Analyzing the Effect of Dynamically Assigned Treatments Using Duration Models, Binary Treatment Models, and Panel Data Models

Jaap H. Abbring
Gerard J. van den Berg

DISCUSSION PAPER SERIES

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Analyzing the Effect of Dynamically Assigned Treatments Using Duration Models, Binary Treatment Models, and Panel Data Models

Jaap H. Abbring
*Free University of Amsterdam, Tinbergen Institute
and IZA Bonn*

Gerard J. van den Berg
*Free University of Amsterdam, Tinbergen Institute,
IFAU-Uppsala, CEPR and IZA Bonn*

ABSTRACT

Analyzing the Effect of Dynamically Assigned Treatments Using Duration Models, Binary Treatment Models, and Panel Data Models[∗]

Often, the moment of a treatment and the moment at which the outcome of interest occurs are realizations of stochastic processes with dependent unobserved determinants. Notably, both treatment and outcome are characterized by the moment they occur. In this paper, we compare different methods of inference of the treatment effect. We argue that the timing of the treatment relative to the outcome conveys useful information on the treatment effect, which is discarded in binary treatment frameworks.

Corresponding author:

Jaap H. Abbring
Department of Economics
Free University of Amsterdam
De Boelelaan 1105
1081 HV Amsterdam
The Netherlands
Email: jabbring@econ.vu.nl

---

1 Introduction

In many real-life treatment situations, the outcome of interest can be characterized by the moment at which it occurs, and both the treatment and the outcome of interest are realized at random points of time. Typical examples include the effect of training programs or punitive benefits reductions on unemployment durations, the effect of the hiring of replacement workers on strike durations, and the effect of promotions on tenure. There is an extensive methodological econometric literature on the evaluation of social programs and treatment effects, but in general this literature does not focus on duration variables as outcome variables and does not exploit the specific information in the timing of successive events (see, for instance, Heckman, LaLonde and Smith, 1999, for an overview of this literature). Empirical research on the effect of dynamic programs on dynamic outcomes has often resorted to the "timing of events" method.[1] In this approach, event-history models along the lines of Kalbfleisch and Prentice (1980), Heckman and Singer (1984), and Lancaster (1990) are used to jointly model the rate at which the treatment occurs and the rate at which the outcome occurs.[2] This approach is firmly rooted in the econometric literature on state dependence and heterogeneity (Heckman and Borjas, 1980, and Heckman, 1981), but has mostly been developed independently from the econometric treatment-effects literature. This paper aims to provide a better understanding of the differences between the various approaches, with a focus on the identification of the treatment effect. Specifically, we make a comparison between, on the one hand, the timing-of-events method, and, on the other hand, latent variable methods with binary treatment indicators and panel data methods.[3]

Our discussion of the timing-of-events approach to program evaluation builds on recent work by Abbring and Van den Berg (2003b) (henceforth AVdB), who explicitly consider the evaluation of treatment effects in a simple event-history (bivariate duration) context. In particular, they discuss the following framework.

---

[1]Examples in economics include Card and Sullivan (1988), Gritz (1993), Lillard (1993), Lillard and Panis (1996), Bonnal, Fougère and Sérandon (1997), Abbring, Van den Berg and Van Ours (1997), Richardson and Van den Berg (2001), Rosholm (2001), Van den Berg, Holm and Van Ours (2002), Brodaty (2003), Røed and Raaum (2003), and Van den Berg, Van der Klaauw and Van Ours (2004). See also Ridder (1986), Ham and LaLonde (1996), and Eberwein, Ham and LaLonde (1997) for strongly related empirical approaches.

[2]In fields like epidemiology, the use of event-history models to analyze treatment effects is widespread (see e.g. Andersen *et al.*, 1993, and Keiding, 1999).

[3]See Maddala (1983), Heckman, LaLonde and Smith (1999), Angrist and Krueger (1999), Blundell and MaCurdy (1999), Baltagi (2001), and Wooldridge (2002) for a comprehensive overview of latent variable methods and panel data methods and their applications.

Consider a subject in a certain state. After a certain stochastic amount of time, the subject leaves this state. The subject may receive a treatment at some stochastic moment before it leaves the state. The parameter of interest is the effect of the treatment on the exit rate out of the state. AVdB adopt an explicit general model framework for the distribution of the durations until treatment and outcome. It allows the duration variables to be dependent by way of dependent unobserved determinants, with each single duration having its own Mixed Proportional Hazard (MPH) model specification.[4] In addition to this, a causal effect of the realized treatment works on the exit rate out of the current state from the moment the treatment is realized onwards

AVdB demonstrate that their baseline model, including the causal treatment effect, is non-parametrically identified from single-spell duration data. This result has a number of notable aspects. First, it does not require exclusion restrictions on observed covariates, so the data need not contain a variable that affects the treatment assignment but does not affect the outcome of interest other than by way of the treatment. Exclusion restrictions are often difficult to justify. If a variable is observed by the analyst then it is often also observable to the individuals under consideration. If the variable affects the probability of treatment and the individual knows this, then he takes his value of the variable into account in determining his optimal strategy. In turn, this strategy affects the rate at which the individual leaves the state of interest. Indeed, if the individual knows that the variable is an important determinant of the treatment assignment process then he may have a strong incentive to inquire its actual value. Second, the result does not require parametric functional form assumptions on the bivariate probability distribution of the unobserved heterogeneity terms or on the duration dependence, the covariate effects, and the treatment effect. Obviously, this is a desirable property. Third, as mentioned above, the AVdB model allows for selection effects by way of unobserved determinants affecting both the treatment assignment and the outcome. In other words, it is not necessary to make a conditional independence assumption stating that the data are able to capture all systematic determinants of the process of treatment assignment so that the remaining observed variation in the treatment assignment is independent of the determinants of the outcome of interest. In applications, such an assumption may be difficult to justify, for example if the treatment assignment is carried out by case workers who use discretionary power, taking individual characteristics of the subject that are unobserved to the analyst into account.

