Faculteit der Economische Wetenschappen en Econometrie

Serie research memoranda

The Non-Parametric Identification of the Mixed Proportional Hazards Competing Risks Model

Jaap H. Abbring
Gerard J. van der Berg

Research Memorandum 2000-24

July 2000

/a/l/e/r/t/

applied
labour
economics
research
team

vrije Universiteit amsterdam

The Non-Parametric Identification of the Mixed Proportional Hazards Competing Risks Model

Jaap H. Abbring
Gerard J. van den Berg*

July 21, 2000

Abstract

We prove identification of dependent competing risks models in which each risk has a mixed proportional hazard specification with regressors, and the risks are dependent by way of the unobserved heterogeneity, or frailty, components. We show that the conditions for non-parametric identification given by Heckman and Honor6 (1989) can be relaxed. We generalize the results for the case in which multiple spells are observed for each subject.

Keywords: competing risks, mixed proportional hazard, non-parametric identification, frailty, duration model, multiple spells.

---

*Department of Economics, Free University Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands.
Email: jabbring@econ.vu.nl (Abbring) and gberg@econ.vu.nl (Van den Berg).

1 Introduction

A spell in a state can often end for a number of reasons. Competing risks models interpret the observed duration or failure time as the minimum of a number of competing latent failure times. The model then specifies the distribution of the observed failure time and the corresponding cause of failure as the distribution of the minimum of the competing latent failure times in combination with the identity of the smallest latent failure time. Suppose there are two competing risks, i.e. competing causes of failure, A and B, with corresponding nonnegative random failure times T_A and T_B. The observed failure time T is $T = \min_{i\in\{A,B\}} T_i$ and the cause of failure I is $I = \arg\min_{i\in\{A,B\}} T_i$. Together, T and I are called the "identified minimum" of T_A and T_B. In this paper we focus on continuously distributed failure times (see Crowder, 1996, for results on discrete time competing risks models).

Competing risks models are very commonly used in empirical research (see e.g. the overviews in Kalbfleisch and Prentice, 1980, Yamaguchi, 1991, Andersen et al., 1993, Klein and Moeschberger, 1999). One may argue that any duration analysis of failure time data subject to right-censoring involves competing risks, where the failure of interest constitutes one risk and the censoring time the other, and where the identified minimum is the smallest of the two, taking into account which one is actually smaller (see e.g. Van den Berg, Lindeboom and Ridder, 1993).

It is well known that the joint distribution of (T_A, T_B) is not identified from the joint distribution of (T, I) (see Cox, 1962 and Tsiatis, 1975; Moeschberger and Klein, 1996, provide a survey of the literature). In particular, for any joint distribution of the latent failure times there is a joint distribution with independent latent failure times that generates the same distribution of the identified minimum. (Note that "identifiability" here concerns the invertability of the map ping from the model to the distribution of T, I, and this should not be confused with "identified" in "identified minimum".) The joint distribution of the latent failure times can only be identified if some structure is imposed on it, for example if it is imposed that T_A and T_B are independent.

A particularly popular class of competing risks models assumes that the hazard rates of the latent failure times each have a mixed proportional hazard (MPH) specification, so they depend multiplicatively on the elapsed duration and a set of regressors (or explanatory variables), part of which may be unobserved (Lancaster, 1990, Van den Berg, 2000). If the unobserved determinants (or frailties) are dependent across the risks then the failure times given the observed determinants are dependent. In practice there is often ample reason to suspect that the unobserved determinants are dependent, especially if the subject is an individual whose behavior may affect all hazard rates. We call this class of models the class of MPH competing risks models. The popularity of this class of models is derived from the popularity of the MPH model for single risks. The latter is by far the most popular class of duration models in econometrics, and is also frequently applied in fields like demography and biostatistics. Also, MPH (competing risks) models often serve as building blocks for models of generalized Markovian processes (see Van den Berg, 2000, for an extensive review and references).

Heckman and Honoré (1989) show that, within this class of competing risks models, the model specification is non-parametrically identified if there is sufficient variation of the latent failure times with the regressors and some regularity conditions are satisfied. Here, "non-parametric" means that no parametric functional forms are assumed for the baseline hazards or the multivariate distribution of frailties, while the identifiability concerns the invertability of the mapping from the model determinants (like the baseline hazards and frailty distribution) to the distribution of T, I (which summarizes the population data). In this paper we show that the conditions of Heckman and Honor6 (1989) can be relaxed considerably. In particular, our results allow for less variation in the regressor values, and as such they are likely to be more relevant for applications? We also provide intuition on the identification of the dependence between the risks.

It is important to know whether, under a certain set of conditions, the MPH competing risks model is non-parametrically identified. First of all, if it is non-parametrically identified then the estimates of the model specification may be less sensitive to parametric functional forms on the model determinants, in the sense that the estimates are not completely driven by these functional forms. Secondly, as noted above, the MPH competing risks model is often nested in a larger multi-state model of failure times. In that case it is useful to know whether the information corresponding to the competing risks part is sufficient to identify certain model determinants or whether the estimates of these determinants are completely driven by the information corresponding to other parts of the model.

