Econometrica, Vol. 71, No. 5 (September, 2003), 1491–1517

THE NONPARAMETRIC IDENTIFICATION OF TREATMENT EFFECTS IN DURATION MODELS

BY JAAP H. ABBRING AND GERARD J. VAN DEN BERG[1]

This paper analyzes the specification and identification of causal multivariate duration models. We focus on the case in which one duration concerns the point in time a treatment is initiated and we are interested in the effect of this treatment on some outcome duration. We define "no anticipation of treatment" and relate it to a common assumption in biostatistics. We show that (i) no anticipation and (ii) randomized treatment assignment can be imposed without restricting the observational data. We impose (i) but not (ii) and prove identification of models that impose some structure. We allow for dependent unobserved heterogeneity and we do not exploit exclusion restrictions on covariates. We provide results for both single-spell and multiple-spell data. The timing of events conveys useful information on the treatment effect.

KEYWORDS: Program evaluation, bivariate duration analysis, selectivity bias, hazard rate, partial likelihood, unobserved heterogeneity, anticipation.

1. INTRODUCTION

THIS PAPER DISCUSSES THE SPECIFICATION and identification of causal multivariate duration models in economics. Consider a subject in a certain state. After a stochastic amount of time, the subject leaves this state. A different type of event may occur at some other random time. We are interested in the causal effect of the latter event on the duration in the state of interest. 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. In biostatistics one is often interested in the effect of a specific medical treatment. We borrow this terminology and refer to training programs, punitive benefits reductions, et cetera, as "treatments," to their causal effects on the outcome durations as "treatment effects," and to the subjects as "individuals."

At the core of the paper is a discussion of the causal relation between two durations, the outcome duration Y and the treatment time S. A duration is simply a nonnegative random variable and it may seem that we can apply standard models for two nonnegative random variables. However, in an explicitly dynamic model framework it is easier to discuss censoring and time aggregation, and handle cases where the treatment may not be observed, or even well-defined, after the subject has left the state of interest. We specify such a

[1]Thanks to Jim Heckman, Bo Honoré, Hide Ichimura, John Kennan, Tony Lancaster, Geert Ridder, Michael Visser, Ed Vytlacil, the editor, and anonymous referees for useful comments and suggestions. Also thanks to seminar participants at UCL London, NYU, IFAU Uppsala, Chicago, Iowa, Stockholm School of Economics, Mannheim, Bristol, INSEE-CREST, Tinbergen Institute Amsterdam, and participants at various conferences. Jaap Abbring acknowledges financial support by the Royal Netherlands Academy of Arts and Sciences.

dynamic model framework in terms of the hazard rates corresponding to Y and S. Hazard rates are the focal points and basic building blocks of econometric duration models. Dynamic economic theories often aim to explain the rate of leaving a state of interest in terms of economic behavior and external conditions (see Van den Berg (2001) for a survey).

We construct a causal model that specifies how the outcome varies if we hypothetically manipulate the treatment time. The model also specifies a (not necessarily randomized) treatment process. Biostatistical studies routinely impose a "consistency" condition that can be interpreted as claiming that each cause should precede its effect (e.g. Robins (1998) and Lok (2001)). Formally, it requires that the outcome (transition) paths corresponding to any two treatment times coincide up to a duration y if the treatment (transition) paths are the same up to y. In more familiar terms, the subject should be in the same outcome state at time y if both treatments start after y. We impose a weak version of this condition that only requires that the corresponding hazard paths coincide up to y. In economics this embodies a substantial informational and behavioral assumption. It requires that individuals either have no access to information on treatment beyond that encoded in the treatment history up to y or simply do not act on that information, at each time y. Otherwise, there are anticipatory effects of future treatment and our consistency condition is violated. Therefore, we refer to this condition as a "no-anticipation" assumption.

Rational forward-looking individuals may be expected to exploit information on the (perceived) properties of the treatment assignment process. In that case, changes in these properties may affect the hazard rate out of the state of interest *before* any actual realization of a treatment. Our model framework enables the specification of such effects. Our analysis, however, focuses on contrasting outcomes between different *realized* treatment times for a given assignment mechanism. This is already an ambitious task. We prove a basic nonidentification theorem that states that no-anticipation and randomized assignment of treatment can be imposed on the causal model without restricting the observational data. This highlights two inference problems. First, we cannot identify anticipation effects of treatment. Second, we have the standard econometric selection problem. Without further assumptions, we cannot distinguish between causal effects and selection effects. In the identification analysis we take the no-anticipation assumption as fundamental and we only address the selection issue. It should be judged in each application whether it makes sense to assume that the observed treatment cannot be anticipated.

Identification requires further structure. We build on the assumption that all selection effects can be captured by related observed and unobserved covariates in the treatment and outcome processes. We examine single-spell and multiple-spell settings. Single-spell data provide information on a single outcome spell and possibly treatment for each individual. We prove identification of two bivariate duration models that allow for time-varying and heterogeneous treatment effects and that borrow structure from the well-known non-

parametric[2] mixed proportional hazard (MPH) model. MPH models are the most popular reduced-form duration models in econometrics (see Van den Berg (2001) for a survey).

With multiple-spell data we observe multiple outcome spells, with or without treatment, for each individual or within groups of individuals who share the same unobservable characteristics. We prove identification of multiple-spell model generalizations. In particular, we can allow for general nonproportionalities. The single-spell results concern bivariate models that embed a marginal model for the treatment process, or a "selection equation." For the multiple-spell case we show that it is sufficient to specify the conditional outcome model in order to identify treatment effects.

This paper contributes to the extensive literature on the econometric evaluation of social programs and treatment effects (see Heckman, LaLonde, and Smith (1999) for an overview). This literature provides relatively little insight in problems with dynamically assigned treatments and duration outcomes. We provide some guidance in tackling such problems, which are frequently encountered in empirical research. We also complement the closely related biostatistical literature (e.g. Robins (1998)). This literature allows for more complex dynamic treatment regimes, but is not tailored to economic problems and data. We offer a dynamic economic perspective on the causal model and identification results for models with unobserved selection effects that cannot be found in the biostatistical literature. Our identification approach is complementary to an approach in which regularity assumptions are made that are insufficient for point-identification but that allow for the identification of bounds for a treatment effect (see, e.g., Manski (1990, 1997)).

Our identification results also extend the existing literature on the nonparametric identification of MPH-type duration models (see the surveys in Heckman and Taber (1994) and Van den Berg (2001)). The conditional outcome parts of our models can be interpreted as univariate duration models with an endogenous time-varying regressor (the treatment process), where the endogeneity originates from a dependence on unobserved characteristics. We demonstrate that these models are identified under similar assumptions as those in the literature.

The empirical literature contains studies in which models are estimated that are similar to the models we consider. For example, Card and Sullivan (1988), Gritz (1993), Bonnal, Fougère, and Sérandon (1997), Abbring, Van den Berg, and Van Ours (1997), and Van den Berg, Van der Klaauw, and Van Ours (2004) study the effect of a treatment of unemployed workers on the transition rate from unemployment to work, allowing for a causal effect on the transition rate as well as for related unobserved heterogeneity in order to deal with selectivity. Lillard (1993) estimates a model for the joint durations of marriage and time until conception of a child, and his model allows the rate at which the marriage

[2]In Subsection 3.2 we justify this terminology.

dissolves to shift to another level at moments of child birth. Lillard and Panis (1996) estimate a model on the joint durations of marriage, nonmarriage, and life, and their model allows the death rate to shift to another level at moments of marriage formation and dissolution. These studies also allow for related unobserved heterogeneity. Below we use the effect of a sudden unemployment benefits reduction on the transition rate to work as an example.

We do not impose exclusion restrictions on covariates, i.e. we do not require that the data contain a variable that affects the treatment assignment but does not affect the outcome of interest other than by way of the treatment. A variable that is observed by the analyst is often also observable to the individuals under consideration. If such a variable affects the treatment process, then a rational individual will take this variable into account in determining his optimal strategy. This behavior affects the rate at which the individual leaves the state of interest. As a result, exclusion restrictions are difficult to justify and instrumental variable methods of inference are not likely to be of help. A practical issue is that the latter methods are often not tailored to our dynamic framework. One may be tempted to time-aggregate the treatment process into a binary indicator for treatment in an interval and apply standard methods for binary treatments. However, this way we would lose track of the parameters of interest and throw away identifying information. Moreover, it is not clear what to do when treatment and outcome occur in the same interval or when observations are right-censored before the end of the interval.

We also avoid conditional-independence assumptions, which require that all selectivity in the assignment of treatment can be controlled by conditioning on observed covariates. In many econometric applications, there may be unobserved heterogeneity in treatment assignment and outcomes. By the argument against instrumental variables assumptions above, any unobserved heterogeneity in treatment assignment is likely to imply unobserved selectivity and violations of conditional independence.

Finally, we show that multiple-spell data are similar to linear panel data in the sense that the intuition for identification in linear panel-data models carries over to our nonlinear models.

Section 2 develops the causal framework and the basic nonidentification result. Sections 3 and 4 derive the identification results for the single-spell and multiple-spell cases, respectively. Section 5 concludes.

2. A FRAMEWORK FOR CAUSAL INFERENCE

2.1. *A Potential Outcome Framework*

Consider a population of individuals flowing into a state of interest and the durations these individuals subsequently spend in that state. We are interested in the causal effect of a single binary treatment that is either assigned at some time in $\mathbb{R}_+ := [0, \infty)$ after entering the state or not assigned at all. We can cast this problem in the standard potential outcome framework pioneered

by Neyman (1923) by recognizing that our dynamically assigned binary treatment can be reinterpreted as a set of mutually exclusive treatments indexed by $\overline{\mathbb{R}}_+ := \mathbb{R}_+ \cup \{\infty\}$. Here, the point ∞ represents the no-treatment case. To each treatment $s \in \overline{\mathbb{R}}_+$ corresponds a random variable $Y^*(s) \geq 0$, the potential outcome in the case that we would intervene and assign treatment s. Causal inference is concerned with contrasting potential outcomes corresponding to different treatments. Because the treatments are mutually exclusive, we can never observe potential outcomes corresponding to different treatments simultaneously. In the words of Dawid (2000), potential outcomes are complementary. This is what Holland (1986) calls the "fundamental problem of causal inference." A solution is to impose sufficient structure on the model to allow for causal inference from actual (experimental or observational) data. The inherently dynamic setting of our model and its application in economics call for a separate discussion.