---

[4]The MPH model is by far the most popular duration model. See Van den Berg (2001) for a survey.

Standard methods of treatment evaluation often rely on exclusion restrictions, parametric functional form assumptions on the joint distribution of the "error terms" in the model, or conditional independence assumptions to identify the treatment effect. In this sense, treatment evaluation with the AVdB model compares favorably to those methods. The aim of the present paper is to provide a better understanding of this. More precisely, we examine which information or variation in the data enables identification of the treatment effect in the AVdB model framework, by comparing the model specification and the data to those used in standard methods of treatment evaluation in the presence of selection effects.

The paper is organized as follows. In Section 2 we present the AVdB model framework. Section 3 makes the comparison to latent variable methods. Section 4 makes the comparison to panel data methods. Section 5 concludes.

2 A duration model framework with dynamically assigned treatments

Consider a subject in a certain state. After a certain amount of time, the subject leaves this state. The subject may receive a treatment at some moment before it leaves the state. We are interested in the determinants of the event of leaving the state, so the latter event is the event of interest, and the duration until this event is the variable of interest. To fix thoughts, consider an individual who is unemployed and who moves to employment at a certain point of time, and who may receive a training at some point during his spell of unemployment. For the sake of convenience, we use the term "individual" in general to denote the subject.

We normalize the point of time at which the individual enters the state to zero. The durations T_m and T_p measure the duration until the event of interest and the duration until treatment, respectively. The population that we consider concerns the inflow into the state, and the unconditional probability distributions that are defined below are distributions over this inflow into the state. Whether this is the inflow at a fixed point of calendar time or the total inflow over some range of inflow dates depends on the application at hand.

The two durations are random variables. We use t_m and t_p to denote their realizations. We assume that all individual differences in the joint distribution of T_m, T_p can be characterized by explanatory variables X, V, where X is observed and V is unobserved to the analyst. This means that any association between

T_m and T_p conditional on X, V has a causal interpretation.[5] We assume that the joint distribution of $T_m, T_p | X, V$ is absolutely continuous, so that it can be characterized in terms of the hazard rates $\theta_p(t|x, V)$ and $\theta_m(t|t_p, x, V)$ of, respectively, $T_p|x, V$ and $T_m|t_p, x, V$.[6] Here and in similar settings below, x denotes the event $\{X = x\}$, and t_p the event $\{T_p = t_p\}$.

As noted in the introduction, we are interested in the causal effect of treatment on the exit out of the current state. The treatment and the exit are characterized by the *moments* at which they occur, and we are interested in the causal effect of treatment at time t_p on the individual outcome hazard $\theta_m(t|t_p, x, V)$. To proceed, we assume that treatment only possibly affects this hazard from the moment of treatment, t_p, onwards. There is no effect of future treatment on the current outcome hazard. More precisely, for given t the individual outcome hazard $\theta_m(t|t_p, x, V)$ is the same for all future treatment times $t_p \geq t$. This level of $\theta_m(t|t_p, x, V)$ is interpreted as the baseline "no treatment" level of the outcome hazard. For $t_p < t$, $\theta_m(t|t_p, x, V)$ may be different from this baseline level. The difference is the effect of the past treatment at t_p on the hazard at $t > t_p$. AVdB call the assumption that future treatment does not affect the current individual outcome hazard a "no anticipation" assumption and show that such an assumption is crucial for identification.[7,8] It should be stressed that this assumption does not exclude that forward-looking individuals have information on properties of

---

[5]AVdB formalize these statements by presenting the model in terms of potential-outcome variables. In the present paper we suppress this for expositional convenience.

[6]For a nonnegative random (duration) variable T, the hazard rate is defined as $\theta(t) = \lim_{dt\downarrow 0} \Pr(T \in [t, t + dt)|T \geq t)/dt$. Somewhat loosely, this is the rate at which the spell is completed at t given that it has not been completed before, as a function of t. It provides a full characterization of the distribution of T (see Lancaster, 1990, and Van den Berg, 2001).

[7]In reality, there is often no information available on the degree to which an actual treatment is anticipated. Even if some anticipation cannot be ruled out, there is virtually never any information on the moment at which the individual receives information on the moment of treatment except in the special case that one knows that the moment of treatment is perfectly foreseeable by the individual. The fact that a realization of the event of interest could be due to the anticipation of a future treatment has haunted the empirical literature on treatment effects. Many standard treatment evaluation studies suffer from a potential bias due to anticipatory effects. This includes studies using "difference-in-differences" methods where one "difference" concerns a comparison between pre- and post-treatment circumstances (see Heckman, LaLonde and Smith, 1999, for an overview).

[8]Note that we attribute the entire association between T_m and T_p conditional on X, V to a causal effect of the treatment on the outcome. We do not consider reverse causal effects of the outcome on the treatment. This is natural within the current asymmetric treatment-outcome setup. Formally, it requires the additional assumption that there are no anticipatory effects of future *outcomes* on the assignment of the treatment. AVdB provide discussion.

the treatment process, and act on this information. Such information, and the induced association between the treatment and outcome processes, can be captured by the covariates X, V. Note, however, that conditioning on X, V typically also serves to control for other selection effects. Thus, we do not attach a causal interpretation to the association between T_m, T_p on the one hand and X, V on the other hand.[9]

Let $V := (V_m, V_p)'$ be a (2×1)-vector of unobserved covariates. Let $T_p \perp\!\!\!\perp V_m | x, V_p$, implying that $\theta_p(t|x, V) = \theta_p(t|x, V_p)$. Furthermore, let $T_m \perp\!\!\!\perp V_p | t_p, x, V_m$, so that $\theta_m(t|t_p, x, V) = \theta_m(t|t_p, x, V_m)$. Somewhat loosely, one may say that V_p (V_m) captures the unobserved determinants of T_p (T_m). Now let us turn to the specifications of the hazard rates $\theta_m(t|t_p, x, V_m)$ and $\theta_p(t|x, V_p)$. We adopt the following model framework,

Model 1.

$$\theta_p(t|x, V_p) \quad = \quad \lambda_p(t)\phi_p(x)V_p \tag{1}$$

$$\theta_m(t|t_p, x, V_m) \quad = \quad \lambda_m(t)\phi_m(x)\delta(t|t_p, x)^{\mathrm{I}(t>t_p)}V_m \tag{2}$$

where I denotes the indicator function, which is 1 if its argument is true and 0 otherwise.

