IFAU – INSTITUTE FOR
LABOUR MARKET POLICY
EVALUATION

Dynamically assigned treatments: duration models, binary treatment models, and panel data models

Jaap H. Abbring
Gerard J. van den Berg

WORKING PAPER 2002:20

The Institute for Labour Market Policy Evaluation (IFAU) is a research institute under the Swedish Ministry of Industry, Employment and Communications, situated in Uppsala. IFAU's objective is to promote, support and carry out: evaluations of the effects of labour market policies, studies of the functioning of the labour market and evaluations of the labour market effects of measures within the educational system. Besides research, IFAU also works on: spreading knowledge about the activities of the institute through publications, seminars, courses, workshops and conferences; creating a library of Swedish evaluational studies; influencing the collection of data and making data easily available to researchers all over the country.

IFAU also provides funding for research projects within its areas of interest. There are two fixed dates for applications every year: April 1 and November 1. Since the researchers at IFAU are mainly economists, researchers from other disciplines are encouraged to apply for funding.

IFAU is run by a Director-General. The authority has a traditional board, consisting of a chairman, the Director-General and eight other members. The tasks of the board are, among other things, to make decisions about external grants and give its views on the activities at IFAU. Reference groups including representatives for employers and employees as well as the ministries and authorities concerned are also connected to the institute.

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

ISSN 1651-1166

Dynamically Assigned Treatments: Duration Models, Binary Treatment Models, and Panel Data Models

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

August 30, 2002

Abstract

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. We compare different methods of inference of the treatment effect, and 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.

*Department of Economics, Free University Amsterdam, and Department of Economics, University College London

†Department of Economics, Free University Amsterdam, De Boelelaan 1105, NL–1081 HV Amsterdam, The Netherlands, Tinbergen Institute, IFAU-Uppsala, and CEPR.

Keywords: program evaluation, treatment effects, bivariate duration analysis, selection bias, hazard rate, unobserved heterogeneity, fixed effects, random effects.
JEL codes: C14, C31, C41.

Thanks to participants at the Tenth Panel Data Conference in Berlin, 2002, in particular to our discussant Bo Honoré, for helpful comments.

1 Introduction

In virtually every real-life treatment situation, both the treatment and the outcome of interest are realized at specific points in 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. An extensive literature on the evaluation of social programs and treatment effects exists, but in general this literature does not address nor exploit the specific information on the timing of events (see, for instance, Heckman, LaLonde and Smith, 1999, for an overview of this literature).

Abbring and Van den Berg (2002) (henceforth AVdB) advance on this literature by explicitly considering the evaluation of treatment effects in a "duration variables" context. 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. (In fact, they present the model in terms of counterfactual variables, but in the present paper we suppress this for expositional convenience.) 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. 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. The MPH model is by far the most popular duration model (see Van den Berg, 2001, for a survey). Models fitting into the AVdB model framework have been estimated by, for example, 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), and Van den Berg, Van der Klaauw and Van Ours (2004).

AVdB demonstrate that their baseline model, including the causal treatment effect, is non-parametrically identified from single-spell duration data, that is, from a random sample of subjects flowing into the state of interest and having subsequently been followed over time. 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 that he may be subject to treatment, then he takes his value of the variable into account to determine his optimal strategy, and 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 the actual value of the variable. Secondly, 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. Thirdly, 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 into account that are unobserved to the analyst.

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. Specifically, we make a comparison to latent variable methods with binary treatment indicators and to panel data methods (see e.g. Wooldridge, 2002, for a textbook overview).

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 in the inflow into the state. Whether this is the inflow at a fixed point of calendar time or the total inflow over time (or the inflow at another 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, for a given individual in the population, the duration variables are absolutely continuous random variables. 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. Of course, the joint distribution of $T_m, T_p|X, V$ can be expressed in terms of the distributions of $T_p|X = x, V$ and $T_m|T_p = t_p, X = x, V$. The latter distributions are in turn characterized by their hazard rates $\theta_p(t|x, V)$ and $\theta_m(t|t_p, x, V)$, respectively.[1]

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 effect of the realization of T_p on the distribution of T_m. To proceed, we assume that,