Unlike Neyman and much of the more recent literature (e.g. Rubin (1974), Holland (1986), and Heckman, LaLonde, and Smith (1999)), we have to deal with infinitely—actually uncountably—many potential outcome variables.[3] For a careful discussion, we explicitly introduce the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which all random variables are defined. $\{Y^*\} := \{Y^*(s); s \geq 0\}$, which collects all potential outcomes, is an $\mathbb{R}_+$-valued stochastic process, which we require to be measurable. Treatments are assigned according to an $\overline{\mathbb{R}}_+$-valued random variable S. The actual outcome is $Y := Y^*(S)$; all other potential outcomes are counterfactual. Measurability of $\{Y^*\}$ ensures that Y is a random variable (by application of Billingsley (1995, Theorem 13.1)). We refer to the pair $(\{Y^*\}, S)$ (or, rather, its distribution) as a causal model specification. A causal model is a set of causal model specifications.

Now consider a sample of individuals i with $i = 1, 2, \ldots, n$, with potential outcomes and treatments $(\{Y_i^*\}, S_i)$ that are distributed like $(\{Y^*\}, S)$ above. We focus on inference of characteristics of the marginal distributions of $Y^*(s)$, like the means $\mu(s) := \mathbb{E}[Y^*(s)]$. For example, suppose that $\mu(s) < \infty$ for all $s \in \overline{\mathbb{R}}_+$ and that we have been able to ensure randomized assignment, so that $\{Y^*\} \perp\!\!\!\perp S$. Then $\mu(S) = \mathbb{E}[Y|S]$ almost surely. So, under some additional smoothness assumptions, we can estimate μ by standard nonparametric regression techniques.

[3]Manski (1997), however, derives nontrivial bounds on the effects of continuous treatments under the assumption of monotone (and bounded) responses. Gill and Robins (2001) extend Robins' earlier work on g-computation to continuous treatments under additional continuity conditions. Also, structural models routinely allow for causal effects of continuous variables. See Heckman (2000) for a review and Pearl (2000) for a discussion of the link between structural models and potential outcome models.

2.2. *The Dynamic Causal Model and the Accumulation of Information*

So far we have ignored the fact that our problem is inherently dynamic. To facilitate an explicit discussion of dynamics, we assume that $Y^*(s)$ is a continuous random variable, and we focus on its hazard rate, which is given almost everywhere by

$$\theta_{Y^*(s)}(y) = \lim_{dy \downarrow 0} \frac{\Pr(Y^*(s) \in [y, y+dy) | Y^*(s) \geq y)}{dy}.$$

The corresponding integrated hazards are denoted by $\Theta_{Y^*(s)}(y) := \int_0^y \theta_{Y^*(s)}(u)du$. We assume that $\Theta_{Y^*(s)}(y) < \infty$ for all $y \in \mathbb{R}_+$, and that $\lim_{y\to\infty} \Theta_{Y^*(s)}(y) = \infty$. The latter assumption entails that $\Pr(Y^*(s) < \infty) = 1$, which may be restrictive in certain applications. It is dropped in Section 3.

It is well known and easy to check that under these assumptions $\Theta_{Y^*(s)}(Y^*(s))$ has a unit exponential distribution. This suggests the following approach towards a causal model. Explicitly including the sample point ω as an argument, let

$$Y^*(s, \omega) = \Theta^{-1}_{Y^*(s)}(E(s, \omega)),$$

with $\Theta^{-1}_{Y^*(s)}(\cdot) = \inf\{y \in \mathbb{R}_+ : \Theta_{Y^*(s)}(y) \geq \cdot\}$ the generalized inverse of $\Theta_{Y^*(s)}$ and $E(s, \cdot)$, $s \in \overline{\mathbb{R}}_+$, unit exponential random variables. Then, it is easy to derive that $Y^*(s, \cdot)$ is indeed a continuous random variable with integrated hazard $\Theta_{Y^*(s)}$, for given $s \in \overline{\mathbb{R}}_+$.

We adopt the following corresponding causal interpretation. First, nature selects an ω according to the model $(\Omega, \mathcal{F}, \mathbb{P})$. Then we intervene and set s, and thus $\Theta_{Y^*(s)}$ and $E(s, \cdot)$.[4] Nature does not re-randomize after our intervention, so ω is fixed throughout the experiment. As the distribution of $E(s, \cdot)$ is invariant to our choice of s, the potential integrated hazards $\Theta_{Y^*(s)}$ have a causal interpretation. Note that this does not require invariance of the actual realization $E(s, \omega)$. Indeed, we are not interested in the structure of $E(s, \cdot)$ beyond distributional invariance.[5] Having established this, it is unnecessary and cumbersome to use notation that explicates the dependence of E on s and to carry along the argument ω. We can simply specify $Y^*(s) := \Theta^{-1}_{Y^*(s)}(E)$, with E a unit exponential random variable, and focus our analysis on $\Theta_{Y^*(s)}$.[6] If we have randomized assignment and observe the distribution of Y conditional on S, then

[4]Note that this explicitly allows for randomness at the individual level like Freedman (2002) and unlike, for example, Holland (1986).

[5]The joint distributions of $\{Y^*\}$ can only be determined if we make further assumptions on the joint distributions of the process $\{E\}$. We would need these distributions to evaluate truly counterfactual statements, notably those involving "causes of effects" (Dawid (2000)). Dawid argues that any such assumptions are fundamentally untestable, except in the unlikely case that, at least in principle, $\{E\}$ can be fully determined by scientific inquiry.

[6]Measurability of Y^* is ensured by requiring that $(s, e) \mapsto \Theta^{-1}_{Y^*(s)}(e)$ is measurable (again by Billingsley (1995, Theorem 13.1)).

$\Theta_{Y^*(s)}$ can be estimated using standard hazard regression techniques (see, e.g., Fleming and Harrington (1991)).

We are now ready to discuss more substantive dynamic issues. Heuristically, $\theta_{Y^*(s)}(y)dy$ is the probability that the outcome spell ends in a small interval $[y, y + dy)$, conditional on survival up to y and under treatment s. For each treatment $s \in \overline{\mathbb{R}}_+$, introduce a dynamic treatment indicator $\{N_s(y); 0 \leq y < \infty\}$ such that $N_s(y) = 0$ if $y < s$ and $N_s(y) = 1$ if $y \geq s$. Then, $\{N_S\}$ is a counting process that tracks the actual treatment status over time. An issue that is central to a behavioral science like economics is the dynamic accumulation of information concerning treatment by the individual. As a routine "consistency" condition, biostatisticians typically impose that the outcomes up to some time y only depend on the treatment history $\{N_s(u); 0 \leq u < y\}$, for a given assigned treatment $s \in \overline{\mathbb{R}}_+$ (Robins (1998) and Lok (2001)). A weak version of this condition requires that hazard rates for any two treatments $s, t \in \overline{\mathbb{R}}_+$ coincide almost everywhere up to the first of the two treatments, $\min\{s, t\}$. This consistency condition seems natural, by simply requiring that cause precedes effect. However, in social sciences, individuals may act on knowledge of the moment of realization of a future treatment. In such cases, individuals are said to *anticipate* the treatment. We can therefore phrase our "consistency" condition as a "no-anticipation" assumption.

ASSUMPTION 1 (No anticipation): *For all* $s, t \in \overline{\mathbb{R}}_+$,

$$\Theta_{Y^*(s)}(y) = \Theta_{Y^*(t)}(y) \quad \textit{for all } y \leq \min\{s, t\}.$$

Suppose that individuals are informed about the treatment time at the start of the spell. If individuals act on this information and their actions affect outcomes, the potential hazards corresponding to different treatments will diverge from time 0 onwards. Similar but subtler patterns arise if individuals dynamically accumulate treatment-specific information that strictly includes the actual treatment history. An example is that the treatment time is revealed at some fixed interval before the treatment time. If the econometrician only observes the moment of the actual treatment, then in these cases the above assumption is violated. Now one may argue that in such cases the real treatment (or, at least, the first treatment) concerns the information shock rather than the actual exposure or participation, and that the real moment of treatment is the moment at which the information about future actual exposure or participation arrives. Information shocks cannot be anticipated. This leads to a somewhat different interpretation of Assumption 1, namely that the econometrician *observes* the real treatment.[7]

[7]If we observe the timing of all information shocks we can use a multiple-treatment extension of our model that is not hard to envision.

It is important to stress that Assumption 1 does not exclude that forward-looking individuals act on properties of the treatment process. It only requires that such effects are the same between interventions manipulating the future treatment status. In that sense, the potential outcome part of our model is a model for contrasting treatments for a given information structure. It does not allow us to contrast outcomes to outcomes in a world without treatment. One of the contributions of this paper is to make these problems explicit in a dynamic setting.[8]

2.3. *A Basic Nonidentification Result*

So far we have assumed that the econometrician always observes S and therefore $\{N_S\}$. However, in many applications treatment is irrelevant and/or unobserved if it occurs after the spell of interest has ended. From now on, we explicitly entertain the possibility that we only observe the treatment history $\{N_S(y); 0 \leq y < Y\}$ realized at Y. Assume that S and $\{Y^*\}$ are such that $\Pr(Y = S) = 0$ (see also Assumption 2 below). In this case, a large data set would provide

$$
\begin{aligned}
&Q_S(y,s) := \Pr(Y > y, S > s, Y > S) \quad \text{and} \\
&Q_Y(y) := \Pr(Y > y, Y < S)
\end{aligned}
\tag{1}
$$

for all $(y,s) \in \mathbb{R}^2_+$. These are the sub-survival-functions of (Y,S) and Y for the sub-populations with respectively $Y > S$ and $Y < S$. So, we have the following definition.