[1] Heckman and Honor6 (1989) require stronger conditions for identification because they examine a class of models that is somewhat more general than the MPH competing risks model.

Admittedly, identifiability corresponds to a weak qualitative notion of data information, and future work should focus on other properties of the mapping from model to data and to quantitative measures of information (see e.g. Hahn, 1994, Heckman and Taber, 1994, and Klaassen and Lenstra, 1998, in the context of a single risk).

In this paper we also generalize the identification results to the case in which we have multiple spell data, i.e. data on more than one identified minimum for each subject. More precisely, these are data that contain multiple independent drawings from the subject-specific distribution of the identified minimum, so that the unobserved determinants are identical across the spells. Such data are frequently available in, for example, econometric applications (Van den Berg, 2000). In the context of a single risk, it is well known that multiple spell data allow for identification under much less stringent conditions than single spell data (see e.g. Honoré, 1993, for some important results, and Van den Berg, 2000, for a survey of the identification literature). We show that this carries over to competing risks models.

The paper is organized as follows. In Section 2, the MPH competing risks model is introduced. Sections 3 and 4 deal with the identification in case of single spell data and multiple spell data, respectively. Section 5 concludes. Appendix A provides the proofs that are omitted from the main text for expositional purposes.

2 The MPH competing risks model

The MPH model is an extension of the Cox (1972) proportional hazard model (it was introduced by Lancaster, 1979, in econometrics and by Vaupel, Manton and Stallard, 1979, in demography). In particular, it allows for observed as well as unobserved regressors. The survivor function of a single duration T, conditional on only on the observed regressors x, is therefore a mixture of the survivor function conditional on observed and unobserved regressors x and V, respectively. As a result, the class of MPH models is characterized by the survivor functions

$$\Pr(T > t|x) = \mathcal{L}_F(Z(t)\phi(x)), \qquad (1)$$

where $\mathcal{L}_F$ is the Laplace transform of a (proper) distribution F of V with sup port on $[0,\infty)$ such that $F(0) < 1$: $\mathcal{L}_F(s) := \int_0^\infty \exp(-sv)dF(v)$. The "integrated baseline hazard" $Z : [0,\infty) \to [0,\infty)$ is assumed to be nondecreasing and differentiable, with derivative Z', and $Z(0) = 0$. The function $\phi : \mathcal{X} \to (0,\infty)$ is the "regressor function", where $\mathcal{X}$ is the support of x. In applications, this regressor function is frequently specified as $\phi(x) = \exp(x'\beta)$, for some vector β of parameters. However, we will not make such parametric assumptions in this paper.

Note that equation (1) is indeed a mixture of

$$\Pr(T > t|x, V) = \exp\left(-Z(t)\phi(x)V\right)$$

over the distribution F of V.[2] The corresponding hazard rate is $Z'(t)\phi(x)V$ for $T|$ (x, V), which explains the terminology "mixed proportional hazard". The Z' function is called the "baseline hazard", which represents duration dependence at the subject level if subjects are characterized by realizations of (x, V). In applications, such duration dependence is often considered of independent interest, as it can frequently be related to the behavior of the subject under study (see e.g. Van den Berg, 2000). The V factor is usually dubbed the unobserved heterogeneity term or frailty, and is treated as a nuisance component.

The multivariate MPH model allows for a convenient structure of the dependence between the failure times. For expositional clarity, we restrict attention to two risks throughout this paper. The extension to more than two risks is trivial. In the case of two failure times T_A and T_B and a vector of regressors x, the MPH competing risks model specifies the joint survivor function of $(T_A, T_B)\,|x$ as

$$S(t_A, t_B|x) := \Pr(T_A > t_A, T_B > t_B|x) = \mathcal{L}_G(Z_A(t_A)\phi_A(x), Z_B(t_B)\phi_B(x)). \tag{2}$$

where $\mathcal{L}_G$ is the Laplace transform of a (proper) bivariate distribution G with support on $[0,\infty)^2$ such that $\lim_{v\to\infty} G(0,v) < 1$ and $\lim_{v\to\infty} G(v, 0) < 1$:

$$\mathcal{L}_G(s_A, s_B) := \int_0^\infty \int_0^\infty \exp(-s_A v_A - s_B v_B) dG(v_A, v_B)$$

The integrated baseline hazards $Z_A : [0,\infty) \to [0,\infty)$ and $Z_B : [0,\infty) \to [0,\infty)$ again satisfy $Z_A(0) = 0$ and $Z_B(0) = 0$. For expositional convenience, we assume that Z_A and Z_B are continuously differentiable on (0, oo), with derivatives $Z'_A > 0$ and $Z'_B > 0$. The results can be extended straightforwardly to allow for intervals on which $Z'_A = 0$ or $Z'_B = 0$, as in Ridder (1990). Finally, $\phi_A : \mathcal{X} \to (0,\infty)$ and $\phi_B : \mathcal{X} \to (0, \infty)$ are the regressor functions.

[2] Here, it is implicitly understood that either V is independent of x, or F is the distribution of V conditional on x. Explicit assumptions are made in Sections 3 and 4.