[1]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). Consider the distribution of a duration variable conditional on some other variables. It is customary to use a vertical "conditioning line" within the argument of a hazard rate in order to distinguish between (on the left-hand side) the value of the duration variable at which the hazard rate is evaluated, and (on the right-hand side) the variables that are conditioned upon.

conditional on X, V, the relation between T_m and T_p is characterized as follows: the realization of T_p affects the shape of the hazard of T_m from t_p onwards, in a deterministic way. This implies that the causal effect is captured by the effect of t_p on $\theta_m(t|t_p, x, V)$ for $t > t_p$. It is ruled out that t_p affects $\theta_m(t|t_p, x, V)$ on $t \in [0, t_p]$. The latter could be called a "no anticipation" assumption, because it basically means that the individual's behavior before the moment of treatment does not depend on the future realization of the moment of treatment. AVdB show that such an assumption is crucial for identification.[2]

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.

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

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

[2] 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 unless the moment of treatment is fully predictable at an individual level. 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).

Now suppose that the determinants of the stochastic process of treatment assignment affect the individual's exit rate out of the state of interest before the actual realization of the treatment. Then the treatment program is said to have an *ex ante* effect on exit out of the state of interest. Such an effect is to be expected in well-established programs. The ex ante effect should not be confused with anticipation of the *realization* of the process of treatment assignment, because in the latter case the individual knows the stochastic outcome rather than the determinants of the process. The ex ante effect can be contrasted to the *ex post* effect of treatment, which is the effect of a realized treatment on the individual exit rate – this is of course the effect we focus on in this paper. We do not deal with ex ante effects of treatment. Identification would require additional information, such as strong functional-form assumptions, instruments for a comparison of a world with a treatment program to a world without the program, or the imposition of an economic-theoretic structure on the model.

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

Apart from the term involving $\delta(t|t_p, x)$, the above hazard rates have Mixed Proportional Hazard (MPH) specifications. The function $\lambda_i(t)$ is called the "baseline hazard" since 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 $\lambda_i(t)$ increases (decreases). The term $\phi_i(x)$ 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. In biostatistics, β 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. Finally, the term V_i is called the "unobserved heterogeneity term". (AVdB also analyze alternative model specifications, notably with δ depending on t, x, and on a third unobserved heterogeneity term, say V_δ.)

The term $\delta(t|t_p, x)^{\mathrm{I}(t>t_p)}$ 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 $\mathrm{I}(t > t_p)$ and redefine $\delta(.)$ such that it is zero 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 silently understood that all statements concerning $\delta(t|t_p, x)$, including identification statements, concern $\delta(t|t_p, x)$ on $\{(t, t_p, x) \in [0, \infty)^2 \times \mathcal{X} : t > t_p\}$. It is useful to separate t from the other arguments of $\delta(.)$ by way of a vertical "conditioning line", because we will occasionally integrate over t.

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. If T_p is realized then the level of the individual exit rate out of the current state increases by a fixed amount. 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.

For convenience, we make a number of normalizations and regularity assumptions on the determinants of 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).

A more substantive assumption is that X is independent from V_m, V_p. 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).

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 T_p is completed before the realization t_m then we also observe the realization 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 a non-random subset from 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 most individuals with high values of V_p have already had the treatment before. 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 for why the individuals who are observed to receive a treatment at a date t_p are a non-random subset 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: *(1)* the individual treatment effect is positive, or *(2)* these individuals have relatively high values of V_m and would have found a job relatively fast 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.

The model specification does not impose parametric functional form restric-

tions on the way in which the distributions of $T_p|x, V_p$ and $T_m|t_p, x, V_m$ (or their properties) depend on their determinants. More specifically, the model specification is nonparametric in the sense that we do not make parametric functional form assumptions on the probability distribution of the unobserved heterogeneity terms, the baseline hazards, the systematic hazards, and the treatment effect.

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 $\phi_p(x)$.

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

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

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 $T_p > T_m$, but 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

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

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

where 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

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

Equation (3) states that the random variable $T_p|x$ is distributed as the sum of the terms on the right-hand side, where V_p and ε_p are random variables. Equation (4) states that the random variable $T_m|t_p, x$ is distributed as the sum of the terms on the right-hand side, where V_m and ε_m are random variables. There holds that

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

but V_p and V_m are allowed to be dependent. 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. A regression equation for, say, $T_p|x, V_p$ states that T_p equals $\Lambda_p^{-1}(\exp(-\log\phi_p(x) - \log V_p + \varepsilon_p))$, where $\Lambda_p^{-1}(.)$ is the inverse of $\Lambda_p(.)$.