DEFINITION 1: Two causal model specifications $(\{Y^*\}, S)$ and $(\{\widetilde{Y}^*\}, \widetilde{S})$ are *observationally equivalent* if $(Q_S, Q_Y) = (\widetilde{Q}_S, \widetilde{Q}_Y)$, where $\widetilde{Q}_S$ and $\widetilde{Q}_Y$ are defined by (1) for $\widetilde{Y} := \widetilde{Y}^*(\widetilde{S})$ and $\widetilde{S}$.

Note that the distribution of the *identified minimum* of (Y,S), i.e. the smallest of Y and S joint with the identity of this smallest duration, is fully characterized by (Q_S^0, Q_Y), with $Q_S^0(s) = Q_S(-\infty, s)$ for all $s \in \mathbb{R}_+$ (Tsiatis (1975)). We will occasionally exploit that our models of (Q_S, Q_Y) embed competing risks models for (Q_S^0, Q_Y).

Now, first note that, under Assumption 1 and randomized assignment, $\Theta_{Y^*(s)}$ can be identified from a standard hazard regression on the treatment history alone. This does not require information on $\{N_S(y); Y < y \leq \infty\}$. We show in the Appendix that any data pair (Q_S, Q_Y) can be supported by a causal model

[8]Heckman, LaLonde, and Smith (1999) have addressed similar issues within the context of the static binary-treatment potential outcome model. They differentiate between the no-treatment outcome in a world with treatments and outcomes in a world without treatments.

specification that satisfies Assumption 1 and randomized assignment.[9] This implies the following proposition.

PROPOSITION 1: *To each causal model specification* $(\{Y^*\}, S)$ *corresponds an observationally equivalent specification* $(\{\widetilde{Y}^*\}, \widetilde{S})$ *that satisfies Assumption* 1 *and randomized assignment* $(\{\widetilde{Y}^*\} \perp\!\!\!\perp \widetilde{S})$.

Note that the "no-anticipation" assumption is made on the marginal distributions of $\{Y^*\}$ whereas the "randomized-assignment" assumption is made on the relation between S and $\{Y^*\}$. One reason for failure of randomized assignment can be that there are causal (in particular, anticipatory) effects of the outcome on treatment assignment.

Proposition 1 provides two insights. First, without further structure, we cannot distinguish between causal effects and spurious selection effects. This is the standard selection problem. Second, we cannot distinguish between potential outcome models with and without anticipation (as defined by Assumption 1). This problem also haunts other methods for inference of treatment effects but is rarely mentioned. Anticipation is often tacitly ruled out. Subsection 2.4 illustrates Proposition 1 by way of an example. Except for that subsection, we assume throughout the remainder of the paper that Assumption 1 is true, and we focus on separating causal and spurious effects. So, we address the first nonidentification problem, but not the second. This requires additional structure. We impose this by explicitly modelling selection effects as related effects of vectors of observed covariates X and unobserved (to the econometrician) covariates V on outcomes and assignment. To this end, define a continuously-distributed potential outcome $Y^\dagger_{sxv}$ for each possible value (s, x, v) of (S, X, V). Similarly, let $S^\dagger_{xv}$ be the random treatment time in the case that we assign covariate values (x, v). The processes $\{Y^\dagger\}$ and $\{S^\dagger\}$ again have to be measurable. For definiteness, we assume that there are no causal effects of S and Y on (X, V). Consistency with the previous model setup requires that $Y^*_s = Y^\dagger_{sXV}$ and $S = S^\dagger_{XV}$. The actual outcome is $Y = Y^\dagger_{SXV}$. We assume that there is no anticipation: Assumption 1 holds for the integrated hazards of $\{Y^\dagger\}$ for all (x, v) in the support of (X, V). In the sequel, it is understood that we mean this slightly stronger assumption if we refer to Assumption 1.

Instead of randomized assignment we impose that all selection effects can be captured by conditioning on (X, V):

ASSUMPTION 2: $\{Y^\dagger\} \perp\!\!\!\perp \{S^\dagger\}$. *The distribution of* $(Y^\dagger_{S^\dagger_{xv}xv}, S^\dagger_{xv})$ *is absolutely continuous with respect to Lebesgue measure on* $\mathbb{R}^2_+$ *for all* (x, v) *in the support of* (X, V).

[9]Gill and Robins (2001) provide a related result for a different model.

This is reminiscent of conditional-independence assumptions but also allows for conditioning on unobserved covariates. The covariates (X, V) should include all joint determinants of outcomes and assignment, including any information that triggers relevant behavioral responses. Note that Assumption 2 excludes the trivial case where the covariates fully determine either the treatment or the outcome. Thus, after controlling for (X, V) there is always some random variation left in both.

The restriction to time-invariant observed covariates is unnecessarily strong for the exposition of the causal model and we could easily extend the analysis to *external* time-varying covariates.[10] However, in our identification analysis we will only deal with time-invariant covariates, which is in line with most of the literature on the identification of duration models (see Heckman and Taber (1994), and Van den Berg (2001) for surveys). The main reason to restrict attention to time-invariant observed covariates in this literature and our paper is expositional convenience. Results by Honoré (1991) and Heckman and Taber (1994) suggest that identification often benefits from any additional variation of the observed covariates over time. In this sense, our analysis provides a baseline that can be enriched by extensions to observed time-varying covariates.

It is more difficult to deal with time-varying unobserved covariates. We do not deal with this except for the case where the covariate paths are common within strata (see Section 4). We recognize that this imposes an important restriction on the model. After all, it is not hard to think of realistic examples with time-varying unobserved covariates. It is a useful topic for further research to study identification of models with such time-varying unobservables. In particular, instrumental variables (or, better, processes), if available, may enable identification.

2.4. *Example: The Effect of Time-varying Benefits on the Unemployment Duration*

In this subsection we illustrate the nonidentification result in Proposition 1 in the context of a specific economic-theoretical (job search) model framework. The treatment is a reduction of the benefits level of an unemployed individual at time S, and the outcome Y is the unemployment duration. In the first model version, individuals have perfect foresight concerning the realization of S, whereas in the second model version they only know the stochastic assignment rule. In other words, in the first version there is full anticipation whereas in the second there is none. We demonstrate that the implied causal models contain observationally equivalent causal model specifications. Note that in both cases the researcher does not observe S unless it precedes Y.

[10]See Kalbfleisch and Prentice (1980, Section 5.3), Heckman and Taber (1994), and Abbring (2001, Section 3.2) for discussions of external covariates in survival analysis.

Consider a standard partial search model describing an unemployed individual (e.g. Mortensen (1986)). Job offers arrive at a rate $\lambda \cdot e$, where $e \geq 0$ is the intensity of search and $\lambda > 0$ is a structural parameter. The corresponding flow of search costs equals $\frac{1}{2}e^2$. The wage offer distribution is degenerate at w. An individual initially receives unemployment benefits equal to w while searching, but these are reduced to b at time S. The individual maximizes her expected lifetime income discounted at the rate $\rho > 0$. For expositional convenience we impose that

$$\lambda = \sqrt{\frac{2\rho + 1}{2(w - b)}} \tag{2}$$

throughout this section.

In the first model version, the individual knows her value s of S from the outset. From the Bellman equations for the expected present value it can be deduced that

$$\theta_{Y^*(s)}(y) = \begin{cases} \dfrac{2\rho}{(2\rho + 1)\mathrm{e}^{\rho(s-y)} - 1} & \text{if } 0 \leq y \leq s \quad \text{and} \\ 1 & \text{if } y > s. \end{cases} \tag{3}$$

For each given s, $\theta_{Y^*(s)}$ increases on $[0, s)$ and is continuous at s. To further facilitate the exposition, we assume that assignment is randomized. We take S to have the density $g(s) \propto (2\rho + 1 - \mathrm{e}^{-\rho s})^2 \mathrm{e}^{-s}$. Recall that the data can be summarized by (Q_S, Q_Y). It can be shown that

$$\begin{aligned} Q_S(y, s) &= \frac{2(\rho + 1)(2\rho + 1)}{4\rho^2 + 10\rho + 5}(1 + \max\{0, y - s\})\mathrm{e}^{-\max\{y,s\}} \quad \text{and} \\ Q_Y(y) &= \frac{4\rho + 3}{4\rho^2 + 10\rho + 5}\,\mathrm{e}^{-y}. \end{aligned} \tag{4}$$

In the second model version, we assume that the benefits reduction arrives at a constant rate (i.e. that S is exponentially distributed) and that assignment is again randomized. The arrival rate of the benefits reduction is known to the individuals, but an individual does not know her realization s of S in advance. Apart from this, the second model version is the same as the first model version. In particular, the wage and initial benefits level w and the reduced benefits level b are the same. We also use identical specifications of search technologies and preferences, but we allow the discount rate ρ and the search-technology parameter λ to be different between the model versions. We make this explicit by using ρ and λ for the first model version only and by denoting the corresponding parameters in the second model version by $\tilde{\rho}$ and $\tilde{\lambda}$, respectively. The restriction in (2) is imposed on both sets of parameters. Note that our intervention involves a manipulation of s without changing the hazard

rate of S. From the Bellman equations it can be deduced that the hazard rate $\theta_{Y^*(s)}(y)$ is constant at some level below 1 before s and equals 1 after s.

For given w, b, and ρ, we can tailor the parameters of the second model version so that it implies the data (Q_S, Q_Y) in (4). In particular, let the arrival rate of the benefits reduction in the second model version be $2(\rho+1)(2\rho+1)/(4\rho^2+10\rho+5)$. Then, for sufficiently large ρ we can pick a discount rate $0<\tilde{\rho}<\rho$ such that the second model version is observationally equivalent to the first.[11] For example, if $w=1$, $b=\frac{1}{2}$, and $\rho=5\%$, then $\lambda=1.049$ by (2) in the first model version. An observationally equivalent model version without anticipation follows if we choose a constant rate 0.419 of benefits reductions and a discount rate $\tilde{\rho}=1.23\%$. In this model version, the hazard rate $\theta_{Y^*(s)}(y)$ jumps from 0.581 to 1 at s. By (2), $\tilde{\lambda}=1.012$.

