abbring, Chiappori and Pinquet (2003)'s model of car-insurance claims under moral hazard and experience rating, individual claim propensities are only causally affected by the past occurrence of claims, but not their timing. Any variation of observed claim intensities with the timing of past claims conditional on their occurrence should therefore be due to selection effects.

[24]Bond and Shaw (2003) call $t \mapsto ((t, \xi_S), (t, \xi_T))$ a "covariate-time transformation". For both risks, $t \mapsto t$ is the (normalized) time-transformation for the first covariate value ($z = 0$).

and Van den Berg (2003a) and the more general model studied by Heckman and Honoré (1989).

Assumption 7 is insufficient for point identification. Assuming in addition that the dependence works by way of multiplicative V in the hazard rates does not help here. However, Bond and Shaw (2003) provide bounds for (ξ_S, ξ_T). Those are all we need to know to rank $F_{T(\infty,z)}$ and $F_{S(z)}$ over z in terms of first-order stochastic dominance. For example, if $\xi_S(t) = t$ for all t then $\overline{F}_{S(0)} = \overline{F}_{S(1)}$. If $\xi_S(t) > t$ for all t, on the other hand, then $\overline{F}_{S(1)} < \overline{F}_{S(0)}$, etcetera.

Define $\overline{Q}_{S;Z=z}(t) := \Pr(S \leq t, T > S | Z = z)$, $\overline{Q}_{T;Z=z}(t) := \Pr(T \leq t, T < S | Z = z)$, and $H_{Z=z}(t) := \Pr(S \leq t, T \leq t | Z = z)$, for $t \geq 0$. These are all observable. Bond and Shaw (2003)'s results immediately imply

Proposition 5. *Suppose that either* $\xi_S \geq \xi_T$ *or* $\xi_S \leq \xi_T$. *Then, either*

$$H_{Z=0}^{-1} \circ H_{Z=1} \quad \leq \xi_S \leq \quad \overline{Q}_{S;Z=0}^{-1} \circ \overline{Q}_{S;Z=1} \quad and$$
$$\overline{Q}_{T;Z=0}^{-1} \circ \overline{Q}_{T;Z=1} \quad \leq \xi_T \leq \quad H_{Z=0}^{-1} \circ H_{Z=1}$$

or

$$H_{Z=0}^{-1} \circ H_{Z=1} \quad \geq \xi_S \geq \quad \overline{Q}_{S;Z=0}^{-1} \circ \overline{Q}_{S;Z=1} \quad and$$
$$\overline{Q}_{T;Z=0}^{-1} \circ \overline{Q}_{T;Z=1} \quad \geq \xi_T \geq \quad H_{Z=0}^{-1} \circ H_{Z=1}.$$

These bounds are tight if $\xi_S = \xi_T$.

Proof. See Appendix 2. $\qquad\qquad\square$

Note that either the lower bound on ξ_S coincides with the upper bound on ξ_T or the upper bound on ξ_S equals the lower bound on ξ_T. As a consequence, either the potential-treatment distributions or the potential-outcome distributions (or both) can be ranked.

4.5.2 Point-identification of ex ante and ex post causal effects

We start by imposing a semi-parametric structure on the embedded competing risks model for $T(\infty, z)$ and $S(z)$, in order to achieve point identification of the two ex ante causal effects. In line with the semi-parametric models in the previous sections, which were based on the MPH framework, we impose an MPH structure, that is,

$$
\begin{aligned}
\theta_{T(\infty,z)}(t|X,V) &= \lambda_1(t)\, V_1\, e^{\beta_1 X + \eta_1 z} \\
\\
\theta_{S(z)}(t|X,V) &= \lambda_2(t)\, V_2\, e^{\beta_2 X + \eta_2 z},
\end{aligned}
\tag{16}
$$

where $V := (V_1, V_2)$. The ex ante causal effect of z on the outcome T is represented by η_1. The ex ante exclusion restriction would impose $\eta_1 = 0$. Note that in this MPH model agents face a positive treatment enrollment rate in both ITT states ($z = 0$ and $z = 1$). Recall from Subsection 4.2 that this implies that neither $\theta_{T(\infty,0)}$ nor $\theta_{T(\infty,1)}$ is likely to equal the ex ante outcome hazard under a hypothetical rule that never assigns treatment. Outcomes under such a rule are not specified in the model and consequently are not identified. The parameter η_2 captures the ex ante causal effect of z on S.