As in the univariate case, equation (2) has a mixture interpretation. Let V_A and V_B be nonnegative random variables such that $\Pr(V_A > 0, V_B > 0) > 0$. Then, equation (2) is a mixture of

$$\Pr(T_A > t_A, T_B > t_B | x, V_A, V_B) = \exp\left(-Z_A(t_A)\phi_A(x)V_A - Z_B(t_B)\phi_B(x)V_B\right)$$

over the joint distribution G of (V_A, V_B), with corresponding hazard rates $Z_i'(t)\phi_i(x)V_i$ for $T_i|(x, V_i)$, $i = A, B$. Thus, the dependence of the latent failure times T_A and T_B, conditional on x, runs by way of the stochastic dependence of the unobserved heterogeneity components V_A and V_B.

An interesting feature of the model is that it allows for two different sources of defectiveness of the mixed duration distribution. First, it allows for mass points of either V_A and/or V_B at 0, in which case some fraction of the population never experiences a realization of the events corresponding to T_A and/or T_B. Second, it does not require that $Z_A(t) \to \infty$ and $Z_B(t) \to \infty$ for $t \to \infty$. In other words, it allows for defectiveness of the duration distribution conditional on the unobserved heterogeneity components. In the latter case, the entire population faces a positive probability of never realizing the events corresponding to T_A and/or T_B.

Heckman and Honoré (1989) do not restrict attention to the class of models captured by (2), but they consider a somewhat more general class,

$$S(t_A, t_B|x) = K\left(\exp(-Z_A(t_A)\phi_A(x)), \exp(-Z_B(t_B)\phi_B(x))\right), \tag{3}$$

where K is a joint cumulative distribution function on $[0, 1]^2$. This more general survivor function reduces to the MPH competing risks survivor function in (2) if

$$K(x_A, x_B) = \int_0^\infty \int_0^\infty x_A^{v_A} x_B^{v_B}\, dG(v_A, v_B) \tag{4}$$

for $(x_A, x_B) \in (0, 1]^2$, $K(0, x) = \lim_{x_A \downarrow 0} K(x_A, x)$ and $K(x, 0) = \lim_{x_B \downarrow 0} K(x, x_B)$ for $x \in (0, 1]$, and $K(0, 0) = \lim_{x \downarrow 0} K(0, x) = \lim_{x \downarrow 0} K(x, 0)$. If either V_A (V_B) has a mass point at 0, then $K(0, 1) > 0$ $(K(1, 0) > 0)$: the relevant marginal distribution corresponding to K has a mass point at 0. Obviously, this corresponds to a defectiveness of the corresponding marginal duration distribution. Heckman and Honoré (1989) do not explicitly mention this possibility, and it is not particularly interesting without the specific mixture interpretation offered by the MPH framework. It should be noted that they do discuss defectiveness due to the functions Z_A and Z_B.

It is not difficult to see that the joint distribution of the identified minimum $(T, I)\,|x$ is fully characterized by the functions

$$\begin{aligned} Q_A\,(t|x) &:= \Pr\,(T_A > t, T_B > T_A|x) \text{ and} \\ Q_B\,(t|x) &:= \Pr\,(T_B > t, T_A > T_B|x)\,. \end{aligned} \tag{5}$$

(see Tsiatis, 1975). In the analysis of identification, these functions are taken to be known. Note that $S(t, t|x) = Q_A(t|x) + Q_B(t|x)$. The functions Q_A and Q_B can be characterized explicitly in terms of their derivatives,

$$\frac{\partial Q_i\,(t|x)}{\partial t} = \phi_i\,(x)\,Z_i'\,(t)\,D_i\mathcal{L}_G\,(\phi_A\,(x)\,Z_A\,(t)\,, \phi_B\,(x)\,Z_B\,(t))\,, \quad i = A, B. \tag{6}$$

Here, $D_i\mathcal{L}_G(s_A, s_B) := \partial\mathcal{L}_G(s_A, s_B)/\partial s_i$.

Before presenting the identification results, it is useful to introduce a general result on completely monotone functions, which are frequently encountered in the analysis of MPH models in the form of (derivatives of) Laplace transforms.

Definition 1. Let Ω be a nonempty open set in $\mathbb{R}^n$. A function $f : \Omega \to \mathbb{R}$ is ***absolutely monotone*** if it is nonnegative and has nonnegative continuous partial derivatives of all orders. f is ***completely monotone*** if $f \circ m$ is absolutely monotone, where $m : x \in \{w \in \mathbb{R}^n : -\omega \in \Omega\} \mapsto -x$.

Note that for $n = 1$ this definition reduces to the familiar definitions in Widder (1946). In the sequel, we occasionally refer to the following result.

Proposition 1. ***Let Ψ be a nonempty open connected set in $\mathbb{R}^n$ and let $f : \Psi \to \mathbb{R}$ and $g : \Psi \to \mathbb{R}$ be completely monotone. If f and g agree on a nonempty open set in Ψ, then $f \equiv g$.***

Proof. The proof exploits two facts that are well known for functions on $\mathbb{R}$: (i) completely monotone functions are real analytic and (ii) real analytic functions are uniquely determined by their values on an open set. See Appendix A for details. cl

3 The main identification result

We make the following assumptions on the MPH competing risks model in (2).