3. SINGLE-SPELL DATA

3.1. *Model Framework*

Single-spell data ideally provide the joint distribution of $(Y, \{N_S(y); 0 \le y < Y\})$, conditional on observed covariates X. In obvious notation, suppose the data come in the form of a collection $\{Q_S, Q_Y\} := \{(Q_S(\cdot|x), Q_Y(\cdot|x)); x \in \mathcal{X}\}$ of conditional sub-survival functions. Here, $\mathcal{X} \subset \mathbb{R}^k$, $1 \le k < \infty$, is the support of X. The results of Subsection 2.3 (applied for given values of X) imply that Assumptions 1 and 2 are not sufficient to identify the causal model from $\{Q_S, Q_Y\}$. In this section, we show that the model is identified under additional separability (proportionality) assumptions. For ease of exposition we omit counterfactual indicators in the notation.

Assumption 2 implies that any dependence between Y and S conditional on (X, V) stems from causal effects of S on Y. Unobserved selection effects are fully captured by the unobserved covariates V. So, it is convenient to specify

[11] To provide some intuition, note that $\theta_{Y^*(s)}(y)$ at $y<s$ cannot be identified from data with $S=s$, because data on S are unavailable if $Y<S$. At the very best one can pin down the mixture hazard rate

$$\int_y^\infty \theta_{Y^*(s)}(y) e^{-\Theta_{Y^*(s)}(y)} g(s)ds \Big/ \int_y^\infty e^{-\Theta_{Y^*(s)}(y)} g(s)ds$$

over the potential hazards corresponding to treatments s in $[y,\infty)$. In both model versions this hazard rate equals

$$\lim_{dy\downarrow 0} \frac{\Pr(y \le Y < y+dy | Y \ge y, S \ge y)}{dy} \equiv \frac{-Q_Y'(y)}{Q_S^0(y)+Q_Y(y)} = \frac{4\rho+3}{4\rho^2+10\rho+5}$$

if we set $\tilde{\rho}$ equal to

$$\frac{32\rho^4+96\rho^3+88\rho^2+24\rho-1}{2(4\rho+3)(4\rho^2+10\rho+5)}.$$

This last expression is positive if $\rho > -\frac{3}{4}+\frac{1}{4}\sqrt{3}+\frac{1}{4}\sqrt{2} \approx 0.037$.

the model as a mixture of the joint distribution of $(Y, S)|(X, V)$ over the distribution of $V|X$. In turn, the joint distribution of $(Y, S)|(X, V)$ is the product of the distributions of $Y|(S, X, V)$ and $S|(X, V)$. So, $\{Q_S, Q_Y\}$ is fully specified by the following.

MODEL 1A: *The hazard rates of* $Y|(S, X = x, V)$ *and* $S|(X = x, V)$ *are given by*

$$\theta_Y(t|S, x, V) = \begin{cases} \lambda_Y(t)\phi_Y(x)V_Y & \text{if } t \leq S, \\ \lambda_Y(t)\phi_Y(x)\delta(t|S, x)V_Y & \text{if } t > S, \end{cases} \tag{5}$$

and

$$\theta_S(t|x, V) = \lambda_S(t)\phi_S(x)V_S, \tag{6}$$

respectively. V is an $\mathbb{R}_+^2$*-valued random vector* $(V_Y, V_S)'$ *distributed with distribution G independent of x. The following regularity conditions and normalizations hold.*

(i) ϕ_Y *and* ϕ_S *are continuous functions* $\phi_S : \mathcal{X} \to (0, \infty)$ *and* $\phi_Y : \mathcal{X} \to (0, \infty)$ *such that* $\phi_Y(x^*) = \phi_S(x^*) = 1$ *for some a priori chosen* $x^* \in \mathcal{X}$.

(ii) *The functions* $\lambda_Y : \mathbb{R}_+ \to (0, \infty)$ *and* $\lambda_S : \mathbb{R}_+ \to (0, \infty)$ *have integrals*

$$\Lambda_Y(t) := \int_0^t \lambda_Y(\tau)d\tau < \infty \quad \text{and} \quad \Lambda_S(t) := \int_0^t \lambda_S(\tau)d\tau < \infty$$

for all $t \in \mathbb{R}_+$. *For some a priori chosen* $t^* \in (0, \infty)$, $\Lambda_Y(t^*) = \Lambda_S(t^*) = 1$.

(iii) *G is such that* $\Pr(V \in (0, \infty)^2) > 0$.

(iv) $\delta : \mathbb{R}_+^2 \times \mathcal{X} \to (0, \infty)$ *is such that*

$$\Delta(t|s, x) := \int_s^t \delta(\tau|s, x)d\tau < \infty \quad \text{and}$$

$$\Upsilon(t|s, x) := \int_s^t \lambda_Y(\tau)\delta(\tau|s, x)d\tau < \infty$$

exist and are continuous on $\{(t, s) \in \mathbb{R}_+^2 : t > s\} \times \mathcal{X}$.

Equations (5) and (6) have mixed proportional hazard (MPH) specifications, except for the factor involving δ. An assumption implicit in equation (5) is that $Y \perp\!\!\!\perp V_S|(S, X, V_Y)$. Similarly, equation (6) embodies the assumption that $S \perp\!\!\!\perp V_Y|(X, V_S)$. One could say that V_Y captures the "unobserved heterogeneity" or "frailty" in Y and V_S captures the unobserved determinants of S.[12] In applied

[12]Here, V_Y and V_S are not necessarily interpreted as unobserved covariates (as in Subsection 2.3) but rather as their (proportional) effects on the conditional hazards.

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. Note that we allow the elements in X to be discrete and/or continuous.

Conditional on (X, V), the variables Y and S are only dependent through $\delta(t|S, X)$. Thus, by Assumption 2, this factor can be given a causal interpretation as the "treatment effect" of S on Y. Note that $\delta(t|s, x)$ only enters $\theta_Y(t|S, x, V)$ at durations $t > S$. So, Model 1a satisfies Assumption 1.[13] If $\delta = 1$, treatment is ineffective. On the other hand, suppose that δ equals some constant larger than one. When S is realized, θ_Y increases by $(\delta - 1) \cdot 100\%$. This stochastically reduces the remaining outcome duration in comparison to the case where the treatment is given at a later point of time. More generally, Model 1a allows the treatment effect to depend on elapsed duration t, the treatment time S, and the observed covariates X. By implication, it may vary with the time $t - S$ since treatment. If the treatment itself takes some time, the effect of $t - S$ may for instance capture a low exit rate during treatment. In Subsection 3.3, we discuss an alternative model that allows δ to depend on a third unobserved component but not on the treatment time.

The functions λ_Y and λ_S are called the "baseline hazards." The corresponding hazard rates are said to be duration dependent if the functions λ_Y and λ_S are nontrivial. The functions Λ_Y, Λ_S, and Υ appear naturally whenever we integrate the hazard rates over their first (duration) argument. We do not require that $\lim_{t\to\infty} \Lambda_Y(t) = \infty$, $\lim_{t\to\infty} \Lambda_S(t) = \infty$, or $\lim_{t\to\infty} \Upsilon(t|s, x) = \infty$. This means we allow for $\Pr(Y = \infty) > 0$ and $\Pr(S = \infty) > 0$, which is important in many applications. Also, we allow V_Y and V_S to have mass points at 0, which entails that some subjects have a hazard rate $\theta_Y(t|S, x, V)$ identically equal to zero and/or $\theta_S(t|x, V) \equiv 0$. In the latter case, a mass of individuals never receives treatment. Note that we allow for two sources of defectiveness (see Abbring (2002) for discussion).[14]

3.2. *Identification*

The components of Model 1a that can be freely varied are Λ_Y, Λ_S, ϕ_Y, ϕ_S, Δ, and G (we take the support $\mathcal{X}$ of X as given). If we specify each of these components, we have fully pinned down the data $\{Q_S, Q_Y\}$. Our identification analysis is nonparametric in the sense that we allow each of the model components to vary in infinite-dimensional function spaces, restricted by the regularity conditions of Subsection 3.1 and some additional assumptions. Let $\mathcal{M}$ be

[13]This also ensures that the stochastic process $\Pr(Y \le t|S, X, V)$ is measurable with respect to $\sigma(\{N_S(s); 0 \le s < t\}, X, V)$ and $\Pr(Y \le t|S, X, V) = \Pr(Y \le t|\{N_S(s); 0 \le s < t\}, X, V)$. We may therefore apply the conventional exponential representation of the survivor function (Fleming and Harrington (1991)).

[14]Model 1a and the analysis below can be straightforwardly extended with causal effects of the outcome on the treatment, introducing the mirror image of δ in equation (6). In applications of this, one typically observes not only $\{Q_S, Q_Y\}$, but the full joint distribution of $(Y, S)|X$. This is needed to identify causal effects in both directions.

the collection of all $(\Lambda_Y, \Lambda_S, \phi_Y, \phi_S, \Delta, G)$ that are admissible under a given set of assumptions. We simply refer to $\mathcal{M}$ as the "model" (under a given set of assumptions and for a given structure as in Model 1a) and to each $M \in \mathcal{M}$ as a "specification" of the model. Each specification $M \in \mathcal{M}$ maps into exactly one data collection $\{Q_S, Q_Y\}$. By analogy to Definition 1, two specifications $M, \widetilde{M} \in \mathcal{M}$ are said to be observationally equivalent if the corresponding data $\{Q_S, Q_Y\}$ and $\{\widetilde{Q}_S, \widetilde{Q}_Y\}$ are equal. Thus, we have the following definition.

DEFINITION 2: A model $\mathcal{M}$ is *identified* (from $\{Q_S, Q_Y\}$) if observational equivalence of any two specifications $M, \widetilde{M} \in \mathcal{M}$ implies that $M = \widetilde{M}$.

If $\mathcal{M}$ is identified, then we also say that the components Λ_Y, et cetera, are identified.