As shown by Abbring and Van den Berg (2003a), this model is identified under some additional conditions, notably (i) $V \perp\!\!\!\perp X$ and (ii) X contains at least two continuous regressors. In particular, the individual ex ante causal effects can be identified from the observable sub-survival functions near 0.

We now turn to point identification of the ex post causal effect of treatment enrollment at s on the outcome T. Consider the following model framework,

$$
\begin{aligned}
\theta_{T(s,z)}(t|X,V) &= \lambda_1(t) \, V_1 \, \gamma^{\mathrm{I}(t>s)} \, e^{\beta_1 X + \eta_1 z \, \mathrm{I}(t \le s)} \\
\\
\theta_{S(z)}(t|X,V) &= \lambda_2(t) \, V_2 \, e^{\beta_2 X + \eta_2 z},
\end{aligned}
\tag{17}
$$

This is the Timing-of-Events model (Abbring and Van den Berg, 2003b) augmented with an ITT or IV variable Z. We make the same underlying assumptions on the model determinants as in Abbring and Van den Berg (2003b). The model embeds the competing risks model (16). If Z is degenerate at say $Z = 0$ then the model reduces to the Timing-of-Events model, and the ex ante causal effects η_1 and η_2 are not identified. Note that the model satisfies Assumption $6(iii)$ (ex post exclusion) since $\theta_{T(s,z)}(t|X,V)$ does not depend on z for $t > s$. We relax this below.

The ex post causal effect in this augmented Timing-of-Events model is represented by

$$
\frac{\theta_{T(s,z)}(t|X,V)}{\theta_{T(\infty,z)}(t|X,V)} = \gamma e^{-\eta_1 z}
$$

for $t > s$. Unlike the ex post causal effect in the Timing-of-Events model, this effect may depend on the ITT status z. Clearly, if Z is dispersed and γ is identified then the augmented model framework allows for richer inference on the ex post causal effect than the Timing-of-Events model. The relevance of this richer specification can be illustrated with the unemployment-insurance example of Subsection 4.4. In this example, z reflects the strictness of monitoring search effort, and s is the time at which benefits are (punitively) reduced. For simplicity,

ignore the punitive nature of benefits reductions and suppose that $z = 1$ simply corresponds to a higher rate of benefits reductions than $z = 0$. Then, an agent assigned to $z = 1$ will search harder ex ante, before a benefits reduction, and increase his search effort less when benefits are actually reduced ($1 \leq \gamma e^{-\eta_1} \leq \gamma$).

In fact, the identification proofs of Abbring and Van den Berg (2003b) can be straightforwardly adapted to prove identification of the full model (17). In particular, we may simply treat Z as another regressor in the embedded competing-risks model, so

Proposition 6. *The Timing-of-Events model augmented by a binary treatment assignment indicator Z (see equations (17)) is identified without ex ante exclusion restrictions.*

Some comments are in order. First, the variation in Z enables us to identify substantially more than Abbring and Van den Berg (2003b). In particular, we can now identify the ex ante effect on the outcome hazard rate, and we can now identify different ex post effects for those assigned to different z.

Second, we may include in (17) an additive term $\eta_3 z\ \mathrm{I}(t > s)$ to the log outcome hazard $\log \theta_{T(s,z)}(t|X,V)$ for $t > s$, capturing a direct (ex post causal ITT) effect of z on the outcome hazard. This does not lead to loss of identification, provided that both $z = 0$ and $z = 1$ lead to a positive probability of being treated before the outcome is realized. Incidentally, this also demonstrates that Assumption 6(iii) is not required to achieve identification of the model with $\eta_3 = 0$.[25,26]

Third, the above model is also over-identified in other directions. Following Abbring and Van den Berg (2003b), γ in (17) can be allowed to depend on t and X and on either s or a multiplicative unobserved heterogeneity term that may itself depend on V_1, V_2. This incorporates heterogeneous treatment effects. The data variation in the moment of treatment is pivotal here.

In addition, if for any given z the corresponding Timing-of-Events model is identified then we may allow for interactions between z on the one hand and the

---

[25]Recall that if we do not impose Assumption 6(iii) then the test on unobserved heterogeneity from Subsection 4.4 can not be applied. One may of course still test for $\eta_3 = 0$ in models (17) that incorporate unobserved heterogeneity.