The fact that we specify the assignment of treatment by way of specifying the hazard rate of a duration distribution (rather than by way of specifying the individual realization of the duration variable) 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 of the outcome of T_p at $t|T_p \geq t$ that remains if the hazard rate at t is specified. 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 the value of $\theta_p(t|x, V_p)dt$, its outcome is completely random. In practice, such randomness may reflect behavior of the institution that supplies or imposes the treatments, or it may reflect purely random external shocks.

Recall that the data ideally provide i.i.d. observations on realizations of $T_m, \mathrm{I}(T_m > T_p)$, and $T_p \cdot \mathrm{I}(T_m > T_p)$, given x. The latter term indicates that if T_p is completed before the realization of T_m then we also observe the realization of T_p, whereas otherwise we merely observe that the realization of T_p exceeds the realization of T_m. To analyze the identification we assume that the data actually provide exact knowledge of the whole joint distribution of $T_m, \mathrm{I}(T_m > T_p)$, and $T_p \cdot \mathrm{I}(T_m > T_p)$, given x.

Identifiability is a property of the mapping from the model determinants ϕ_i, Λ_i $(i = m, p)$, δ and G to the data as summarized in the joint distribution of

$T_m, \mathrm{I}(T_m > T_p)$, and $T_p \cdot \mathrm{I}(T_m > T_p)$, given x. The model specification defines the unique mapping from the domain to the data. The model is non-parametrically identified if this mapping has an inverse, *i.e.* if for given data there is a unique set of functions ϕ_i, Λ_i $(i = m, p)$, δ and G in the domain that is able to generate these data (of course, these data must be in the image of the mapping).

Note that this approach treats the individual-specific unobserved heterogeneity terms as realizations of random variables. In terms of panel data analysis, this means that V_m, V_p are treated as "random effects" when estimating the model with, for example, Maximum Likelihood. An alternative approach treats the individual values of V_m, V_p as unknown individual-specific parameters (or, equivalently, as "incidental" parameters or "fixed effects").

AVdB demonstrate that Model 1 (as characterized by the functions ϕ_m, ϕ_p, Λ_m, Λ_p, δ, and G) 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_m, V_p. For a given individual, the different spells provide independent drawings from the joint distribution of the $T_m, \mathrm{I}(T_m > T_p)$, and $T_p \cdot \mathrm{I}(T_m > T_p)$, given x, V_m and V_p. 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_m, V_p and we observe one duration for each of these individuals. Such a group of individuals is usually called a stratum.

Since V_m and V_p are unobserved, the duration variables given x are not independent across spells. 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_m, V_p in the inflow on x, and it does not require $\mathrm{E}(V) < \infty$ anymore. 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 size of the treatment effect may also be different across the two spells. The individual values of x in

may also differ across 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 (see *e.g.* Heckman (1990) for identification results based on full information).

Consider

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

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 between the means of $[Y|Z > 0, x]$ and $[Y|Z \leq 0, x]$, as a function of $x := x_Y, x_Z$. We have that

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

If we do not make parametric functional form assumptions on the joint distribution of $\varepsilon_Y, \varepsilon_Z$ then the sum of the second and third term on the right-hand side of

(6) can be linear in $x'_Z\beta_Z$. If in addition we do not make 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.