Recall that, in the case of Model 1a, Δ is only defined, and only needs to be identified, on $\{(t, s) \in \mathbb{R}_+^2 : t > s\} \times \mathcal{X}$. Once we know Δ, we can identify $\delta(t|s, x)$ for almost all $t > s$, for all $s \in \mathbb{R}_+$ and $x \in \mathcal{X}$. This covers the relevant area, as $\delta(t|s, x)$ only enters the model through the factor $\delta(t|S, x)^{N_S(t-)}$. Also note that concerning the effects of x we focus on nonparametric identification of the maps ϕ_Y and ϕ_S. In practice, one may start off with a parametric specification of, for example, ϕ_Y and require that all parameters can be recovered from its graph $\{(x, \phi_Y(x)); x \in \mathcal{X}\}$. In the case where $\phi_Y(x)$ is log-linear in $x'\beta_i$, this requires that $\mathcal{X}$ is not contained in a proper linear subspace of the k-dimensional Euclidian space $\mathbb{R}^k$.

Now, consider the following assumptions.

ASSUMPTION 3: $\{(\phi_Y(x), \phi_S(x)); x \in \mathcal{X}\}$ *contains a nonempty open set in* $\mathbb{R}^2$.

ASSUMPTION 4: $\mathbb{E}[V_Y] < \infty$ *and* $\mathbb{E}[V_S] < \infty$.

Note that Assumption 3 does not impose exclusion restrictions on the structural components of the hazards, but does require independent variation in these components. If $k = 2$, $\phi_Y(x) = \exp(x'\beta_Y)$, and $\phi_S(x) = \exp(x'\beta_S)$ for some β_Y and β_S such that $(\beta_Y \ \beta_S)$ has full rank, then it is sufficient that $\mathcal{X}$ contains a nonempty open set in $\mathbb{R}^2$. Assumption 4 is a common assumption in the analysis of single-spell duration data with MPH-type models (e.g. Elbers and Ridder (1982)).

The identification analysis in this section exploits that Model 1a embeds a competing risks model for the distribution $\{Q_S^0, Q_Y\}$ of the identified minimum. This embedded model is a standard MPH competing risks model that does not involve the treatment effect δ. Conditional on the observed covariates X, the competing risks are only dependent through the unobserved covariates V. Heckman and Honoré (1989) study the identifiability of similar competing risk models. Here, we rely on the following result of Abbring and Van den Berg (2003).

PROPOSITION 2: *Under Assumptions* 3 *and* 4, Λ_Y, Λ_S, ϕ_Y, ϕ_S, *and* G *in Model* 1*a are identified from* $\{Q_S^0, Q_Y\}$.

This leads to the main result.

PROPOSITION 3: *Under Assumptions* 3 *and* 4, Λ_Y, Λ_S, ϕ_Y, ϕ_S, Δ, *and* G *in Model* 1*a are identified from* $\{Q_S, Q_Y\}$.

The proof of this proposition provides some intuition on what drives the identification of the treatment effect. Consider individuals who receive a treatment at s. The natural control group consists of individuals whose spells have not ended at s and who have not yet received a treatment. A necessary condition for a meaningful comparison of these groups is that there is some randomization in the treatment assignment at s. The model allows for such randomization because we specify assignment by way of the *hazard rate* of S. The integrated hazard rate $\int_0^S \theta_S(\tau|x,V)d\tau$ has a unit exponential distribution independently of (x,V) (see, e.g., Ridder (1990) and Horowitz (1999)). Thus, there is a random component in the assignment that is independent of (x,V). This component affects Y only by way of the treatment.[15] However, the presence of this component is not sufficient for a meaningful comparison. We still have to deal with a selection effect: the unobserved heterogeneity distribution is different between the treatment and control groups at s. This can be corrected by exploiting the information in the competing risks part of the data. The competing risks part of the model determines the distribution of V_Y in the two groups at s and the competing risks part of the data enables the identification of this.

The results in our companion paper, Abbring and Van den Berg (2002b), provide some additional intuition for the case of a constant treatment effect δ. We show that single-spell data are informative on the sign of $\ln(\delta)$ even if there is no variation in X. If treatment and outcome are typically realized very quickly after each other, no matter how long the elapsed duration in the state of interest before the treatment, 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. If there is no such pattern but yet there is a correlation between the points in time of treatment and outcome, then this is evidence of a dominating selection effect. Somewhat loosely, local dependence indicates a causal effect whereas global dependence indicates a selection effect.

[15] It is tempting to interpret this random component in terms of the randomness of the outcome of S at t that remains if the hazard rate $\theta_S(t|x,V)$ at t is specified. Consider an individual who has not yet been given a treatment and whose spell has not yet ended, at t. Basically, in a small time interval $[t, t+dt)$, the probability of treatment is $\theta_S(t|x,V)dt$, and the probability of no treatment is $1-\theta_S(t|x,V)dt$. This is a Bernoulli trial. Given the value of $\theta_S(t|x,V)dt$, its outcome is completely random.

This is not surprising given that the treatment only works after it has been realized, whereas the unobserved heterogeneity values affect the hazard rates everywhere. This illustrates the usefulness of the information on the *timing* of events to assess treatment effects. There is a similarity to the identification of MPH models with exogenous time-varying explanatory variables (Honoré (1991)).

Further evidence is provided by a comparison to the identification of the treatment effect in standard latent variable models with a regression specification for the outcome of interest, a binary treatment indicator, and a regression specification for a corresponding latent variable (Abbring and Van den Berg (2002a)). By analogy to Ridder (1990) and Horowitz (1999), the distributions of $Y|(S, X, V)$ and $S|(X, V)$ in our Model 1a can be written in a regression form. The most fundamental difference between our model and latent variable models concerns the fact that the former incorporates the timing of the treatment whereas the latter do not. It is generally acknowledged that an exclusion restriction is required for identification of the treatment-effect parameter in latent variable models. Since we do not need exclusion restrictions to identify our model, it follows that the information on the timing of treatment is very useful.

3.3. *Heterogeneous Treatment Effects*

Model 1a excludes unobserved heterogeneity in the treatment effect. Recent contributions to the literature on treatment evaluation are often concerned with treatment effects that may depend on unobservables (see, e.g., Heckman, LaLonde, and Smith (1999) for a survey). The following model allows δ to depend on unobservables, but, unlike Model 1a, excludes variation of δ with S.

MODEL 1B: *The hazard rates of* $Y|(S, X = x, V)$ *and* $S|(X = x, V)$ *are given by*

$$\theta_Y(t|S, x, V) = \begin{cases} \lambda_Y(t)\phi_Y(x)V_Y & \text{if } t \leq S, \\ \lambda_\Delta(t)\phi_\Delta(x)V_\Delta & \text{if } t > S, \end{cases} \tag{7}$$

and

$$\theta_S(t|x, V) = \lambda_S(t)\phi_S(x)V_S, \tag{8}$$

respectively. V *is now a* $\mathbb{R}^3_+$*-valued random vector* $(V_Y, V_\Delta, V_S)'$ *distributed with distribution* G_Δ *independent of* x*. The relevant regularity conditions and normalizations of Model* 1*a hold and*:

(i) $\phi_\Delta : \mathcal{X} \to (0, \infty)$ *is a continuous function such that* $\phi_\Delta(x^*) = 1$ *for some a priori chosen* $x^* \in \mathcal{X}$;

(ii) *the function* $\lambda_\Delta : \mathbb{R}_+ \to (0, \infty)$ *has an integral*

$$\Lambda_\Delta(t) := \int_0^t \lambda_\Delta(\tau) d\tau < \infty$$

for all $t \in \mathbb{R}_+$; *for some a priori chosen* $t^* \in (0, \infty)$, $\Lambda_\Delta(t^*) = 1$;

(iii) G_Δ *is such that* $\Pr(V \in (0, \infty)^3) > 0$.

The treatment effect,

$$\delta(t|x, V_Y, V_\Delta) := \frac{\lambda_\Delta(t)\phi_\Delta(x)V_\Delta}{\lambda_Y(t)\phi_Y(x)V_Y},$$

indeed depends on unobservables, but not on S. Note that δ is only well-defined for $V_Y > 0$ here. We could use the point ∞ to represent the case $V_Y = 0$ but, instead, we focus directly on identification of λ_Y, λ_Δ, ϕ_Y, ϕ_Δ, and G_Δ.

First, note that Model 1b embeds the same MPH competing risks model as Model 1a, so that Proposition 2 again applies. We need some additional assumptions.

ASSUMPTION 5: $\{\phi_\Delta(x); x \in \mathcal{X}\}$ *contains a nonempty open interval in* $\mathbb{R}$.

ASSUMPTION 6: $\mathbb{E}[V_\Delta V_S] < \infty$.

These new assumptions are simply standard MPH assumptions (e.g. Elbers and Ridder (1982)) on the embedded model for $(Y - S)|(S, Y > S, X = x, V)$. After all, this is a MPH model with hazard $\lambda_\Delta(y)\phi_\Delta(x)V_\Delta$ at time $y - S$ (since S). The finite-mean Assumption 6 applies to $V_\Delta V_S$ rather than V_Δ because of the conditioning on S.

PROPOSITION 4: *Under Assumptions* 3–6, Λ_Y, Λ_Δ, Λ_S, ϕ_Y, ϕ_Δ, ϕ_S, *and* G_Δ *in Model* 1*b are identified from* $\{Q_Y, Q_S\}$.

The results in this section require separability assumptions, independence of X and V, sufficient variation in the observed covariates, and finiteness of the means of V. Although it is common to make such assumptions in applied duration analysis, they are sometimes hard to justify (see, for example, Van den Berg (2001)). To deal with this, we extend our analysis in two directions. In the next subsection we show that observed covariates, and most of the assumptions above, are not needed for identification if we have multiple-spell data. In Abbring and Van den Berg (2002b) we show that single spells are informative on aspects of δ and G even if there is no variation in X.