Assumption 1. (Independence between observed and unobserved regressors.) G does not depend on x.

Assumption 2. (Variation in observed regressors.) The function $(\phi_A(x), \phi_B(x))$ can attain all values in a nonempty open set $\Phi \subset (0,\infty)^2$ when x varies over X.

Assumption 3. (Normalizations.) For some a priori chosen $t^* \in (0, \infty)$, $Z_A(t^*) = Z_B(t^*) = 1$. For some a priori chosen $x^* \in \mathcal{X}$, $\phi_A(x^*) = \phi_B(x^*) = 1$.

Assumption 4. (Tail of the frailty distribution.) $\lim_{s\downarrow 0} D_A\mathcal{L}_G(s, s) < \infty$ and $\lim_{s\downarrow 0} D_B\mathcal{L}_G(s, s) < \infty$.

Assumption 1 is standard in the MPH literature, and reduces to the stochastic independence assumption $(V_A, V_B) \perp\!\!\!\perp x$ in the mixture interpretation with stochastic V_i. If $\phi_i(x) = \exp(x'\beta_i)$, then it is sufficient for Assumption *2* that x has ***two*** continuous covariates which affect the hazard rates of both risks but with different nonzero coefficients, and which are not perfectly collinear. Note that this assumption is fundamentally weaker than exclusion restrictions of the sort encountered in instrumental variable analysis, where there is a covariate which affects one endogenous variable but not the other.[3] Assumption 3 concerns innocuous normalizations. In the mixture interpretation, Assumption 4 is equivalent to $E(V_i) < \infty$, $i = A$, B, which is a standard assumption in the single spell MPH literature (e.g., Elbers and Ridder, 1982). Ridder (1990) shows that this assumption cannot be omitted without loss of identification.

We have the following result.

Proposition 2. ***If Assumptions 1–4 are satisfied, then the MPH competing risks model (which is characterized by the junctions ϕ_A, ϕ_B, Z_A, Z_B, and $\mathcal{L}_G$) is nonparametrically identified from the distribution of $(T, I)|x$.***

Proof Take an arbitrary $x \in \mathcal{X} : x \neq x^*$, and compute the ratios of both $\partial Q_A(t|x)/\partial t$ and $\partial Q_B(t|x)/\partial t$ at x and x^*. For $i = A, B$, this gives

$$\frac{D_i\mathcal{L}_G\left[\phi_A(x)Z_A(t), \phi_B(x)Z_B(t)\right] \phi_i(x)Z_i'(t)}{D_i\mathcal{L}_G\left[\phi_A(x^*)Z_A(t), \phi_B(x^*)Z_B(t)\right] \phi_i(x^*)Z_i'(t)}. \tag{7}$$

[3] In the case of binary data on the "identified minimum" (i.e., it is observed which duration ends first but not when), exclusion restrictions are necessary to achieve identification. This illustrates the fact that the timing of events in duration data provides a valuable source of information concerning the underlying model.

Cancel $Z_i'(t)$ and let $t \downarrow 0$. Then, by Assumptions 4 and 3, (7) reduces to $\phi_i(x)$.

Next, in equation (2), let $(\phi_A(x), \phi_B(x))$ range over the open set Φ of Assumption 2, for $t_A = t_B$. Then, as we observe $S(t, t|x)$ and because of the complete monotonicity of the bivariate Laplace transform, we can trace out $\mathcal{L}_G$ on $(0, \infty)^2$ by Proposition 1.

Finally, for any given $x \in \mathcal{X}$, we can rewrite equation (6) as a system of two differential equations, in (Z_A, Z_B), (Z_A', Z_B') and t:

$$Z_i' = \frac{\partial Q_i(t|x)}{\partial t} \left[\phi_i(x) D_i \mathcal{L}_G(\phi_A(x) Z_A, \phi_B(x) Z_B)\right]^{-1}, \quad i = A, B \tag{8}$$

with initial conditions that are provided by the normalizations on Z_i in Assumption 3: $Z_A(t^*) = Z_B(t^*) = 1$.

Let the function $f : (0,\infty)^3 \to (0,\infty)^2$ denote the right-hand side of the system of differential equations in (8), as a function of (t, Z_A, Z_B), so that the system can be written as $(Z_A', Z_B') = f(t, Z_A, Z_B)$. Note that f is continuous. By construction, a solution $(Z_A, Z_B) : (0,\infty) \to (0,\infty)^2$ of this system exists. Furthermore, continuity of $\partial f/\partial Z_i$ $(i = m, p)$ on its domain $(0,\infty)^3$ implies Lipschitz continuity of f with respect to Z_A and Z_B. This implies local uniqueness of the solution to the initial conditions. As we already know that a solution exists on $(0, \infty)$, this in turn implies that there is a unique solution on (0, oo). See *e.g.* Walter (1998), Theorem 10.VI. As $Z_A(0) = 0$ and $Z_B(0) = 0$, this implies that $(Z_A(t), Z_B(t))$ is uniquely determined on $[0, \infty)$. □

Note that $\mathcal{L}_G$ in turn identifies G by the uniqueness of the bivariate Laplace transform. Also, note that we can solve equation (8) uniquely for any given $x \in$ X. If we have solutions for any two x, $x' \in$ X, our model restricts these two solutions to be identically the same. This provides overidentifying restrictions similar to those discussed by Melino and Sueyoshi (1990) for the single risk MPH model.