Apart from the factor involving $\delta(t|t_p, x)$, the above hazard rates have Mixed Proportional Hazard (MPH) specifications. The function λ_i is called the "baseline hazard", because it gives the shape of the hazard rate θ_i for any given individual. The hazard rate is said to be duration dependent if its value changes over t. Positive (negative) duration dependence means that λ_i is increasing (decreasing). The factor $\phi_i(x)$ is called the "systematic part" of the hazard. In applied duration analysis, it is common to specify $\phi_i(x) = \exp(x'\beta_i)$, so that the hazard function is multiplicative in all separate elements of x.[10] Finally, the factor V_i is called the "unobserved heterogeneity" factor.[11]

The factor $\delta(t|t_p, x)^{\mathrm{I}(t>t_p)}$ captures the treatment effect. The notation used here requires some discussion. First, note that there are alternative but observationally equivalent ways of capturing this effect. For example, one may suppress

---

[9]As a consequence, we cannot predict outcomes under counterfactual treatment-assignment mechanisms. Identification of such outcomes would require additional information, such as strong functional-form assumptions, instrumental variables, or the imposition of an economic-theoretic structure on the model.

[10]In biostatistics, β_i is often called the "treatment effect" if x captures whether the subject has received a treatment at the beginning of the spell, but we avoid such confusing terminology.

[11]AVdB also analyze alternative model specifications, notably with δ depending on t, x, and on a third unobserved heterogeneity factor, say V_δ.

I$(t > t_p)$ and redefine δ such that it equals 1 if $t \leq t_p$. This is however less attractive from an expositional point of view. For example, in our setup a constant treatment effect is relatively easy to capture. The function $\delta(t|t_p, x)$ is not identified on $t \in [0, t_p]$, so by not restricting its values on this interval we create an uninteresting identification problem. In the sequel, it is tacitly understood that all statements concerning $\delta(t|t_p, x)$, including identification statements, concern $\delta(t|t_p, x)$ on $\{(t, t_p, x) \in [0, \infty)^2 \times \mathcal{X} : t > t_p\}$.

Clearly, treatment is ineffective if and only if $\delta(t|t_p, x) \equiv 1$. Now suppose $\delta(t|t_p, x)$ is equal to a constant larger than one. When T_p is realized then the level of the individual exit rate out of the current state increases by a fixed percentage. This stochastically reduces the remaining duration in that state, in comparison to the case where the treatment is given at a later point of time. More in general, we allow the treatment effect to vary with the moment of treatment t_p, with x, and with the elapsed time t in the current state. As a result, the individual effect may also vary with the time $t - t_p$ since (the onset of) treatment.

We assume that V is independent of X with distribution G. We also assume that $\mathrm{E}(V_i) < \infty$. This is a common assumption in the analysis of single-spell duration data with MPH-type models (see Heckman and Taber, 1994, and Van den Berg, 2001, for surveys). Note that we treat V as a "random effect". An alternative approach treats the individual values of V as unknown individual-specific parameters (or, equivalently, as "incidental" parameters or "fixed effects"). For convenience, we also impose a number of normalizations and regularity assumptions on the model (see AVdB). Notably, the individual values of x are taken to be time-invariant. For our results it is useful to point out that "binary treatment" analyses often define the issue of time-varying covariates away, by assuming that there is only a single point of time at which possible treatment and outcome may occur. Time-varying covariates are potentially very useful for the identification of duration models (see Honoré, 1991, and Heckman and Taber, 1994).

The model does not impose parametric functional form restrictions on the distributions of $T_p|x, V_p$ and $T_m|t_p, x, V_m$. More specifically, the model is nonparametric in the sense that we do not make parametric functional form assumptions on the probability distribution of the unobserved heterogeneity factors, the baseline hazards, the systematic hazards, and the treatment effect. Furthermore, we do not impose that there are observed explanatory variables that do affect T_p but do not affect T_m other than by way of t_p. In other words, we do not exclude elements of x from $\phi_m(x)$ that are included in $\phi_p(x)$.

It is useful to phrase the problem of the identification of the treatment effect in the presence of "selectivity", in the context of our Model 1. First, note that

the data typically provide observations on realizations of T_m and X. In addition, if treatment is initiated before the realization t_m then we also observe the treatment time t_p, otherwise we merely observe that T_p exceeds t_m. Now consider the (sub)population of individuals with a given value of x. The individuals who are observed to receive a treatment at a date t_p are not representative (in terms of their distribution of T_m, T_p, V) for this population. The most important reason for this is that the distribution of V_p among them does not equal the corresponding population distribution, because individuals with high values of V_p are more likely to have been treated before at time t_p. If V_p and V_m are dependent, then by implication the distribution of V_m among them does not equal the corresponding population distribution either. A second reason is that, in order to *observe* the fact that treatment occurs at t_p, the individual should not have left the state of interest before t_p. Because of all this, the treatment effect cannot be inferred from a direct comparison of realized durations t_m of these individuals to the realized durations of other individuals. If the individuals with a treatment at t_p have relatively short durations then this can be for two reasons: *(i)* the individual treatment effect is positive, or *(ii)* these individuals have relatively high values of V_m and would have relatively short durations anyway. The second relation is called a *spurious* relation as it is merely due to the limited observability of the set of explanatory variables. This relation is referred to as "selectivity". If V_m and V_p are independent then $\mathrm{I}(t > t_p)$ is an "ordinary" exogenous time-varying covariate for T_m, and one may infer the treatment effect from a univariate duration analysis based on the distribution of $T_m | t_p, x, V_m$ mixed over the distribution of V_m. However, in general there is no reason to assume independence of V_m and V_p, and if this possible dependence is ignored then the estimate of the treatment effect may be inconsistent.