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 while the latter 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 $T_m \cdot \mathrm{I}(T_m > T_p)$ 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 the first to be realized. 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 terms (see Heckman and Honoré, 1989, Lancaster, 1990). Note that the specification of the competing risks model does not depend on δ, which is intuitively obvious. 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 more than the identified minimum and x, namely $T_m \cdot \mathrm{I}(T_m > T_p)$. 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 rates θ_m at t of those who are treated exactly at t to those who are not yet treated at t:

$$\log \theta_m(t|T_p = t, x) - \log \theta_m(t|T_p > t, x)$$

There holds that

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

so, as a result,

$$\log\theta_m(t|T_p=t,x)-\log\theta_m(t|T_p>t,x)= \tag{7}$$
$$\log\delta(t|t,x)+\log \mathrm{E}(V_m|T_m\geq t,T_p=t,x)-\log \mathrm{E}(V_m|T_m\geq t,T_p>t,x)$$

The selection effect $\log \mathrm{E}(V_m|T_m \geq t, T_p = t, x) - \log \mathrm{E}(V_m|T_m \geq t, T_p > t, x)$ is identified from the competing risks part of the model, that is, from information on events in $[0, t)$. In words, we know the average "quality" (V_m) among treatment and control groups at t from the competing risks part of the model. This enables identification of $\delta(t|t, x)$. 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 day 1.[3]

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. Secondly, the MPH specification is important, because this is assumed for the identification of the competing risks model (see also Abbring and Van den Berg, 2001a, for intuition behind the identification of the competing risks model).

4 Comparison to panel data models

Consider the dynamic panel data model

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

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}, $x_{W,t}$, $x_{W,t-1}$, $x_{Z,t}$, $x_{Z,t-1}$, $\mathrm{I}(Z_t > 0)$ and $\mathrm{I}(Z_{t-1} > 0)$ for each subject. We take $\epsilon_{j,t}$ to be i.i.d. across time and across $j = W, Z$, so that

[3] In most econometric evaluation approaches, a treatment effect can only be identified if some of the variation in the assignment of treatment is not fully collinear with the variation in the other determinants of the outcome of interest. In our framework, this is taken care of by the random component ε_p in the "regression" equation (3) for T_p. The variation in ε_p affects T_m only by way of the treatment.

the spurious dependence between treatment and outcome (the "selection effect") runs by way of the relation between V_W and V_Z. We take $\mathrm{E}[\epsilon_{j,t}|x_j] = 0$, where for sake of brevity we use $x_j := (x_{j,t}, x_{j,t-1})$. We can treat V_W as a fixed effect by taking first differences of W_t across time,

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

Clearly, δ_0 is identified.

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 $\mathrm{E}[W_t - W_{t-1}|Z_{t-1} \leq 0, Z_t > 0, x_W, x_Z]$. A "difference-in-differences" however gives

$$\begin{aligned}\mathrm{E}[W_t - W_{t-1}|Z_{t-1} \leq 0, &Z_t > 0, x_W, x_Z] \\ &- \mathrm{E}[W_t - W_{t-1}|Z_{t-1} \leq 0, Z_t \leq 0, x_W, x_Z] = x'_{W,t}\delta_0.\end{aligned} \tag{10}$$

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 translate the dynamic panel data model towards 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 $W_{t-1} = 1 \Longrightarrow W_t = 1$, 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, x_Z. This variation distinguishes panel data models from our duration model. As a result, we cannot apply (9) and, in particular, we cannot do difference-in-differences like 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 (2001b) 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 (2001b) is the observation that difference-in-differences amounts to an examination of a specific interaction term: the extent to which the outcome over time differs between treated and controls. Abbring and Van den Berg (2001b) 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 term 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 in 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 treatment effect and 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 effects varies with t and t_p).

These separability assumptions at the individual level enable an empirical distinction between the treatment effect, that works from a specific point of time, and the selection effect that 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 the translation of the dynamic panel data model towards 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 develop an analogue of the fixed-effect panel data estimator based on (10). Both estimators do not require 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 taken as 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. Secondly, 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. Thirdly, it is an important topic for further research to investigate to what extent the identification results for the single-spell duration model framework are robust with respect to separability assumptions embedded in the model 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 round-off such information into a binary treatment indicator. Secondly, 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 (2001a), "The identifiability of the mixed proportional hazards competing risks model", Working paper, Free University Amsterdam, Amsterdam.

Abbring, J.H. and G.J. van den Berg (2001b), "A simple procedure for inference on treatment effects in duration models", Working paper, Free University Amsterdam, Amsterdam.

Abbring, J.H. and G.J. van den Berg (2002), "The non-parametric identification of treatment effects in duration models", Working paper, Free University Amsterdam, Amsterdam.

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.

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.

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.

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

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

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 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.

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.