The main difference between our Proposition 2 and the identification result of Heckman and Honoré (1989) is that they tighten Assumption 2 by imposing that $\Phi = (0, \infty)^2$, which is often unlikely to be satisfied in applications. The restriction to MPH competing risks models provides us with the latitude to relax this strong assumption on the regressor effects.[4]

[4] In fact, for identification we only need that $(u, v) \mapsto K(\exp(-u), \exp(-v))$ in equation (4) is real analytic, and not that it is actually a Laplace transform, as in the MPH model. However, as stated before the MPH model is frequently applied and has an attractive mixture interpretation.

It is interesting to obtain some insight into the identification of whether the durations are dependent or not, since this distinguishes the above identification result from the literature in which competing risks models without regressors are examined. We define

$$\theta_A(t_A|x, T_B > t_A)$$

to be the hazard rate of the duration T_A at the value t_A, conditional on x and conditional on the duration T_B exceeding t_A. More generally, the hazard $\theta_A(t_A|x,\ T_B > t_B)$ corresponds to the conditional distribution of $T_A|x,\ T_B > t_B$. We evaluate this hazard for given t_A and t_B, and in fact we take $t_B = t_A$. Obviously, the hazard $\theta_B(t_B|x,\ T_A > t_B)$ can be defined analogically. It is important that the "conditional" hazard rates $\theta_A(t_A|x,\ T_B > t_A)$ and $\theta_B(t_B|x, T_A > t_B)$ can be expressed in terms of the distribution of T, I, so that, in the analysis of identification, these rates are taken to be known.

Assumption 2 implies that $\phi_A(x)$ and $\phi_B(x)$ are not perfectly related, and that there is some independent variation in both. Now suppose that V_A and V_B are independent. Then, $\theta_A(t_A|x,\ T_B > t_A)$ does not vary with $\phi_B(x)$ if $\phi_A(x)$ is held constant. Similarly, $\phi_A(x)$ does not affect $\theta_B(t_B\,|x,\ T_A > t_B)$.

Now let us examine what happens if V_A and V_B are dependent. It is straightforward to show that

$$\theta_A(t_A|x, T_B > t_A) = \frac{\mathrm{E}_V\left[Z'_A(t_A)\phi_A(x)V_A \exp(-Z_A(t_A)\phi_A(x)V_A - Z_B(t_A)\phi_B(x)V_B)\right]}{\mathrm{E}_V\left[\exp(-Z_A(t_A)\phi_A(x)V_A - Z_B(t_A)\phi_B(x)V_B)\right]}$$

with E_V denoting the expectation with respect to the bivariate distribution $G(v_A, v_B)$. If we differentiate this with respect to $\phi_B(x)$ then the resulting expression has the same sign as

$$-\mathrm{Cov}(V_A, V_B|x, T_A > t_A, T_B > t_A)$$

(provided that $t_A > 0$). If V_A and V_B are dependent then in general there are many values of t_A such that the above expression is nonzero. If $\phi_B(x)$ is large then the dynamic selection of individuals with high V_B occurs relatively fast. By conditioning on $T_B > t_A$, we therefore condition on a sub-population with relatively low values of V_B. If V_A and V_B are positively related then this subpopulation also has relatively low values of V_A, and hence a low hazard rate for risk A.

In sum, the derivative of $\theta_A(t_A|x, T_B > t_A)$ with respect to $\phi_B(x)$ and its mirror image for T_B are informative on the dependence or independence of the unobserved heterogeneity terms. This is intuitively very plausible. If the regressor part of the hazard rate of T_B does not directly affect the individual hazard rate of T_A but does affect the observed hazard rate of T_A then this indicates that there is a spurious relation between the durations by way of their unobserved determinants. It should again be stressed that this is not based on an exclusion restriction in the usual sense of the word. All explanatory variables are allowed to affect both duration variables – they are just not allowed to affect both duration distributions in the same way.[5]

4 Identification with multiple spells

So far, we have focused on "single spell" competing risks models, which specify the distribution of the identified minimum (T, I) of a single pair of latent failure times (T_A, T_B) (conditional on regressors x). Instead, assume that for each subject we observe two spells, with identified minima (T_1, I_1, T_2, I_2), with $T_1 = \min_{i\in\{A,B\}} T_{i,1}$, $I_1 = \arg\min_{i\in\{A,B\}} T_{i,1}$, $T_2 = \min_{i\in\{A,B\}} T_{i,2}$, and $I_2 = \arg\min_{i\in\{A,B\}} T_{i,2}$, and with corresponding latent failure times $(T_{A,1}, T_{B,1})$ and $(T_{A,2}, T_{B,2})$.