4. IDENTIFICATION IN THE CASE OF MULTIPLE-SPELL DATA

4.1. *Multiple-spell Data*

In this section, we study identification from data that cover multiple outcome and treatment spells for each individual. We focus on the case in which the data provide information on the distribution of two full spells over the population of individuals. The extension to more than two spells is trivial. Actually, the use of the term "individual" is not very appropriate here, because 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. It is convenient to refer to either such a group of individuals or a single individual for which we have multiple spells as a *stratum*. For each stratum, we observe a random draw from the joint distribution of $(Y_1, \{N_{S_1}(y); 0 \leq y < Y_1\}, Y_2, \{N_{S_2}(y); 0 \leq y < Y_2\})$. We add subscripts 1 and 2 to random variables and functions corresponding to respectively spell 1 and 2 in a stratum. So, Y_1 is the outcome duration in the first spell, Y_2 the outcome duration in the second spell in a stratum, et cetera. By analogy to (1), the multiple-spell data distribution can be characterized by

$$
\begin{aligned}
Q_{SS}(y_1, y_2, s_1, s_2) &:= \Pr(Y_1 > y_1, Y_2 > y_2, \quad S_1 > s_1, S_2 > s_2, \quad Y_1 > S_1, Y_2 > S_2),\\
Q_{YS}(y_1, y_2, s_2) &:= \Pr(Y_1 > y_1, Y_2 > y_2, \qquad\qquad S_2 > s_2, \quad S_1 > Y_1, Y_2 > S_2),\\
Q_{SY}(y_1, y_2, s_1) &:= \Pr(Y_1 > y_1, Y_2 > y_2, \quad S_1 > s_1, \qquad\quad Y_1 > S_1, S_2 > Y_2)
\end{aligned}
$$

and

$$
Q_{YY}(y_1, y_2) := \Pr(Y_1 > y_1, Y_2 > y_2, \qquad\qquad S_1 > Y_1, S_2 > Y_2)
$$

for all $(y_1, y_2, s_1, s_2) \in \mathbb{R}^4_+$. Throughout this section we suppress observed covariates. The identification results do not require variation in observed covariates. Moreover, the model allows for full interaction of such covariates with all other determinants. All results can be interpreted as being conditional on observed covariates.

As with linear panel data, the fact that we observe multiple outcomes for given unobserved heterogeneity values can be exploited to deal with unobserved heterogeneity under conditions that are mild relative to the single-spell (i.e. cross-sectional) case. Intuitively, we need some separability of the outcome hazards, in treatment effect and unobserved covariate components. Then, if the unobserved components are constant between spells in a stratum, variation between spells and within strata can be used to control for selection effects and identify the treatment effects.

In the next subsection, we study the multiple-spell analogues of Models 1a and 1b. These models allow for different baseline hazards and treatment effects between spells within a stratum, but have multiplicative unobserved heterogeneity in the hazards. In the vein of the literature on identification of multiple-spell duration models (Honoré (1993)), we explore identifiability of fully specified bivariate models. In Subsection 4.3, we focus on identification

of the treatment effect in models that only specify the conditional distribution of the outcome spells. Stratified partial likelihood (SPL) arguments are applied to prove identification of the treatment effects in a model that allows for interactions between duration effects and unobserved heterogeneity.

4.2. *Identification of the Full Model*

Consider the following multiple-spell analogues of Models 1a and 1b:

MODEL 2A: *Multiple spells within a stratum are only dependent through the covariates* V, *i.e.* $(Y_1, S_1) \perp\!\!\!\perp (Y_2, S_2)|V$. *The hazard rates of* $Y_k|(S_k, V)$ *and* $S_k|V$ *are*

$$\theta_{Y_k}(t|S_k, V) = \begin{cases} \lambda_{Y_k}(t)V_Y & \text{if } t \le S_k, \\ \lambda_{Y_k}(t)\delta_k(t|S_k)V_Y & \text{if } t > S_k, \end{cases} \quad \text{and}$$

$$\theta_{S_k}(t|V) = \lambda_{S_k}(t)V_S \qquad (k = 1, 2),$$

respectively. $V := (V_Y, V_S)$ *has distribution* G. *Regularity conditions as in Model* 1*a apply. For some a priori chosen* $t^* \in (0, \infty)$, $\Lambda_{Y_1}(t^*) = \Lambda_{S_1}(t^*) = 1$ (*in obvious notation*).

MODEL 2B: *Same as Model 2a except that*

$$\theta_{Y_k}(t|S_k, V) = \begin{cases} \lambda_{Y_k}(t)V_Y & \text{if } t \le S_k, \\ \lambda_{\Delta_k}(t)V_\Delta & \text{if } t > S_k, \end{cases} \qquad (k = 1, 2),$$

and $V := (V_Y, V_\Delta, V_S)$ *has distribution* G_Δ. *For some a priori chosen* $t^* \in (0, \infty)$, $\Lambda_{\Delta_1}(t^*) = 1$.

The first line excludes any causal effects across different spells within a stratum. Outcomes and treatments may be related between spells, but only through common unobservable components. The normalizations are innocuous because the unobserved covariates can capture the scale of first-period hazards. In both models, the same unobservables enter the first and second period hazards multiplicatively. This suggests that we can exploit the panel structure of the data, even though the baseline hazards and treatment effects are not specified to be the same in the first and second spell of a stratum. Indeed, we have the following result.

PROPOSITION 5: Λ_{Y_1}, Λ_{Y_2}, Λ_{S_1}, Λ_{S_2} *and, respectively,* Δ_1, Δ_2, *and* G *in Model* 2*a and* Λ_{Δ_1}, Λ_{Δ_2}, *and* G_Δ *in Model* 2*b are identified from* $(Q_{SS}, Q_{YS}, Q_{SY}, Q_{YY})$.

The proof exploits that both models embed a two-spell version of the competing risks model with hazards that are proportional in V and t. Abbring and

Van den Berg (2003) show that Λ_{Y_1}, Λ_{Y_2}, Λ_{S_1}, and Λ_{S_2} (but not G) in Model 2a or 2b are identified. Unlike the single-spell case, we do not invoke a competing risks result that also provides identification of G. This would impose additional conditions on the model that we can circumvent here by exploiting the additional (non-competing-risk) information in $(Q_{SS}, Q_{YS}, Q_{SY}, Q_{YY})$.

Note that we do not need finite $\mathbb{E}[V]$. Also, recall that we implicitly allow observed covariates X to enter the model in a general way. We could add X as an argument to each of the model components in Models 2a and 2b and allow V and X to be dependent. The gain (in terms of identification) from using multiple-spell data is similar to the gain from using such data for MPH models without treatment effects (Honoré (1993)).

4.3. *Identification and Estimation from a Stratified Partial Likelihood*

The results in the previous subsection show that multiple-spell data allow for robust inference on the treatment effects under the assumption of proportional unobserved heterogeneity common within strata. In panel regression models with individual-specific effects, a "within estimator" can be used to estimate the regression parameters. This analogy suggests an analysis of the multiple-spell model of $Y|(S, V)$ only, i.e. without specifying the marginal distribution of (S, V). In this subsection, we analyze identification of treatment effects in such a conditional model. The analysis is based on a Cox (1972) partial likelihood for stratified data.[16]

First, we adapt the notation. The SPL method allows for general interactions between time and unobserved covariates. Therefore, we replace V_Y by a (for now) stochastic process $\{V_Y\}$. It is implicitly understood that $\{V_Y\}$ replaces V_Y in V. We have the following model.

MODEL 2C: *Multiple outcome spells within a stratum are only dependent through the covariates V and related treatments S_1 and S_2, i.e. $Y_1 \perp\!\!\!\perp (Y_2, S_2)|(S_1, V)$ and $Y_2 \perp\!\!\!\perp (Y_1, S_1)|(S_2, V)$. The hazard rate of $Y_k|(S_k, V)$ is*

$$\theta_{Y_k}(t|S_k, V) = \begin{cases} \lambda_{Y_k}(t)V_Y(t) & \text{if } t \le S_k, \\ \lambda_{Y_k}(t)\delta_k(t|S_k)V_Y(t) & \text{if } t > S_k, \end{cases} \qquad (k = 1, 2).$$

Regularity conditions as in Model 1a apply. In particular $\lambda_{Y_k}(\cdot)V_Y(\cdot)$ is (pathwise) integrable on bounded intervals. We normalize $\lambda_{Y_1} = 1$.

[16]The analysis of a partial likelihood for stratified data was pioneered by Holt and Prentice (1974). See Ridder and Tunalı (1999) for an econometric SPL application with ample attention for practical problems like censoring. In medical science, SPL methods have been applied in a similar context in so-called "co-twin control" analyses of mortality (see Lichtenstein, Gatz, and Berg (1998)).

The unobserved heterogeneity $\{V_Y\}$ only enters hazard rates at t through its current value $V_Y(t)$. It is crucial that $\{V_Y\}$ is the same between the spells in a stratum. The spell-specific baseline hazards λ_{Y_1} and λ_{Y_2} allow for different duration dependence patterns between the spells. The normalization $\lambda_{Y_1} = 1$ is innocuous, as $\{V_Y\}$ can absorb the duration dependence and scale of θ_{Y_1}.

Let $Y_{(1)} := \min\{Y_1, Y_2\}$ be given. Intuitively, the probability that, say, the first spell in the stratum is the shortest, i.e. $Y_1 = Y_{(1)}$, does not depend on the factor $\{V_Y\}$ that is common between the spells in the stratum. It only depends on the duration dependence factor λ_{Y_2} and the treatment effects, which may be different between the spells if the treatment histories $\{N_{S_1}(\tau); 0 \leq \tau < Y_{(1)}\}$ and $\{N_{S_2}(\tau); 0 \leq \tau < Y_{(1)}\}$ are. Formally,

$$
\begin{aligned}
(9) \qquad & \Pr(Y_1 = Y_{(1)} | Y_{(1)}, \{N_{S_1}(\tau); 0 \leq \tau < Y_{(1)}\}, \{N_{S_2}(\tau); 0 \leq \tau < Y_{(1)}\}, V) \\
&= \frac{\delta_1(Y_{(1)}|S_1)^{N_{S_1}(Y_{(1)}-)}}{\delta_1(Y_{(1)}|S_1)^{N_{S_1}(Y_{(1)}-)} + \lambda_{Y_2}(Y_{(1)})\delta_2(Y_{(1)}|S_2)^{N_{S_2}(Y_{(1)}-)}} \\
&= \begin{cases}
\dfrac{1}{1 + \lambda_{Y_2}(Y_{(1)})} & \text{if } S_1, S_2 > Y_{(1)}, \\
\dfrac{\delta_1(Y_{(1)}|S_1)}{\delta_1(Y_{(1)}|S_1) + \lambda_{Y_2}(Y_{(1)})} & \text{if } S_1 < Y_{(1)} < S_2, \\
\dfrac{1}{1 + \lambda_{Y_2}(Y_{(1)})\delta_2(Y_{(1)}|S_2)} & \text{if } S_1 > Y_{(1)} > S_2, \\
\dfrac{\delta_1(Y_{(1)}|S_1)}{\delta_1(Y_{(1)}|S_1) + \lambda_{Y_2}(Y_{(1)})\delta_2(Y_{(1)}|S_2)} & \text{if } S_1, S_2 < Y_{(1)}.
\end{cases}
\end{aligned}
$$

The right-hand side does not depend on V. Indeed, the only unknown components are the second-period baseline hazard, which captures the duration-dependence difference between spells, and the treatment effects. For the first sub-population of strata in (9), with $S_1, S_2 > Y_{(1)}$, the only reason the spells may not be equally likely to be the shortest is a difference in the baselines at $Y_{(1)}$. Treatment effects are irrelevant because in both spells treatment has not started at time $Y_{(1)}$. The second and third sub-populations are directly informative on the treatment effects as the treatment status of the spells at $Y_{(1)}$ is different. Finally, the strata in which $S_1, S_2 < Y_{(1)}$ provide some more diffuse information on the parameters because the treatment effect enters for both spells.