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

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., B. van der Klaauw, and J.C. van Ours (2004), "Punitive sanctions 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.

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

Rapport (some of the reports are written in English)

2002:1 Hemström Maria & Sara Martinson ”Att följa upp och utvärdera arbetsmarknadspolitiska program”

2002:2 Fröberg Daniela & Kristian Persson ”Genomförandet av aktivitetsgarantin”

2002:3 Ackum Agell Susanne, Anders Forslund, Maria Hemström, Oskar Nordström Skans, Caroline Runeson & Björn Öckert “Follow-up of EU’s recommendations on labour market policies”

2002:4 Åslund Olof & Caroline Runeson “Follow-up of EU’s recommendations for integrating immigrants into the labour market”

2002:5 Fredriksson Peter & Caroline Runeson “Follow-up of EU’s recommendations on the tax and benefit systems”

2002:6 Sundström Marianne & Caroline Runeson “Follow-up of EU’s recommendations on equal opportunities”

2002:7 Ericson Thomas ”Individuellt kompetenssparande: undanträngning eller komplement?”

2002:8 Calmfors Lars, Anders Forslund & Maria Hemström ”Vad vet vi om den svenska arbetsmarknadspolitikens sysselsättningseffekter?”

2002:9 Harkman Anders ”Vilka motiv styr deltagandet i arbetsmarknadspolitiska program?”

2002:10 Hallsten Lennart, Kerstin Isaksson & Helene Andersson ”Rinkeby Arbetscentrum – verksamhetsidéer, genomförande och sysselsättningseffekter av ett projekt för långtidsarbetslösa invandrare”

2002:11 Fröberg Daniela & Linus Lindqvist ”Deltagarna i aktivitetsgarantin”

Working Paper

2002:1 Blundell Richard & Costas Meghir “Active labour market policy vs employment tax credits: lessons from recent UK reforms”

2002:2 Carneiro Pedro, Karsten T Hansen & James J Heckman “Removing the veil of ignorance in assessing the distributional impacts of social policies”

2002:3 Johansson Kerstin “Do labor market programs affect labor force participation?”

2002:4 Calmfors Lars, Anders Forslund & Maria Hemström “Does active labour market policy work? Lessons from the Swedish experiences”

2002:5 Sianesi Barbara "Differential effects of Swedish active labour market programmes for unemployed adults during the 1990s"

2002:6 Larsson Laura "Sick of being unemployed? Interactions between unemployment and sickness insurance in Sweden"

2002:7 Sacklén Hans "An evaluation of the Swedish trainee replacement schemes"

2002:8 Richardson Katarina & Gerard J van den Berg "The effect of vocational employment training on the individual transition rate from unemployment to work"

2002:9 Johansson Kerstin "Labor market programs, the discouraged-worker effect, and labor force participation"

2002:10 Carling Kenneth & Laura Larsson "Does early intervention help the unemployed youth?"

2002:11 Nordström Skans Oskar "Age effects in Swedish local labour markets"

2002:12 Agell Jonas & Helge Bennmarker "Wage policy and endogenous wage rigidity: a representative view from the inside"

2002:13 Johansson Per & Mårten Palme "Assessing the effect of public policy on worker absenteeism"

2002:14 Broström Göran, Per Johansson & Mårten Palme "Economic incentives and gender differences in work absence behavior"

2002:15 Andrén Thomas & Björn Gustafsson "Income effects from market training programs in Sweden during the 80's and 90's"

2002:16 Öckert Björn "Do university enrollment constrainsts affect education and earnings? "

2002:17 Eriksson Stefan "Imperfect information, wage formation, and the employability of the unemployed"

2002:18 Skedinger Per "Minimum wages and employment in Swedish hotels and restaurants"

2002:19 Lindeboom Maarten, France Portrait & Gerard J. van den Berg "An econometric analysis of the mental-health effects of major events in the life of older individuals"

2002:20 Abbring Jaap H., Gerard J. van den Berg **"**Dynamically assigned treatments: duration models, binary treatment models, and panel data models"

Dissertation Series

2002:1 Larsson Laura "Evaluating social programs: active labor market policies and social insurance"

2002:2 Nordström Skans Oskar "Labour market effects of working time reductions and demographic changes"