The survivor function of $(T_{A,1}, T_{B,1}, T_{A,2}, T_{B,2})\,|x$ is given by

$$\begin{aligned} &S(t_{A,1}, t_{B,1}, t_{A,2}, t_{B,2}|x) \\ &:= \Pr(T_{A,1} > t_{A,1}, T_{B,1} > t_{B,1}, T_{A,2} > t_{A,2}, T_{B,2} > t_{B,2}|x) \\ &= \mathcal{L}_G(Z_{A,1}(t_{A,1}|x) + Z_{A,2}(t_{A,2}|x), Z_{B,1}(t_{B,1}|x) + Z_{B,2}(t_{B,2}|x)), \end{aligned} \tag{9}$$

where the distribution G of (V_A, V_B) is now more generally allowed to depend on x. The functions $Z_{i,j} : [0,\infty) \times \mathcal{X} \to [0,\infty)$ $(i = A, B; j = 1,2)$ are increasing in their first argument, with $Z_{i,j}(0|x) = 0$, for all $x \in \mathcal{X}$. Also, for any given $x \in \mathcal{X}$, $Z_{i,j}(t|x)$ is assumed to be continuously differentiable on $(0,\infty)$. In the sequel, we will still refer to the $Z_{i,j}$ as the "integrated baseline hazards", even though these now include regressor effects. It is important to point out that we

[5] Note that the intuitive argument does not use all assumptions we made for full identification. Of course, the $\phi_i(x)$ are not directly observed. We identify these by examining data at zero durations. It is a topic for further research to expand on this by constructing a useful test statistic on independence.

allow the risk-specific baseline hazards, including the way they depend on x, to differ across spells.

With V_A and V_B again nonnegative random variables, we can interpret equation **(4)** as a mixture of

$$\begin{aligned}&\Pr(T_{A,1} > t_{A,1}, T_{B,1} > t_{B,1}, T_{A,2} > t_{A,2}, T_{B,2} > t_{B,2}|x, V_A, V_B)\\&= \exp\left(-Z_{A,1}(t_{A,1}|x)V_A - Z_{B,1}(t_{B,1}|x)V_B - Z_{A,2}(t_{A,2}|x)V_A - Z_{B,2}(t_{B,2}|x)V_B\right)\end{aligned}$$

over the distribution G, which is now the joint conditional distribution G of $(V_A, V_B)|x$. The corresponding hazard rates are $Z'_{i,j}(t|x)V_i$ for $T_{i,j}|(x, V_i)$, where $Z'_{i,j}(t|x) := \partial Z_{i,j}(t|x)/\partial t$ $(i = A, B; j = 1, 2)$. So, conditional on (x, V_A, V_B), the pairs of latent failure times $(T_{A,1}, T_{B,1})$ and $(T_{A,2}, T_{B,2})$ are independent realizations. Thus, we can interpret the model as a model for two spells in a "stratum" that is characterized by a single realization of (V_A, V_B), and where the spells are independent conditional on (x, V_A, V_B).

The stratum could either correspond to a single physical unit, like an individual, for which we observe multiple spells in exactly the same state, or it could consist of single spells corresponding to multiple similar physical units, like for instance a pair of twins. In either case, multiple spell information, i.e. stratification of the data with respect to (V_A, V_B), provides us with multiple realizations of T, I conditional on the same values of the unobservables. It is intuitively clear that such multiple spell data facilitate identification of our model. The analogy with panel data suggests that we can deal with unobserved heterogeneity in multiple spell data by a conditional likelihood approach or a first-differencing approach. However, in our case this is non-trivial because of the nonlinearity of the model. In the remainder of this section, we formally analyze the identification of the multiple spell model.

Consider the following assumption for the multiple spell model:

Assumption 5. (Normalizations.) For some *a priori* chosen $t^* \in (0, \infty)$, $Z_{A,1}(t^*|x) = Z_{B,1}(t^*|x) = 1$, for all $x \in \mathcal{X}$.

This normalization precludes variation of the conditional integrated baseline hazards at $t = t^*$ with x. It is necessary for identification as we allow for general scale effects of x on the conditional distribution G of $(V_A, V_B)|x$. At first sight this might seem restrictive. Consider for example a model with

$$\Pr(T_{A,1} > t|x, V_A) = \exp\left(-Z_{A,1}(t_{A,1})\phi_A(x)V_A\right), \quad \text{and} \quad V_A, V_B|x \sim G(v_A, v_B|x), \tag{10}$$

where ϕ_A is a non-constant positive function on $\mathcal{X}$, and where $Z_{A,1}(t)$ satisfies the part of Assumption 5 that concerns $Z_{A,1}(t|x)$. Then, as $\phi_A(x)$ is not constant, $Z^*_{A,1}(t(z) := Z_{A,1}(t)\phi_A(x)$ does not satisfy the part of Assumption 5 that concerns $Z_{A,1}(t|x)$. Thus, the model in equation (10) does not satisfy Assumption 5. However, there is an observationally equivalent model that does satisfy the assumption. Changing variables $V^*_A := \phi_A(x)V_A$ *in equation* (10) *gives*

$$\Pr(T_{A,1} > t|x, V_A) = \Pr(T_{A,1} > t|x, V^*_A) = \exp\left(-Z_{A,1}(t_{A,1})V^*_A\right), \quad \text{and}$$

$$V^*_A, V_B|x \sim G(v_A/\phi_A(x), v_B|x) =: G^*(v^*_A, v_B|x) \tag{11}$$

This model does satisfy Assumption 5, and it can always be translated back into model (10).[6] We prefer Assumption 5 over an alternative normalization that restricts the dependence of (a scale parameter of) G on x, for the sole reason that we believe that the former is more convenient from an expositional point of view. Note that the issue here is somewhat reminiscent of the role of time-constant regressors in linear panel data models with fixed effects.