We now examine Model 1 from a number of different angles. First, the probability $\Pr(T_m > t_m, T_p > t_p | x)$ can be expressed as

$$\Pr(T_m > t, T_p > t_p | x) =$$
$$\int_0^\infty \int_0^\infty \exp\left(-\phi_m(x)v_m\left[\Lambda_m(\min\{t, t_p\}) + \mathrm{I}(t > t_p)\Delta(t | t_p, x)\right]\right.$$
$$\left. -\phi_p(x)v_p\Lambda_p(t_p)\right) dG(v_m, v_p)$$

with

$$\Lambda_i(t) := \int_0^t \lambda_i(\tau) d\tau$$

$$\Delta(t | t_p, x) := \int_{t_p}^t \lambda_m(\tau)\delta(\tau | t_p, x) d\tau.$$

The joint density of $T_m, T_p|x$ follows from differentiation with respect to t_m and t_p. Note that Model 1 and the above include a specification of the distribution of T_p for $T_p > T_m$. This specification is immaterial, as it does not play any role in the paper or indeed in any empirical analysis.

To make comparisons to other models, it is useful to rewrite Model 1 as a regression-type model. It is well known that the integrated hazard of a duration distribution has an exponential distribution with parameter 1 (see *e.g.* Ridder, 1990, Horowitz, 1999). From this it follows that we can write

$$\log \Lambda_p(T_p) = -\log \phi_p(x) - \log V_p + \varepsilon_p \tag{3}$$

$$\begin{aligned} \log \left[\Lambda_m(\min\{T_m, t_p\}) + \mathrm{I}(T_m > t_p)\Delta(T_m|t_p, x) \right] \\ = -\log \phi_m(x) - \log V_m + \varepsilon_m, \end{aligned} \tag{4}$$

where

$$\varepsilon_m \quad \perp\!\!\!\perp \quad \varepsilon_p, \qquad \text{and} \qquad (\varepsilon_m, \varepsilon_p) \quad \perp\!\!\!\perp \quad (X, V)$$

and ε_p and ε_m have an Extreme Value – Type I (EV1) probability distribution. This distribution does not have any unknown parameters; its density equals

$$f(\varepsilon_i) = e^{\varepsilon_i} e^{-e^{\varepsilon_i}}, \qquad -\infty < \varepsilon_i < \infty.$$

Note that V_p and V_m are allowed to be dependent. Also, it is important to stress that the fact that ε_p and ε_m have a fully specified distribution does not mean that we make a parametric functional form assumption on the distributions of $T_p|x, V_p$ and $T_m|t_p, x, V_m$. This is because the left-hand sides of the above regression-type equations specify unknown transformations of the dependent variables: the integrated baseline hazard Λ_p for T_p, and a generalized integrated baseline hazard (including treatment effects depending on t_p) for T_m. Equation (3), for example, can be rewritten as

$$T_p = \Lambda_p^{-1} \left(\frac{e^{\varepsilon_p}}{\phi_p(x)V_p} \right),$$

where Λ_p^{-1} is the (generalized) inverse of Λ_p.

The fact that we specify the assignment of treatment by specifying the hazard rate of a duration distribution rather than the individual realization of the duration variable itself implies that there is a random component in the assignment that is by definition independent of all other variables. This random component is represented by the term ε_p in the "regression" equation (3) for T_p. One may interpret this random component in terms of the randomness in the treatment

8

assignment at t that remains if we take $\theta_p(t|x, V_p)$ as given. Consider an individual who has not yet been given a treatment and who has not yet left the state of interest, at time t. Basically, in a small time interval $[t, t+dt)$, the probability of treatment is $\theta_p(t|x, V_p)dt$, and the probability of no treatment is $1 - \theta_p(t|x, V_p)dt$. This is a Bernoulli trial. Given $\theta_p(t|x, V_p)dt$, its outcome is randomized. In practice, such randomization may reflect behavior of the institution that supplies or imposes the treatments, or it may reflect random external shocks.

We first discuss identification from single-spell data. The components of Model 1 that can be freely specified are $\Lambda_m, \Lambda_p, \phi_m, \phi_p, \Delta$, and G. To each specification of $\Lambda_m, \Lambda_p, \phi_m, \phi_p, \Delta, G$ corresponds a joint distribution of $T_m, \mathrm{I}(T_m > T_p)$ and $T_p \, \mathrm{I}(T_m > T_p)$ given x. Ideally, we can directly infer the latter from a large data set and in the sequel we will simply refer to it as the "data". We say that the model is "identified" under the given assumptions if no two model specifications that satisfy these assumptions generate the same data. AVdB demonstrate that Model 1 is non-parametrically identified from single-spell data. The result implies that the treatment effect is identified without exclusion restrictions or parametric functional form restrictions on the distribution of unobserved heterogeneity.

We now turn to the case where the data cover multiple spells of an individual in the state of interest, $i.e.$ if the data are "multiple-spell" data. We assume that an individual has a fixed value of V. For a given individual, the different spells provide independent draws from the joint distribution of $T_m, \mathrm{I}(T_m > T_p)$ and $T_p \, \mathrm{I}(T_m > T_p)$ given x and V. The extension to more than two spells is trivial. Also note that the setup includes cases in which physically different individuals share the same value of V and we observe one duration for each of these individuals. Such a group of individuals is usually called a "stratum".

Since V is unobserved, the duration variables are not independent across spells given x. In fact, any stochastic dependence across spells can only be due to the presence of heterogeneity. It is ruled out by assumption that realizations of T_m or T_p in one spell affect the distributions of durations in another spell. This may be a strong assumption in some applications. For example, participation in a training program for the unemployed may have an effect on the durations of future unemployment spells.