Because V does not enter the right-hand side of (9), the same expression is valid for $\Pr(Y_1 = Y_{(1)} | Y_{(1)}, \{N_{S_1}(\tau); 0 \leq \tau < Y_{(1)}\}, \{N_{S_2}(\tau); 0 \leq \tau < Y_{(1)}\})$, which follows from data. So, we can first identify Λ_{Y_2} from data on the first sub-population of strata and then Δ_1 and Δ_2 from the second and third sub-populations, respectively. So, we have the following result.

PROPOSITION 6: *Λ_{Y_2}, Δ_1, and Δ_2 in Model 2c are identified from* $(Q_{SS}, Q_{YS}, Q_{SY}, Q_{YY})$.

Equation (9) has obvious implications for estimation.

5. CONCLUSION

In this paper we have analyzed the effect of a dynamically assigned binary treatment on a duration outcome. At the core of our analysis is a basic nonidentification result: (i) no anticipation and (ii) randomized treatment assignment can be imposed without restricting the observational data. Throughout the paper, we have taken the no-anticipation assumption to be fundamental. Cases in which agents are provided with and act on information on future treatments can be handled in a multiple-treatment extension of our framework in which both the actual treatment and information shocks are modelled as treatments.

The paper thus focuses on identifiability of the effects of a single binary treatment under the no-anticipation assumption, but allowing for general unobserved selectivity in the assignment process. We show that, under some separability assumptions, the (possibly heterogeneous) treatment effect is identified from single-spell data without the need to rely on exclusion restrictions, conditional independence, parametric functional-form assumptions, or multiple observations on the same individual. These results show that the timing of events conveys useful information on the treatment effect. In a companion paper (Abbring and Van den Berg (2002b)), we develop simple procedures for assessing the direction of the treatment effect that exploit this idea.

With multiple-spell data, the model is identified under even weaker conditions. In particular, we do not need some of the separability assumptions that are important for identifiability with single-spell data.

Dept. of Economics, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands; jabbring@feweb.vu.nl; http://staff.feweb.vu.nl/jabbring

and

Dept. of Economics, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands; gberg@feweb.vu.nl; http://staff.feweb.vu.nl/gberg.

Manuscript received April, 2000; final revision received February, 2003.

APPENDIX

PROOF OF PROPOSITION 1: Take some (Y, S) and the corresponding (Q_S, Q_Y) as given. By Miller (1977, Theorem 1), there exists a $(\widetilde{Y}, \widetilde{S})$ such that $\widetilde{S} \perp\!\!\!\perp \widetilde{Y}$, $\widetilde{Q}_S^0 = Q_S^0$, and $\widetilde{Q}_Y = Q_Y$. (This theorem extends the nonidentification results by Cox (1959, 1962) and Tsiatis (1975) for continuous competing risks models to general mixed discrete-continuous models.) Let

$$\Theta_{\widetilde{Y}^{*(s)}}(y) := \begin{cases} \Theta_{\widetilde{Y}}(y) & \text{if } 0 \leq y \leq s, \\ \Theta_{\widetilde{Y}}(s) - \ln[\Pr(Y > y | Y > s, S = s)] & \text{if } y > s. \end{cases}$$

Let E be a unit exponential random variable such that $E \perp\!\!\!\perp \widehat{S}$. Set $\widetilde{S} = \widehat{S}$, $\widetilde{Y}^*(s) = \Theta^{-1}_{\widetilde{Y}^*(s)}(E)$ for all $s \in \overline{\mathbb{R}}_+$ and $\widetilde{Y} = \widetilde{Y}^*(\widetilde{S})$. Clearly, $\{\widetilde{Y}^*\}$ satisfies Assumption 1 and $\{\widetilde{Y}^*\} \perp\!\!\!\perp \widetilde{S}$. Direct computations confirm that $(\{\widetilde{Y}^*\}, \widetilde{S})$ is observationally equivalent to $(\{Y^*\}, S)$. In particular, it is easy to check that $\widetilde{Q}_Y = \widehat{Q}_Y$ and $\widetilde{Q}^0_S = \widehat{Q}^0_S$, so that $\widetilde{Q}_Y = Q_Y$ and $\widetilde{Q}^0_S = Q^0_S$ by construction. *Q.E.D.*

PROOF OF PROPOSITION 3: By Proposition 2, Λ_Y, Λ_S, ϕ_Y, ϕ_S, and G are identified from $\{Q^0_S, Q_Y\}$. For fixed $x \in \mathcal{X}$,

$$\frac{\partial Q_S(y, s|x)}{\partial s} = \phi_S(x)\lambda_S(s)\mathcal{L}^{(S)}_G(\phi_Y(x)[\Lambda_Y(s) + Y(y|s, x)], \phi_S(x)\Lambda_S(s)) \tag{10}$$

for all $y \in \mathbb{R}_+$ and almost all $s \in \mathbb{R}_+$ such that $s < y$. $\mathcal{L}_G$ is the bivariate Laplace transform of G,

$$\mathcal{L}_G(z_1, z_2) := \int_0^\infty \int_0^\infty \exp(-z_1 v_Y - z_2 v_S) dG(v_Y, v_S)$$

and $\mathcal{L}^{(S)}_G(z_1, z_2) := \partial \mathcal{L}_G(z_1, z_2)/\partial z_2$ for all $(z_1, z_2) \in (0, \infty)^2$. For given s and x, the right-hand side of (10) is a strictly increasing and identified function of $Y(y|s, x)$. This identifies $Y(y|s, x)$ and therefore

$$\Delta(y|s, x) = \int_s^y \frac{\partial Y(\tau|s, x)/\partial \tau}{\lambda_Y(\tau)} d\tau$$

for all $x \in \mathcal{X}$ and all $y, s \in \mathbb{R}_+$ such that $s < y$ (using continuity). *Q.E.D.*

PROOF OF PROPOSITION 4: By Proposition 2, Λ_Y, Λ_S, ϕ_Y, and ϕ_S are identified from $\{Q^0_S, Q_Y\}$. For fixed $x \in \mathcal{X}$, $\partial^2 Q_S(y, s|x)/\partial y \partial s$ and $\partial^2 Q_S(y, s|x^*)/\partial y \partial s$ exist for almost all $y, s \in \mathbb{R}_+$ such that $s < y$. For these (y, s)

$$\begin{aligned} &\frac{\partial^2 Q_S(y, s|x)/\partial y \partial s}{\partial^2 Q_S(y, s|x^*)\partial y \partial s} \\ &\quad = \phi_S(x)\phi_\Delta(x) \frac{\mathcal{L}^{(\Delta S)}_{G_\Delta}(\phi_Y(x)\Lambda_Y(s), \phi_\Delta(x)(\Lambda_\Delta(y) - \Lambda_\Delta(s)), \phi_S(x)\Lambda_S(s))}{\mathcal{L}^{(\Delta S)}_{G_\Delta}(\Lambda_Y(s), \Lambda_\Delta(y) - \Lambda_\Delta(s), \Lambda_S(s))}. \end{aligned} \tag{11}$$

$\mathcal{L}_{G_\Delta}$ is the trivariate Laplace transform of the distribution G_Δ of (V_Y, V_Δ, V_S) and $\mathcal{L}^{(\Delta S)}_{G_\Delta}(z_1, z_2, z_3) := \partial^2 \mathcal{L}_{G_\Delta}(z_1, z_2, z_3)/\partial z_2 \partial z_3$ for all $(z_1, z_2, z_3) \in (0, \infty)^3$. Letting $y \downarrow 0$ and $s \downarrow 0$, both sides of (11) above reduce to $\phi_S(x)\phi_\Delta(x)$ because $\mathbb{E}[V_\Delta V_S] = \lim_{z \downarrow (0,0,0)} \mathcal{L}^{(\Delta S)}_{G_\Delta}(z) < \infty$. As we have identified ϕ_S and the left-hand side consists of data, this identifies ϕ_Δ.

Again for arbitrary $x \in \mathcal{X}$ and $y, s \in \mathbb{R}_+$ such that $s < y$,

$$\frac{\partial Q_S(y, s|x)/\partial s}{\lambda_S(s)\phi_S(x)} = \mathcal{L}^{(S)}_{G_\Delta}(\phi_Y(x)\Lambda_Y(s), \phi_\Delta(x)\big(\Lambda_\Delta(y) - \Lambda_\Delta(s)\big), \phi_S(x)\Lambda_S(s)\big), \tag{12}$$

with $\mathcal{L}^{(S)}_{G_\Delta}(z_1, z_2, z_3) := \partial \mathcal{L}_{G_\Delta}(z_1, z_2, z_3)/\partial z_3$. Note that the left-hand side of this equation is already identified. As $s \downarrow 0$, the right-hand side reduces to

$$\mathcal{L}^{(S)}_{G_\Delta}(0, \phi_\Delta(x)\Lambda_\Delta(y), 0). \tag{13}$$

After imposing $y = t^*$, we can identify the completely monotone function $-\mathcal{L}^{(S)}_{G_\Delta}(0, \cdot, 0)$ on a nonempty open set in $\mathbb{R}$ by appropriately varying x in (13). This identifies $-\mathcal{L}^{(S)}_{G_\Delta}(0, z, 0)$ for all $z \in (0, \infty)$ because of the real analyticity of $-\mathcal{L}^{(S)}_{G_\Delta}(0, \cdot, 0)$ (see below). Subsequently, Λ_Δ is identified from (13), because the right-hand side is strictly monotone in Λ_Δ.