We have the following result.

Proposition *3. If Assumption 5 is satisfied, then the functions* $Z_{A,1}, Z_{B,1}, Z_{A,2}$, *and* $Z_{B,2}$ *are non-parametrically identified from the distribution of* $(T_1, I_1, T_2, I_2)\ |x$.

Proof. Pick an arbitrary $x \in \mathcal{X}$. From the distribution of $T_1, I_1, T_2, I_2|x$ we can derive

$$\int_0^{t^*} \frac{\partial \Pr\left(T_{A,1} > \tau, T_{B,1} > T_{A,1}, T_{A,2} > t, T_{B,2} > t|x\right)/\partial\tau}{\partial \Pr\left(T_{A,1} > \tau, T_{B,1} > \tau, T_{A,2} > t, T_{B,2} > T_{A,2}|x\right)/\partial t} d\tau = \frac{1}{Z'_{A,2}(t|x)},$$

[6]In (10) the individual hazard rate varies over x whereas in (11) the frailty distribution among individuals with a given x varies over x. This difference is semantic, except if a physical interpretation is given to what constitutes the frailty, but there is often no reason to do so.

using Assumption 5. This identifies $Z_{A,2}$. In turn, $Z_{A,1}$ is then identified from

$$\int_0^{t^*} \frac{\partial \Pr\left(T_{A,1} > t, T_{B,1} > t, T_{A,2} > \tau, T_{B,2} > T_{A,2}|x\right)/\partial\tau}{\partial \Pr\left(T_{A,1} > t, T_{B,1} > T_{A,1}, T_{A,2} > \tau, T_{B,2} > \tau|x\right)/\partial t} d\tau = \frac{Z_{A,2}(t^*|x)}{Z'_{A,1}(t|x)}.$$

Similarly, we can identify $Z_{B,1}$ and $Z_{B,2}$. cl

Having identified the integrated baseline hazards, the natural next step is to use these in identifying $\mathcal{L}_G$. It is not difficult to see that, for any given $x \in \mathcal{X}$, we can identify $\mathcal{L}_G$ and its first and second order partial derivatives on $\mathcal{Z}_x := \{\zeta(t_1, t_2|x) : (t_1, t_2) \in (0, \infty)^2\} \subset (0, \infty)^2$, where $\zeta(t_1, t_2|x) := (Z_{A,1}(t_1|x) + Z_{A,2}(t_2|x), Z_{B,1}(t_1|x) + Z_{B,2}(t_2|x))$. Note that ζ is identified under Assumption 5. As $\lim_{t\downarrow 0} \zeta(t,t|x) = (0,0)$, we can identify the first and second moments of G, for each $x \in \mathcal{X}$. However, without further assumptions on the effects of the covariates x, we cannot exploit variation in x as in the single spell case, and we have to identify G from variation in ζ for given $x \in X$. Without further restrictions on the integrated baseline hazards, G may not be identified, as the following counter-example shows.

For given $x \in \mathcal{X}$, suppose that $Z_{B,1}(t|x) \equiv kZ_{A,1}(t|x)$ and $Z_{B,2}(t|x) \equiv kZ_{A,2}(t|x)$, for some constant $k > 0$. This implies that $\mathcal{Z}_x = \{(z, kz) : z \in (0, \infty)\}$ is simply a curve in $(0,\infty)^2$. Then, for given $x \in \mathcal{X}$, we can only identify the bivariate transform $\mathcal{L}_G$ and its first and second derivatives on this single curve, which cannot be extended to the entire $(0,\infty)^2$ as required for identification of $\mathcal{L}_G$.

The following assumption excludes such cases, without exploiting variation in x.

Assumption 6. (Variation in baseline hazards.) For each $x \in \mathcal{X}$, there is a $(\tau_1, \tau_2) \in (0, \infty)^2$, which may depend on x, such that

$$Z'_{A,1}(\tau_1|x)Z'_{B,2}(\tau_2|x) \neq Z'_{B,1}(\tau_1|x)Z'_{A,2}(\tau_2|x).$$

Assumption 6 is not very restrictive. For example, suppose that, for given $x \in \mathcal{X}$, both $Z_{A,1}(t|x) \equiv t$ and $Z_{A,2}(t|x) \equiv t$, and $Z_{B,1}(t|x) \equiv Z_{B,2}(t|x)$. Then, it requires that $Z_{B,1}$ is not linear on all of $(0,\infty)$. In general, Assumption 6 ensures that, for each $x \in \mathcal{X}$, there is a $(\tau_1, \tau_2) \in (0,\infty)^2$ such that $\zeta(t_1, t_2|x)$ is an open mapping locally around (τ_1, τ_2). In turn this implies that we can trace $\mathcal{L}_G$ on an open set that contains $\zeta(\tau_1, \tau_2|x)$ by varying (t_1, t_2) over an open set containing