Compared to Model 1, the AVdB model for multiple spells is much more flexible. It allows for interaction between t and x in the individual hazard rates, so that the MPH structure is relaxed substantially. It also allows for dependence of V and x, and it does not require that $\mathrm{E}(V_i) < \infty$. It is allowed that the individual hazard rates in the second spell depend on t, x in a different way than they do in the first spell. The treatment effect and the systematic components may also be

different across the two spells, and x may even be completely absent. However, it still requires that the treatment effect and the unobservables affect the hazard rates multiplicatively. AVdB demonstrate that this model is non-parametrically identified from multiple-spell data.

3 Comparison to latent variable models with binary treatment assignment

We start this section by considering identification of a treatment effect in standard latent variable models with binary treatment assignment (see Maddala, 1983, and Wooldridge, 2002, for overviews). We restrict attention to the relations between the *mean* of the endogenous variables on the one hand and the explanatory variables on the other. This is in line with the spirit in which these models are interpreted.[12]

Consider

$$
\begin{aligned}
Y &= x_Y' \beta_Y + x_Y' \delta_0 \, \mathrm{I}(Z > 0) + \varepsilon_Y \\
Z &= x_Z' \beta_Z + \varepsilon_Z
\end{aligned}
\tag{5}
$$

where Y, Z, ε_Y, and ε_Z are continuous random variables, $Z > 0$ indicates treatment, $\mathrm{I}(Z > 0)$ is a binary treatment indicator, and $x_Y' \delta_0$ is the treatment effect. We assume that inference is based on a random sample of subjects with information on Y, x_Y, x_Z, and $\mathrm{I}(Z > 0)$ for each subject. Furthermore, $x := (x_Y, x_Z)$ is independent of $\varepsilon_Y, \varepsilon_Z$. We take the parameter β_Z to be identified from the data on $\mathrm{I}(Z > 0)$ and x_Z. It is generally acknowledged that either exclusion restrictions (stating that some covariates in x_Z with non-zero parameters in β_Z are not included in x_Y) or parametric functional-form assumptions on the joint distribution of $\varepsilon_Y, \varepsilon_Z$ are required for identification of the treatment-effect parameter δ_0. To illustrate this, suppose one aims to identify δ_0 from the difference in the mean outcomes of treated and untreated individuals,

$$
\begin{aligned}
\mathrm{E}(Y|Z > 0, x) &- \mathrm{E}(Y|Z \leq 0, x) = \\
x_Y' \delta_0 + \mathrm{E}(\varepsilon_Y | \varepsilon_Z > -x_Z' \beta_Z, x_Z) &- \mathrm{E}(\varepsilon_Y | \varepsilon_Z \leq -x_Z' \beta_Z, x_Z).
\end{aligned}
\tag{6}
$$

If we do not make parametric functional form assumptions on the joint distribution of $\varepsilon_Y, \varepsilon_Z$ then the difference in the conditional mean errors in the right-hand

---

[12]See *e.g.* Heckman (1990) for identification results based on full information.

side of (6) can be linear in $x'_Z \beta_Z$ on its support. If, in addition, we do not impose an exclusion restriction then obviously δ_0 is not identified from this linear expression.

Now suppose that we do not observe Y but only $\mathrm{I}(Y > 0)$. This is the "binary treatment – binary outcome" specification. For example, $Z > 0$ may indicate whether an unemployed individual has participated in a training program, and $Y > 0$ may indicate whether he has found a job. The binary specification for Y effectively reduces the information in the data, so identification needs at least as many assumptions as above (see Cameron and Heckman, 1998, for results).

Before we contrast the identification results, it should be noted that using latent variable models with binary treatment assignment or discrete-time panel data models when Y is in fact a (function of a) duration variable leads to a number of intractable practical problems. First, it requires aggregation over time. Often, it is recorded whether a treatment occurs in a baseline period, and it is recorded whether the outcome (*i.e.* the duration variable) is realized in the subsequent period. Then, the question arises what to do when the treatment and the outcome occur in the same period. In addition, it is not clear how to deal with observations that are right-censored before the end of the observation window. It is common that a spell in a state of interest can end in different ways. Usually, only a subset of these are deemed interesting. For example, an unemployment spell can end because of a transition to work, but also because of a transition into education, military service, etcetera. A study of the transition rate to work may treat transitions to other destinations as independently right-censoring the duration until work. The treatment effect estimate may be biased if such observations are discarded or if they are treated as observations of exit to work.

Now let us consider the similarities and differences between the above model and the duration model of Section 2 in its regression representation (3)–(4), and the corresponding identification results. To shape thoughts, one may, in the above latent variable model, interpret Y as $\log T_m$ and $Z > 0$ as $T_p < T_m$. Of course, alternative interpretations are possible, and each of them is imperfect. Moreover, if Model 1 is the true model then in general the parameter δ_0 in the above model specification is a complicated function of all parameters of Model 1.

The most notable similarity between the models concerns the proportionality assumptions in Model 1 and the additivity assumption in regression equations. Both impose some "smoothness" by excluding certain interactions of time and explanatory variables at the individual level. The MPH specification does impose more structure than a regression specification, because the former decomposes

the "error term" into two terms.[13]

The most fundamental difference between the data used for the latent variable model and the data used for Model 1 concerns the fact that the latter incorporate the timing of the treatment whereas the former do not. At the same time, the most fundamental difference between the identification result for the latent variable model and the identification result for Model 1 concerns the fact that the latter does not need exclusion restrictions and even allows the treatment effect to have a time dimension. This suggests that the timing of events provides potentially very useful information on the treatment effect.