[26]Because X affects S, one may claim that the effect of X on the outcome hazard rate before s incorporates an ex ante effect. In that case, it seems restrictive to impose that the effect of X on the outcome hazard after s is the same as before s. A similar case can be made for λ_1. The identification results in Abbring and Van den Berg (2003b) can be applied to prove that these extensions do not lead to loss of identification. Also note that the interpretation of η_3 affects the interpretation of γ. For example, one may interpret η_3 as an indicator of effect heterogeneity. We do not pursue this further here.

covariate effects and duration dependence terms in (17) on the other. Note that this automatically leads to violation of Assumption 6(*iii*), although this can be remedied by imposing absence of such interactions if $t > s$. Finally, the presence of exogenous time-varying explanatory variables generally allows for identification of less restrictive duration models (see Honoré, 1991, and Heckman and Taber, 1994).

Does it help to impose restrictions on the modeling of non-compliance? Consider the interpretation of Z as a randomized ITT. In the above model, non-compliance is incorporated in the specification of $\{S(z)\}$. However, in practical applications, the experimental design may lead to types of non-compliance that justify restrictions on $\{S(z)\}$. For example, like in Subsection 3.3, non-compliance may be asymmetric in the sense that agents always comply if assigned to the control group ($z = 0$), i.e. $\Pr(S(0) = \infty) = 1$. As another example, $S(z)$ may have a mass point at ∞ with $\Pr(S(z) = \infty)$ dependent on z and with $[S(z)|S(z) < \infty]$ having a distribution that does not depend on z. Unfortunately, such restrictions do not enable identification under much weaker conditions than above, because they do not substantially facilitate the identification of the embedded dependent competing risks model for $T(\infty, z)$ and $S(z)$.

We conclude this section by noting that even though ITT is randomly assigned, not much can be learned about either (*i*) ex ante ITT effects or (*ii*) ex post treatment effects, without imposing semi-parametric structure. This means that results based on actual social (or, for that sake, laboratory) experiments depend on the chosen semi-parametric structure.

5   Conclusion

Social experiments in which the outcome of interest is a duration variable are more difficult to analyze than social experiments with time-independent outcomes. First, the outcome may be censored. Second, the randomization occurs at time zero but the composition of survivors changes over time in different ways in the treatment and control groups, and this complicates inference of average effects on the individual hazard rate. The paper studies the three most important benchmark cases, distinguished by whether treatment is immediate or not and whether compliance is perfect or not.

In the intermediate case of imperfect compliance and immediate treatment, one can make non-parametric inference on local average treatment effects on survival probabilities, and we derive the relevant asymptotic results. To infer average effects on hazard rates, one has to resort to semi-parametric models. As

a by-product, the paper establishes identification of MPH models with a binary endogenous regressor and a binary instrumental variable. In the most complicated case (imperfect compliance, later treatment), non-parametric analyses are not informative on (ex ante) effects on survival before actual treatment, nor on ex post effects. Non-parametric bounds are so wide that ex ante effects typically can not be signed. Again, semi-parametric models provide identification, and the paper establishes identification of Timing-of-Events models augmented with ex ante effects.

The results of this paper lead to the conclusions that (1) while it is possible to make non-parametric inferences on additive effects on survival probabilities, the degree to which these are informative decreases with the complexity of the case at hand, and (2) to study the more interesting average effects on individual hazard rates and conditional exit probabilities one needs a semi-parametric structure, despite the randomization at time zero. In sum, despite the randomization not much can be learned without a semi-parametric structure.

One approach to these problems is to use more complex dynamic experimental designs than the simple design considered in this paper. However, the main arguments and results carry over to such designs in which randomization takes place repeatedly at discrete (possibly random) times. If actual treatment enrollment takes place at the same times, an extension of our first two cases applies. If agents enroll in treatment more frequently, say continuously, then an extension of our third case applies. With continuous outcomes and treatment enrollment, repeated randomization can at best reduce the inference problems, e.g. by narrowing bounds on some parameters of interests, but not solve these problems altogether. A topic for further research is to design and explore the use of experiments in which individuals are on purpose misinformed about the regime they are assigned to.