Finally, by appropriately varying x and y in equation (12), we can trace $\mathcal{L}^{(S)}_{G_\Delta}$ on a nonempty open subset of $\mathbb{R}^3$. This identifies $\mathcal{L}^{(S)}_{G_\Delta}$ on $(0, \infty)^3$ because $-\mathcal{L}^{(S)}_{G_\Delta}$ is real analytic. A proof of these

statements is in a 2000 working paper version of this paper and is available upon request. With $\mathcal{L}_{G_\Delta}(0,0,0)=1$, this identifies $\mathcal{L}_{G_\Delta}$. *Q.E.D.*

PROOF OF PROPOSITION 5: By Abbring and Van den Berg (2003), Λ_{Y_1}, Λ_{S_1}, Λ_{Y_2}, and Λ_{S_2} are identified.

(i) For Model 2a, define $Y_k(y|s) := \int_s^y \lambda_{Y_k}(\tau)\delta_k(\tau|s)d\tau$, $k=1,2$. Note that

$$\Lambda_{Y_2}(t^*)\left\{\int_0^{t^*}\left[\int_s^y \frac{\partial^2[Q_{SY}(\tau,t,s)+Q_{SS}(\tau,t,s,t)]/\partial\tau\partial s}{\partial^2 Q_{SY}(\tau,t,s)/\partial s\partial t}d\tau\right]^{-1}dt\right\}^{-1} = Y_1(y|s).$$

Because we know $\Lambda_{Y_2}(t^*)$ and Λ_{Y_1}, this identifies Δ_1. Similarly, we can identify Y_2 and Δ_2 from Q_{YS} and Q_{SS}.

Now, consider

$$\frac{\partial[Q_{SY}(y,0,s)+Q_{SS}(y,0,s,0)]/\partial s}{\lambda_{S_1}(s)} = \mathcal{L}_G^{(S)}(\Lambda_{Y_1}(s)+Y_1(y|s),\Lambda_{S_1}(s)),$$

for $y>s$. As we know Λ_{Y_1}, Λ_{S_1} and Y_1, we can find a nonempty open set in $\mathbb{R}^2$ on which we know $\mathcal{L}_G^{(S)}$. The function $-\mathcal{L}_G^{(S)}$ is real analytic, so this determines $\mathcal{L}_G^{(S)}$ on $(0,\infty)^2$ (recall the discussion in the proof of Proposition 4). Together with $\mathcal{L}_G(0,0)=1$, this identifies $\mathcal{L}_G$.

(ii) For Model 2b, $t_1, s_1 \in \mathbb{R}_+$ such that $s_1 < t_1$ and $t_2, s_2 \in \mathbb{R}_+$ such that $s_2 < t_2$,

$$\int_{s_1}^{t_1}\left[\int_{s_2}^{t_2}\frac{\partial^3 Q_{SS}(y_1,y_2,s_1,s_2)/\partial s_1\partial s_2\partial y_2}{\partial^3 Q_{SS}(y_1,y_2,s_1,s_2)/\partial s_1\partial s_2\partial y_1}dy_2\right]^{-1}dy_1 = \frac{\Lambda_{\Delta_1}(t_1)-\Lambda_{\Delta_1}(s_1)}{\Lambda_{\Delta_2}(t_2)-\Lambda_{\Delta_2}(s_2)}.$$

Setting $s_1=s_2=0$ and consecutively $t_1=t^*$ and $t_2=t^*$ identifies Λ_{Δ_2} and Λ_{Δ_1}. Identification of G_Δ follows as under (i). *Q.E.D.*

REFERENCES

ABBRING, J. H. (2001): "Econometric Duration and Event-History Analysis," Lecture Notes, Department of Economics, The University of Chicago.

——— (2002): "Stayers versus Defecting Movers: A Note on the Identification of Defective Duration Models," *Economics Letters*, 74, 327–331.

ABBRING, J. H., AND G. J. VAN DEN BERG (2002a): "Analyzing the Effect of Dynamically Assigned Treatments Using Duration Models, Binary Treatment Models, and Panel Data Models," Mimeo, Free University, Amsterdam.

——— (2002b): "A Simple Procedure for Inference on Treatment Effects in Duration Models," Mimeo, Free University, Amsterdam.

——— (2003): "The Identifiability of the Mixed Proportional Hazards Competing Risks Model," *Journal of the Royal Statistical Society, Series B*, 65, 701–710.

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.

BILLINGSLEY, P. (1995): *Probability and Measure*, Third Edition. New York: Wiley.

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.

CARD, D., AND D. SULLIVAN (1988): "Measuring the Effect of Subsidized Training Programs on Movements In and Out of Employment," *Econometrica*, 56, 497–530.

COX, D. R. (1959): "The Analysis of Exponentially Distributed Life-times with Two Types of Failure," *Journal of the Royal Statistical Society, Series B*, 21, 411–421.

——— (1962): *Renewal Theory*. London: Methuen.

——— (1972): "Regression Models and Life-tables (with Discussion)," *Journal of the Royal Statistical Society, Series B*, 34, 187–202.

DAWID, A. P. (2000): "Causal Inference without Counterfactuals," *Journal of the American Statistical Association*, 95, 407–448.

ELBERS, C., AND G. RIDDER (1982): "True and Spurious Duration Dependence: The Identifiability of the Proportional Hazard Model," *Review of Economic Studies*, 64, 403–409.

FLEMING, T. R., AND D. P. HARRINGTON (1991): *Counting Processes and Survival Analysis*. New York: Wiley.

FREEDMAN, D. A. (2002): "On Specifying Graphical Models for Causation, and the Identification Problem," Technical Report 601, Department of Statistics, University of California, Berkeley, CA.

GILL, R. D., AND J. M. ROBINS (2001): "Causal Inference for Complex Longitudinal Data: The Continuous Case," *Annals of Statistics*, 29, 1785–1811.

GRITZ, R. M. (1993): "The Impact of Training on the Frequency and Duration of Employment," *Journal of Econometrics*, 57, 21–51.

HECKMAN, J. J. (2000): "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective," *Quarterly Journal of Economics*, 115, 45–97.

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 *Handbook of Labor Economics, Volume III*, ed. by O. Ashenfelter and D. Card. Amsterdam: North-Holland.

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.

HOLLAND, P. W. (1986): "Statistics and Causal Inference," *Journal of the American Statistical Association*, 81, 945–960.

HOLT, J. D., AND R. L. PRENTICE (1974): "Survival Analysis in Twin Studies and Matched Pair Experiments," *Biometrika*, 61, 17–30.

HONORÉ, B. E. (1991): "Identification Results for Duration Models with Multiple Spells or Time-varying Covariates," Working Paper, Northwestern University, Evanston.

——— (1993): "Identification Results for Duration Models with Multiple Spells," *Review of Economic Studies*, 60, 241–246.

HOROWITZ, J. L. (1999): "Semiparametric Estimation of a Proportional Hazard Model with Unobserved Heterogeneity," *Econometrica*, 67, 1001–1028.

KALBFLEISCH, J. D., AND R. L. PRENTICE (1980): *The Statistical Analysis of Failure Time Data*. New York: Wiley.

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

LICHTENSTEIN, P., M. GATZ, AND S. BERG (1998): "A Twin Study of Mortality after Spousal Bereavement," *Psychological Medicine*, 28, 635–643.

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.

LOK, J. J. (2001): "Statistical Modelling of Causal Effects in Time," Ph.D. Thesis, Division of Mathematics and Computer Science, Faculty of Sciences, Free University, Amsterdam.

MANSKI, C. F. (1990): "Nonparametric Bounds on Treatment Effects," *American Economic Review*, 80, 319–323.

——— (1997): "Monotone Treatment Response," *Econometrica*, 65, 1311–1334.

MILLER, D. R. (1977): "A Note on Independence of Multivariate Lifetimes in Competing Risks Models," *Annals of Statistics*, 5, 576–579.

MORTENSEN, D. T. (1986): "Job Search and Labor Market Analysis," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and R. Layard. Amsterdam: North-Holland.

NEYMAN, J. (1923): "On the Application of Probability Theory to Agricultural Experiments. Essay on Principles," *Roczniki Nauk Rolniczych*, 10, 1–51 (in Polish; edited and translated version of Section 9 by D. M. Dabrowska and T. P. Speed (1990), *Statistical Science*, 5, 465–472).

PEARL, J. (2000): *Causality: Models, Reasoning, and Inference*. Cambridge, UK: Cambridge University Press.

RIDDER, G. (1990): "The Non-parametric Identification of Generalized Accelerated Failure-time Models," *Review of Economic Studies*, 57, 167–182.

RIDDER, G., AND I. TUNALI (1999): "Stratified Partial Likelihood Estimation," *Journal of Econometrics*, 92, 193–232.

ROBINS, J. M. (1998): "Structural Nested Failure Time Models," in *The Encyclopedia of Biostatistics*, ed. by P. Armitage and T. Colton. Chichester: John Wiley and Sons.

RUBIN, D. B. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66, 688–701.

TSIATIS, A. (1975): "A Nonidentifiability Aspect of the Problem of Competing Risks," *Proceedings of the National Academy of Sciences*, 72, 20–22.

VAN DEN BERG, G. J. (2001): "Duration Models: Specification, Identification, and Multiple Durations," in *Handbook of Econometrics, Volume V*, ed. by J. J. Heckman and E. Leamer. Amsterdam: North Holland.

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.