We shed more light on this by returning to Model 1. A single observation is equivalent to an observation of a realization of $\min\{T_m, T_p\}, \mathrm{I}(T_m > T_p)$ and $T_m \mathrm{I}(T_m > T_p)$ given x. The pair $\min\{T_m, T_p\}, \mathrm{I}(T_m > T_p)$ is called the *identified minimum* of T_m, T_p. The data and the model can be decomposed into two parts: the identified minimum given x, and the duration from T_p until T_m if T_p is smallest. Let us examine the part of the model that specifies the distribution of this identified minimum given x. This is in fact the well-known competing risks model with dependent risks, where the dependence runs by way of related unobserved heterogeneity factors (see Heckman and Honoré, 1989, Lancaster, 1990). Note that, because of the no-anticipation assumption, the specification of the competing risks model does not depend on δ. Heckman and Honoré (1989) show that, under a number of assumptions, the dependent competing risks model is identified from single-spell data on the identified minimum and x. So if the data provide exact knowledge of the joint distribution of the identified minimum given x then $\phi_m, \phi_p, \Lambda_m, \Lambda_p$, and G can be deduced. Of course, we observe not only the identified minimum and x, but also $T_m \mathrm{I}(T_m > T_p)$. The latter, which is intimately linked to the time distance between the moment of outcome and the moment of treatment, can be used to identify $\delta(t|t_p, x)$. Intuitively, one can compare the actual distribution of this time distance to the distribution that would prevail if $\delta = 1$. The latter distribution is identified from the competing risks model.

To clarify this further, we compare the observable exit rate θ_m at time t for individuals who are treated at some time $t_p < t$,

$$\theta_m(t|T_p = t_p, x) = \lambda_m(t)\phi_m(x)\delta(t|t_p, x)\mathrm{E}(V_m|T_m \geq t, T_p = t_p, x),$$

---

[13]Because of this, identification of an MPH model requires data on the whole distribution of the duration variable given X instead of just the mean given X, and without the proportionality assumption embedded in the MPH specification the model is not identified (Van den Berg, 2001). All of this is also essential for the identification of the shape of δ as a function of t or $t - t_p$ in Model 1, which is why a comparison of identification of the treatment effect to that in the latent variable model is fairer if one restricts attention to constant δ.

to the exit rate θ_m at time t for individuals who have not yet been treated at t,

$$\theta_m(t|T_p > t, x) = \lambda_m(t)\phi_m(x)\mathrm{E}(V_m|T_m \geq t, T_p > t, x).$$

Note that

$$\log \theta_m(t|T_p = t_p, x) - \log \theta_m(t|T_p > t, x) =$$
$$\log \delta(t|t_p, x) + \log \mathrm{E}(V_m|T_m \geq t, T_p = t_p, x) - \log \mathrm{E}(V_m|T_m \geq t, T_p > t, x).$$

So, for t near t_p we have that

$$\lim_{t\downarrow t_p} \left[\log \theta_m(t|T_p = t_p, x) - \log \theta_m(t|T_p > t, x)\right] = \qquad (7)$$
$$\log \delta(t_p^+|t_p, x) + \log \mathrm{E}(V_m|T_m \geq t_p, T_p = t_p, x) - \log \mathrm{E}(V_m|T_m \geq t_p, T_p > t_p, x),$$

where $\delta(t_p^+|t_p, x) := \lim_{t\downarrow t_p} \delta(t|t_p)$ is the effect of treatment at t_p on the outcome hazard just after t_p. The selection effect $\log \mathrm{E}(V_m|T_m \geq t_p, T_p = t_p, x) - \log \mathrm{E}(V_m|T_m \geq t_p, T_p > t_p, x)$ is identified from the competing risks part of the model. In words, we know the average "quality" (V_m) among treatment and control groups at t_p from the competing risks part of the model. This enables identification of $\delta(t_p^+|t_p, x)$. A given treatment works and only works from the day the subject is exposed to its possible effects, whereas unobserved heterogeneity affects the observed exit rate out of the current state from the start.[14]

The above results based on the partitioning of model and data reinforce the conclusions from the comparison between the binary treatment approach and the duration model approach. First of all, the timing of events conveys useful information on the treatment effect. Second, the MPH specification is important, because this is assumed for the identification of the competing risks model.[15]

4 Comparison to panel data models

Consider the dynamic panel data model

$$W_t = x'_{W,t}\beta_W + x'_{W,t}\delta_0 \, \mathrm{I}(Z_t > 0) + V_W + \epsilon_{W,t}$$
$$Z_t = x'_{Z,t}\beta_Z + V_Z + \epsilon_{Z,t}, \qquad (8)$$

---

[14]Econometric evaluation approaches typically require some independent variation in the assignment of treatment for identification of a treatment effect. In our framework, this is taken care of by the random component ε_p in the "regression" equation (3) for T_p, which only affects T_m by way of the treatment.

[15]See Abbring and Van den Berg (2003a) for intuition behind the identification of the competing risks model.

where the index t denotes time. We deliberately introduce new notation W_t for the outcome because below we adopt two different interpretations of W_t in terms of duration outcomes. Inference is based on a random sample of subjects with information on W_t, W_{t-1}, $x_{W,t}$, $x_{W,t-1}$, $x_{Z,t}$, $x_{Z,t-1}$, $I(Z_t > 0)$, and $I(Z_{t-1} > 0)$ for each subject. We take $E[\epsilon_{W,t-1}|Z_{t-1}, Z_t, x_{W,t-1}, x_{W,t}] = E[\epsilon_{W,t}|Z_{t-1}, Z_t, x_{W,t-1}, x_{W,t}] = 0$. Thus, the spurious dependence between treatment and outcome (the "selection effect") runs by way of the relation between V_W and V_Z. We can treat V_W as a fixed effect and control for it by taking first differences of W_t across time,

$$E[W_t - W_{t-1}|Z_{t-1} \leq 0, Z_t > 0, x_{W,t-1}, x_{W,t}] = (x_{W,t} - x_{W,t-1})'\beta_W + x'_{W,t}\delta_0. \quad (9)$$

Clearly, δ_0 is identified (under standard support conditions on the regressors).