The results have some implications for the design of social experiments and laboratory experiments. First, experimental inference requires semi-parametric structure, so results depend on the chosen semi-parametric structure. Second, it is useful to collect as many explanatory variables on the subjects as possible, for two reasons: it serves to reduce the magnitude of unobserved heterogeneity, and it facilitates the semi-parametric inference. Third, it is advisable to minimize imperfect compliance.

Appendix

Appendix 1   Proofs of statements in Section 3

Proof of Proposition 1. First write

$$\sqrt{n}\left(\widehat{\overline{F}}_{0;Q} - \overline{F}_{0;Q}\right) = \frac{1 - \widehat{p}(0)}{\widehat{p}(1) - \widehat{p}(0)}\sqrt{n}\left(\widehat{\overline{F}}_{T;S=0,Z=0} - \overline{F}_{T;S=0,Z=0}\right)$$

$$- \frac{1 - \widehat{p}(1)}{\widehat{p}(1) - \widehat{p}(0)}\sqrt{n}\left(\widehat{\overline{F}}_{T;S=0,Z=1} - \overline{F}_{T;S=0,Z=1}\right)$$

$$+ \overline{F}_{T;S=0,Z=0}\sqrt{n}\left(\frac{1 - \widehat{p}(0)}{\widehat{p}(1) - \widehat{p}(0)} - \frac{1 - p(0)}{p(1) - p(0)}\right)$$

$$- \overline{F}_{T;S=0,Z=1}\sqrt{n}\left(\frac{1 - \widehat{p}(1)}{\widehat{p}(1) - \widehat{p}(0)} - \frac{1 - p(1)}{p(1) - p(0)}\right) \quad \text{and}$$

$$\sqrt{n}\left(\widehat{\overline{F}}_{1;Q} - \overline{F}_{1;Q}\right) = \frac{\widehat{p}(1)}{\widehat{p}(1) - \widehat{p}(0)}\sqrt{n}\left(\widehat{\overline{F}}_{T;S=1,Z=1} - \overline{F}_{T;S=1,Z=1}\right)$$

$$- \frac{\widehat{p}(0)}{\widehat{p}(1) - \widehat{p}(0)}\sqrt{n}\left(\widehat{\overline{F}}_{T;S=1,Z=0} - \overline{F}_{T;S=1,Z=0}\right)$$

$$+ \overline{F}_{T;S=1,Z=1}\sqrt{n}\left(\frac{\widehat{p}(1)}{\widehat{p}(1) - \widehat{p}(0)} - \frac{p(1)}{p(1) - p(0)}\right)$$

$$- \overline{F}_{T;S=1,Z=0}\sqrt{n}\left(\frac{\widehat{p}(0)}{\widehat{p}(1) - \widehat{p}(0)} - \frac{p(0)}{p(1) - p(0)}\right).$$

Next, note that $\sqrt{n}\left(\widehat{\overline{F}}_{T;S=0,Z=z} - \overline{F}_{T;S=0,Z=z}\right)$ $(z = 0, 1)$, $\sqrt{n}\left(\widehat{\overline{F}}_{T;S=1,Z=z} - \overline{F}_{T;S=1,Z=z}\right)$ $(z = 0, 1)$, and $\sqrt{n}\left(\widehat{p}(z) - p(z)\right)$ $(z = 0, 1)$ converge jointly in distribution to $-\overline{F}_{T;S=0,Z=z}\mathbb{G}_{0z}/\sqrt{(1 - p(z))q_z}$ $(z = 0, 1)$, $-\overline{F}_{T;S=1,Z=z}\mathbb{G}_{1z}/\sqrt{p(z)q_z}$ $(z = 0, 1)$, and $\mathcal{N}_z$ $(z = 0, 1)$, with $\mathcal{N}_1$, $\mathcal{N}_0$, $\mathbb{G}_{11}$, $\mathbb{G}_{01}$, $\mathbb{G}_{10}$, and $\mathbb{G}_{00}$ mutually independent with distributions as given in Proposition 1. Furthermore, $\widehat{p}(z) \xrightarrow{\mathcal{D}} p(z)$ by the law of large numbers. The claimed result follows from consecutively applying Slutsky's lemma, the delta method, and the continuous-mapping theorem (see e.g. Van der Vaart and Wellner, 1996). □

Derivation of equation (11). Note that the right-hand side of (11) equals

$$\lim_{t \downarrow 0} \frac{p(1)f_{T;S=1,Z=1}(t) - p(0)f_{T;S=1,Z=0}(t)}{[1 - p(0)]f_{T;S=0,Z=0}(t) - [1 - p(1)]f_{T;S=0,Z=1}(t)}$$

by De l'Hospital's rule, and that

$$f_{T;S=s,Z=z}(t) = \lambda(t)\gamma^s\mathbb{E}\left[V\exp\left(-\gamma^s V\int_0^t \lambda(\tau)d\tau\right)|S = s, Z = z, T \geq t\right].$$

Next, we substitute the equation below, which follows from $\mathbb{E}[V|Z = 1] = \mathbb{E}[V|Z = 0]$,

$$p(1)\mathbb{E}[V|S = 1, Z = 1] - p(0)\mathbb{E}[V|S = 1, Z = 0] = [1 - p(0)]\mathbb{E}[V|S = 0, Z = 0] - [1 - p(1)]\mathbb{E}[V|S = 0, Z = 1]$$

□

Proof of Proposition 3. The X variables only play a role in the identification of λ, ϕ and F_V from the outcomes for $Z = 0$. We therefore proceed conditional on X, subsume $\phi(X)$ into λ, and suppress X in the notation.

34

Among agents with $S = 0, Z = 1$ there holds that

$$\overline{F}_{T;S=0,Z=1}(t) = \mathcal{L}_{V;S=0,Z=1}\left(\int_0^t \lambda(\tau)d\tau\right),$$

with $\mathcal{L}_{V;S=0,Z=1}$ being the Laplace transform of $[V|S=0,Z=1]$. Its argument $\int_0^t \lambda(\tau)d\tau$ is an already identified function. Thus, $\mathcal{L}_{V;S=0,Z=1}$ is identified, and therefore $F_{V;S=0,Z=1}$.

From

$$F_V(v) = F_{V;Z=1}(v) = p(1)F_{V;S=1,Z=1}(v) + [1-p(1)]F_{V;S=0,Z=1}(v)$$

we can now also identify $F_{V;S=1,Z=1}$, since all other quantities in this equation are observed or identified. Note that this requires $p(1) > 0$, which is the present formalization of Assumption $3(i)$.

Among agents with $S = 1, Z = 1$ there holds that

$$\overline{F}_{T;S=1,Z=1}(t) = \mathcal{L}_{V;S=1,Z=1}\left(\int_0^t \lambda(\tau)\gamma(\tau)d\tau\right),$$

with $\mathcal{L}_{V;S=1,Z=1}$ being the Laplace transform of $[V|S=1,Z=1]$. Because this is already identified, and the left-hand side is observed, we can back out the argument of this Laplace transform. Since λ is also already (almost everywhere) identified, it follows that γ is (almost everywhere) identified. $\qquad\square$

Appendix 2   Proofs of statements in Section 4

Proof of Proposition 5. This proof follows as a special case of Bond and Shaw (2003). First, suppose that $\xi_S \geq \xi_T$. Then,

$$\begin{aligned}
\overline{Q}_{S;Z=1}(t) &= \Pr\left(S(1) \leq t, T(\infty,1) > S(1)\right) \\
&= \Pr\left(S(0) \leq \xi_S(t), \xi_T^{-1}(T(\infty,0)) > \xi_S^{-1}(S(0))\right) \\
&\geq \Pr\left(S(0) \leq \xi_S(t), T(\infty,0) > S(0)\right) = \overline{Q}_{S;Z=0}(\xi_S(t)),
\end{aligned}$$

$$\begin{aligned}
\overline{Q}_{T;Z=1}(t) &= \Pr\left(T(\infty,1) \leq t, T(\infty,1) < S(1)\right) \\
&= \Pr\left(T(\infty,0) \leq \xi_T(t), \xi_T^{-1}(T(\infty,0)) < \xi_S^{-1}(S(0))\right) \\
&\leq \Pr\left(T(\infty,0) \leq \xi_T(t), T(\infty,0) < S(0)\right) = \overline{Q}_{T;Z=0}(\xi_T(t)),
\end{aligned}$$

and

$$\begin{aligned}
H_{Z=0}(\xi_T(t)) \leq H_{Z=1}(t) &= \Pr(T(\infty,1) \leq t, S(1) \leq t) \\
&= \Pr\left(T(\infty,0) \leq \xi_T(t), S(0) \leq \xi_S(t)\right) \leq H_{Z=0}(\xi_S(t)).
\end{aligned}$$