Now suppose that β_W varies over time. The equivalent of the right-hand side of equation (9) equals $x'_{W,t}\beta_{W,t} - x'_{W,t-1}\beta_{W,t-1} + x'_{W,t}\delta_0$. As a result, δ_0 is unidentified from $E[W_t - W_{t-1}|Z_{t-1} \leq 0, Z_t > 0, x_{W,t-1}, x_{W,t}]$. A "difference-in-differences" however gives

$$\begin{aligned} E[W_t - W_{t-1}|Z_{t-1} \leq 0, Z_t > 0, x_{W,t-1}, x_{W,t}] \\ - E[W_t - W_{t-1}|Z_{t-1} \leq 0, Z_t \leq 0, x_{W,t-1}, x_{W,t}] = x'_{W,t}\delta_0, \end{aligned} \quad (10)$$

so δ_0 is again identified. Note that this does not require the specification of a model equation for the treatment assignment Z_t. Moreover, Z_t and $X_{W,t}$ may be dependent, and the method even works in the absence of observed covariates. Also, clearly, δ_0 can be allowed to depend on t.

We can relate the dynamic panel data model to our model in two ways. We can either (i) let W_t be a survival indicator for a single outcome duration T_m and set $W_t = 1 \iff T_m < t$, or (ii) relate each W_t to a separate outcome duration, say $T_{m,t}$.

Option (i) suggests a link between our single-spell duration model and a multi-period panel data model. In fact, because a duration variable has a time unit, single-spell duration models are often thought to be similar to multi-period panel data models. However, this apparent similarity is deceptive. In the case of (i), we only observe one duration outcome per individual, so $W_{t-1} = 1 \implies W_t = 1$, which is not necessarily so in the general case of specification (8). The latter specification allows for more variation in W_t and W_{t-1} given $Z_{t-1}, Z_t, x_{W,t-1}, x_{W,t}$. This variation distinguishes panel data models from our duration model. As a result, we cannot apply (9) and, in particular, we cannot apply difference-in-differences as in (10).

Nevertheless, the results at the end of the previous section (see equation (7)) suggest that an approach in the spirit of the difference-in-differences approach

might be possible, provided that one can usefully compare different sets of individuals across time, and provided that the treatment assignment process is specified. In Abbring and Van den Berg (2002) we develop such an approach. Not surprisingly, it is much more involved than straightforward difference-in-differences.

The point of departure in Abbring and Van den Berg (2002) is the observation that difference-in-differences amounts to an examination of a specific interaction effect: the extent to which the outcome over time differs between treated and controls. Abbring and Van den Berg (2002) focus on the log rate at which a treatment is given conditional on the moment of exit ($\log \theta_p(t|x, t_m)$ with $t < t_m$) and examine whether this behaves differently as a function of t when t_m is different. Intuitively, if $\delta > 1$ then many of those who "die" at t_m received a treatment shortly before t_m, so $\theta_p(t|x, t_m)$ will tend to increase shortly before t_m, relative to $\theta_p(t|x, t_m)$ for larger t_m. This amounts to examining the interaction effect between the moment of treatment t and the outcome t_m in $\log \theta_p(t|x, t_m)$. It turns out that this approach allows one to distinguish between a causal treatment effect and selectivity. Intuitively, if treatment and outcome are typically realized very quickly after each other, no matter what the values of the other outcome determinants are, then this is evidence of a positive causal treatment effect. The selection effect does not give rise to the same type of quick succession of events. The x variables play no role here. However, δ is assumed to be constant over t and t_p.

This result again illustrates the usefulness of the information on the timing of events to assess the treatment effect. Both in the panel data approach and in our duration model approach, the treatment effect works from a specific point of time onwards, whereas the selection effect works at all points of time in a more permanent way. In both approaches, separability assumptions, ruling out certain interaction effects of the determinants of the individual outcome of interest, are needed. In the panel data approach, additivity of treatment effect and unobserved heterogeneity in the outcome equation (8) is crucial. The previous section argued that in the duration model approach the additivity of the determinants of the individual log outcome hazard rate $\log \theta_m(t|t_p, x, V_m)$ (equation (2)) is crucial. The results in this section so far emphasize that what is particularly crucial is the additivity of the treatment effect and the unobserved heterogeneity in this log hazard rate (although this by itself seems to be insufficient to identify the whole duration model including the way the treatment effect varies with t and t_p).

These separability assumptions at the individual level enable an empirical

distinction between the treatment effect, which works from a specific point of time, and the selection effect, which works at all points of time. The duration approach is much more involved because of the dynamic nature of selectivity in duration analysis: as time proceeds, the composition of the survivors changes, so the selection effect changes.

Now consider option (ii) for relating the dynamic panel data model to our duration model framework. Let Z_t indicate whether a treatment is given during the t-th spell. This interpretation allows for a straightforward comparison with our multiple-spell duration model. In both cases we observe multiple outcomes for an individual with the same unobserved covariates. This suggests that we can remove the role of unobserved heterogeneity in multiple-spell data by some sort of conditional likelihood or first-differencing approach. AVdB discuss an analogue of the fixed-effect panel data estimator based on (10). Neither estimator requires any observed explanatory variables or a specification of the treatment assignment process.

5   Conclusions

Variation in the duration until treatment relative to the duration until the outcome of interest conveys useful information on the causal treatment effect in the presence of selection effects. This information is discarded in a binary treatment framework. Analysis of the duration variables allows for inference on causal treatment effects if no valid instruments are available and if conditional independence assumptions can not be justified. With single-spell duration data, this works as follows. If treatment and outcome are typically realized very quickly after each other, no matter what the values of the other outcome determinants are, then this is evidence of a positive causal treatment effect. The selection effect does not give rise to the same type of quick succession of events.

We make a number of qualifications. First, it is pivotal that individuals do not anticipate the realization of the moment of treatment, because then the treatment works from a moment in time that precedes the actual participation. It is obvious that this would lead to incorrect inference. See AVdB for an extensive methodological analysis of this issue. Second, the information in the timing of events is useful for inference on how the causal treatment effect varies with time and with the time since treatment. This provides potentially very useful insights into the workings of the treatment, and this is important from a policy point of view. Third, an important topic for further research is the robustness of the identification results for the single-spell duration model framework to relaxation

of the separability assumptions embedded in this framework.

The results lead to some suggestions for empirical work. First, it is useful not to discard information on the timing of the treatment, and, in particular, not to aggregate such information into a binary treatment indicator. Second, it is potentially useful to exploit multiple-spell data, as this leads to inference under much weaker assumptions than single-spell data.