Taken together, these inequalities imply that

$$\begin{aligned}
H_{Z=0}^{-1} \circ H_{Z=1} &\leq \xi_S \leq \overline{Q}_{S;Z=0}^{-1} \circ \overline{Q}_{S;Z=1} \quad\text{and} \\
\overline{Q}_{T;Z=0}^{-1} \circ \overline{Q}_{T;Z=1} &\leq \xi_T \leq H_{Z=0}^{-1} \circ H_{Z=1}.
\end{aligned}$$

Similarly, if $\xi_S \leq \xi_T$ we have that

$$\begin{aligned}
H_{Z=0}^{-1} \circ H_{Z=1} &\geq \xi_S \geq \overline{Q}_{S;Z=0}^{-1} \circ \overline{Q}_{S;Z=1} \quad\text{and} \\
\overline{Q}_{T;Z=0}^{-1} \circ \overline{Q}_{T;Z=1} &\geq \xi_T \geq H_{Z=0}^{-1} \circ H_{Z=1}.
\end{aligned}$$

$\qquad\square$

References

Abadie, A. (2002), "Bootstrap tests for distributional treatment effects in instrumental variable models", *Journal of the American Statistical Association*, 97, 284–292.

Abadie, A., J. Angrist, and G. Imbens (2002), "Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings", *Econometrica*, 70, 91–117.

Abbring, J.H. (2003), "Dynamic econometric program evaluation", Working paper, Free University, Amsterdam [Paper prepared for the H. Theil Memorial Conference, Amsterdam, 16-18 August 2002].

Abbring, J.H., P.A. Chiappori, and J. Pinquet (2003), "Moral hazard and dynamic insurance data", *Journal of the European Economic Association*, 1, 767–820.

Abbring, J.H. and G.J. van den Berg (2003a), "The identifiability of the mixed proportional hazards competing risks model", *Journal of the Royal Statistical Society Series B*, 65, 701–710.

Abbring, J.H. and G.J. van den Berg (2003b), "The non-parametric identification of treatment effects in duration models", *Econometrica*, 71, 1491–1517.

Abbring, J.H., G.J. van den Berg, and J.C. van Ours (2005), "The effect of unemployment insurance sanctions on the transition rate from unemployment to employment", *Economic Journal*, 115, forthcoming.

Andersen, P.K., Ø. Borgan, R.D. Gill, and N. Keiding (1993), *Statistical Models Based on Counting Processes*, Springer-Verlag, New York.

Bijwaard, G. (2003), "Instrumental variable estimation for duration data", Working paper, Erasmus University, Rotterdam.

Bijwaard, G. and G. Ridder (2005), "Correcting for selective compliance in a re-employment bonus experiment", *Journal of Econometrics*, 125, 77–111.

Black, D.A., J.A. Smith, M.C. Berger, and B.J. Noel (2003), "Is the threat of reemployment services more effective than the services themselves? Evidence from random assignment in the UI system", *American Economic Review*, 93, 1313–1327.

Bond, S.J. and J.E.H. Shaw (2003), "Bounds on the covariate-time transformation for competing-risks survival analysis", Working paper, University of Warwick, Coventry.

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

Cox, D.R. (1959), "The analysis of exponentially distributed life-times with two types of failure", *Journal of the Royal Statistical Society Series B*, 21, 411–421.

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

Eberwein, C., J.C. Ham, and R.J. LaLonde (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*, 64, 655–682.

Elbers, C. and G. Ridder (1982), "True and spurious duration dependence: The identifiability of the proportional hazard model", *Review of Economic Studies*, 49, 403–410.

Ferrall, C. (2002), "Estimation and inference in social experiments", Working paper, Queen's University, Kingston.

Feins, J.D., M.J. Holin, and A.A. Phipps (1996), *MTO: Moving to Opportunity for Fair Housing Demonstration. Program Operations Manual (Revised)*, Abt Associates Inc., Cambridge, MA.

Fleming, T.R. and D.P. Harrington (1991), *Counting Processes and Survival Analysis*, Wiley, New York.

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.J. (1997), "Instrumental variables: A study of implicit behavioral assumptions used in making program evaluations", *Journal of Human Resources*, 32, 441–462.

Heckman, J.J. and B.E. Honoré (1989), "The identifiability of the competing risks model", *Biometrika*, 76, 325–330.