References

Abbring, J.H. and G.J. van den Berg (2002), "A simple procedure for inference on treatment effects in duration models", Working paper, Free University, Amsterdam.

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*, forthcoming.

Abbring, J.H., G.J. van den Berg, and J.C. van Ours (1997), "The effect of unemployment insurance sanctions on the transition rate from unemployment to employment", Working paper, Tinbergen Institute, Amsterdam.

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

Angrist, J.D. and A.B. Krueger (1999), "Empirical strategies in labor economics", in O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics, Volume III*, North-Holland, Amsterdam.

Baltagi, B.H. (2001), *Econometric Analysis of Panel Data*, Wiley, New York.

Blundell, R. and T. MaCurdy (1999), "Labor supply", in O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics, Volume III*, North-Holland, Amsterdam.

Bonnal, L., D. Fougère, and A. Sérandon (1997), "Evaluating the impact of French employment policies on individual labour market histories", *Review of Economic Studies*, 64, 683–713.

Brodaty, T. (2003), "Do (and how) active labor market programs improve the reemployment process? A multiple programs and outcomes approach", Working paper, INSEE, Paris.

Cameron, S.V. and J.J. Heckman (1998), "Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males", *Journal of Political Economy*, 106, 262–333.

Card, D. and D. Sullivan (1988), "Measuring the effect of subsidized training programs on movements in and out of employment", *Econometrica*, 56, 497–530.

Carneiro, P., K.T. Hansen, and J.J. Heckman (2003), "Estimating distribution of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice", Working Paper 9546, National Bureau of Economic Research, Cambridge, MA.

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.

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

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. (1981), "Heterogeneity and state dependence", in S. Rosen, editor, *Studies in Labor Markets*, University of Chicago Press.

Heckman, J.J. (1990), "Varieties of selection bias", *American Economic Review*, 80 (supplement), 313–318.

Heckman, J.J. and G.J. Borjas (1980), "Does unemployment cause future unemployment? Definitions, questions and answers from a continuous time model of heterogeneity and state dependence", *Economica*, 47, 247–283.

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. and B. Singer (1984), "Econometric duration analysis", *Journal of Econometrics*, 24, 63–132.

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.

Horowitz, J.L. (1999), "Semiparametric estimation of a proportional hazard model with unobserved heterogeneity", *Econometrica*, 67, 1001–1029.

Kalbfleisch, J.D. and R.L. Prentice (1980), *The Statistical Analysis of Failure Time Data*, Wiley, New York.

Keiding, N. (1999), "Event history analysis and inference from observational epidemiology", *Statistics in Medicine*, 18, 2353–2363.

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

Lillard, L.A. (1993), "Simultaneous equations for hazards", *Journal of Econometrics*, 56, 189–217.

Lillard, L.A. and C.W.A. Panis (1996), "Marital status and mortality: The role of health", *Demography*, 33, 313–327.

Maddala, G.S. (1983), *Limited-dependent and qualitative variables in econometrics*, Cambridge University Press, Cambridge.

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.

Ridder, G. (1986), "An event history approach to the evaluation of training, recruitment and employment programmes", *Journal of Applied Econometrics*, 1, 109–126.

Ridder, G. (1990), "The non-parametric identification of generalized accelerated failure-time models", *Review of Economic Studies*, 57, 167–182.

Røed, K. and O. Raaum (2003), "The effect of programme participation on the transition rate from unemployment to employment", Working paper, University of Oslo, Oslo.

Rosholm, M. (2001), "An analysis of the processes of labour market exclusion and (re-)inclusion", Working paper, University of Aarhus, Aarhus.

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., A. Holm, and J.C. van Ours (2002), "Do stepping-stone jobs exist? Early career paths in the medical profession", *Journal of Population Economics*, 15, 647–665.

Van den Berg, G.J., B. van der Klaauw, and J.C. van Ours (2004), "Punitive sacntions and the transition rate from welfare to work", *Journal of Labor Economics*, 22, forthcoming.

Wooldridge, J. (2002), *Econometric Analysis of Cross Section and Panel Data*, MIT Press, Cambridge, MA.

IZA Discussion Papers

No.	Author(s)	Title	Area	Date
817	T. M. Andersen T. T. Herbertsson	Measuring Globalization	2	07/03
818	J. Pencavel	The Surprising Retreat of Union Britain	3	07/03
819	M. Beine F. Docquier H. Rapoport	Brain Drain and LDCs' Growth: Winners and Losers	1	07/03
820	C. M. Cornwell K. H. Lee D. B. Mustard	The Effects of Merit-Based Financial Aid on Course Enrollment, Withdrawal and Completion in College	6	07/03
821	P. Carneiro J. J. Heckman	Human Capital Policy	6	07/03
822	D. Weichselbaumer R. Winter-Ebmer	The Effects of Competition and Equal Treatment Laws on the Gender Wage Differential	6	07/03
823	A. Filippin	Discrimination and Workers' Expectations	5	07/03
824	A. Filippin	Discrimination and Workers' Expectations: Experimental Evidence	5	07/03
825	A. Filippin A. Ichino	Gender Wage Gap in Expectations and Realizations	5	07/03
826	K. T. Hansen J. J. Heckman K. J. Mullen	The Effect of Schooling and Ability on Achievement Test Scores	6	07/03
827	H. Buddelmeyer E. Skoufias	An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA	6	07/03
828	D. Checchi T. Jappelli	School Choice and Quality	3	07/03
829	J. J. Heckman X. Li	Selection Bias, Comparative Advantage and Heterogeneous Returns to Education: Evidence from China in 2000	6	07/03
830	T. J. Hatton	Emigration from the UK, 1870-1913 and 1950-1998	1	07/03
831	J. H. Abbring G. J. van den Berg	Analyzing the Effect of Dynamically Assigned Treatments Using Duration Models, Binary Treatment Models, and Panel Data Models	6	07/03

An updated list of IZA Discussion Papers is available on the center's homepage www.iza.org.