Heckman, J.J., R.J. LaLonde, and J.A. Smith (1999), "The economics and econometrics of active labor market programs", in O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics, Volume III*, North-Holland, Amsterdam.

Heckman, J.J., J.A. Smith, and C. Taber (1998), "Accounting for dropouts in evaluations of social programs", *Review of Economics and Statistics*, 80, 1–14.

Heckman, J.J. and C.R. Taber (1994), "Econometric mixture models and more general models for unobservables in duration analysis", *Statistical Methods in Medical Research*, 3, 279–302.

Honoré, B.E. (1991), "Identification results for duration models with multiple spells or time-varying covariates", Working paper, Northwestern University, Evanston.

Imbens, G.W. and J.D. Angrist (1994), "Identification and estimation of local average treatment effects", *Econometrica*, 62, 467–475.

Imbens, G.W. and D.B. Rubin (1997), "Estimating outcome distributions for compliers in instrumental variables models", *Review of Economic Studies*, 64, 555–574.

Katz, L.F., J.R. Kling, and J.B. Liebman (2001), "Moving to Opportunity in Boston: Early results of a randomized mobility experiment", *Quarterly Journal of Economics*, 116, 607–654.

Kortram, R.A., A.J. Lenstra, G. Ridder, and A.C.M. van Rooij (1995), "Constructive identification of the mixed proportional hazards model", *Statistica Neerlandica*, 49, 269–281.

Lancaster, T. (1990), *The Econometric Analysis of Transition Data*, Cambridge University Press, Cambridge.

Lenstra, A.J. and A.C.M. van Rooij (1998), "Nonparametric estimation of the mixed proportional hazards model", Working paper, Free University, Amsterdam.

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

Meyer, B.D. (1996), "What have we learned from the Illinois reemployment bonus experiment?", *Journal of Labor Economics*, 14, 26–51.

Neyman, J. (1923), "On the application of probability theory to agricultural experiments. Essay on principles", *Roczniki Nauk Rolniczych*, 10, 1–51 [in Polish; edited and translated version of Section 9 by D.M. Dabrowska and T.P Speed (1990), *Statistical Science*, **5**, 465–472].

Orr, L., J.D. Feins, R. Jacob, E. Beecroft, L. Sanbonmatsu, L.F. Katz, J.B. Liebman, and J.R. Kling (2003), *Moving to Opportunity for Fair Housing Demonstration Program. Interim Impacts Evaluation*, U.S. Department of Housing and Urban Development, Office of Policy Development and Research, Washington, D.C.

Peterson, A.V. (1976), "Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks", *Proceedings of the National Academy of Sciences*, 73, 11–13.

Richardson, K. and G.J. van den Berg (2001), "The effect of vocational employment training on the individual transition rate from unemployment to work", *Swedish Economic Policy Review*, 8, 175–213.

Robins, J.M. and A.A. Tsiatis (1991), "Correcting for non-compliance in randomized trials using rank preserving structural failure time models", *Communications in Statistics–Theory and Methods*, 20, 2609–2631.

Rubin, D. (2000), "Discussion of causal inference without counterfactuals", *Journal of the American Statistical Association*, 95, 435–438.

Van den Berg, G.J. (2001), "Duration models: Specification, identification, and multiple durations", in J.J. Heckman and E. Leamer, editors, *Handbook of Econometrics, Volume V*, North Holland, Amsterdam.

Van den Berg, G.J. (2004), "An economic analysis of exclusion restrictions for instrumental variable estimation", Working paper, Free University Amsterdam, Amsterdam.

Van den Berg, G.J. and B. van der Klaauw (2005), "Counseling and monitoring of unemployed workers: Theory and evidence from a controlled social experiment", *International Economic Review*, forthcoming.

Van den Berg, G.J. and J.C. van Ours (1997), "Eyeballing state dependence and unobserved heterogeneity in aggregate unemployment duration data", *Research in Labor Economics*, 16, 369–394.

Van der Vaart, A.W. and J.A. Wellner (1996), *Weak Convergence and Empirical Processes*, Springer, New York.

Vaupel, J.W. and A.I. Yashin (1985), "The deviant dynamics of death in heterogeneous populations", in N.B. Tuma, editor, *Sociological Methodology 1985*, Jossey-Bass, San Francisco.