(τ_1, τ_2). Then, variation in the baseline hazards can replace regressor variation in the multiple spell case. Formally, we have

Proposition 4. ***If Assumptions* 5 *and* 6 *are satisfied, then the multiple spell MPH competing risks model (which is characterized by the functions*** $Z_{A,1}$, $Z_{B,1}$, $Z_{A,2}$, $Z_{B,2}$, ***and*** $\mathcal{L}_G$***) is non-parametrically identified from the distribution of*** $(T_1, I_1, T_2, I_2)|x$.

Proof. $Z_{A,1}$, $Z_{B,1}$, $Z_{A,2}$, $Z_{B,2}$ are identified by Proposition 3. As a consequence, ζ is identified. Next, as, for given $x \in \mathcal{X}$, ζ is continuously differentiable, it is an open mapping locally around (τ_1, τ_2) as in Assumption 6, by direct implication of the inverse function theorem. Thus, for each $x \in \mathcal{X}$, we can trace $\mathcal{L}_G$ on an open set by varying (t_1, t_2) over an open set that contains (τ_1, τ_2). For each $x \in \mathcal{X}$, this identifies $\mathcal{L}_G$ by Proposition **1.** cl

It follows from Proposition 3 that Assumption 6 is identified, i.e. can be tested, under Assumption 5. If Assumption 6 is not satisfied, we have to rely on alternative assumptions, which guarantee that we can exploit variation in x as in the single spell case.

First, we need independence of (V, V_B) and x as in Assumption **1.**

Assumption 7. (Independence between observed and unobserved regressors, up to a scale factor.) There are functions $\phi_A : \mathcal{X} \to (0, \infty)$ and $\phi_B : \mathcal{X} \to (0, \infty)$, and a distribution function $G^* : [0, \infty)^2 \to [0, 1]$ that does not depend on x, such that $G(u/\phi_A(x), v/\phi_B(x)) \equiv G^*(u, v)$. For some a priori chosen $x^* \in \mathcal{X}, \phi_A(x^*) = \phi_B(x^*) = 1$.

It should be noted that we cannot simply require independence of (V_A, V_B) and x, as G is supposed to absorb multiplicative regressor effects at $t = t^*$ (see the discussion of Assumption 5). The functions ϕ_A and ϕ_B in Assumption 7 can be thought of as the multiplicative regressor effects at $t = t^*$. If we rewrite the model in terms of G^*, we get

$$S(t_{A,1}, t_{B,1}, t_{A,2}, t_{B,2}|x) =$$
$$\mathcal{L}_{G^*}\left((Z_{A,1}(t_{A,1}|x) + Z_{A,2}(t_{A,2}|x))\phi_A(x), (Z_{B,1}(t_{B,1}|x) + Z_{B,2}(t_{B,2}|x))\phi_B(x)\right).$$

The regressor functions ϕ_A and ϕ_B enter proportionally in the conditional hazard rates of the transformed model. Thus, Assumption 7 reduces the second step identification problem to the identification of ϕ_A, ϕ_B, and a distribution G^* that does not depend on x. The normalization is innocuous, as we leave the scale of G* unnormalized.

Define $\mathcal{Z}^* := \{\zeta^*(t_1,t_2|x) : (t_1,t_2) \in (0, \infty)^2, x \in X\}$, with $\zeta^*(t_1,t_2|x) := ((Z_{A,1}(t_1|x)+Z_{A,2}(t_2|x))\phi_A(x), (Z_{B,1}(t_1|x)+Z_{B,2}(t_2|x))\phi_B(x))$. If we can identify ζ^*, we can trace $\mathcal{L}_{G^*}$ on $\mathcal{Z}^*$. The following assumption ensures that there is sufficient variation of ζ^*.

Assumption 8. (Variation in observed regressors.) There is a nonempty open set Φ^* such that $\Phi^* \subset \mathcal{Z}^*$.

A sufficient condition for Assumption 8 is that (ϕ_A, ϕ_B) satisfies Assumption 2, but more subtle conditions are obviously sufficient.

We cannot directly apply Assumption 8, as we have not shown that ϕ_A and ϕ_B, and therefore ζ^*, are identified. Indeed, it is clear that these functions are not identified from within-stratum variation, as ϕ_A and ϕ_B are time-invariant and appear proportionally in the hazard rates of each of the spells in a stratum. This can be solved by also imposing the finite means Assumption 4, so that we can identify ϕ_A and ϕ_B by evaluating the mixture hazard rates near *0*, as in the single spell case.

Thus, we have

Proposition 5. *If Assumptions 4, 5, 7, and 8 are satisfied, then the multiple spell MPH competing risks model (which is characterized by the functions $Z_{A,1}$, $Z_{B,1}$, $Z_{A,2}$, $Z_{B,2}$, and $\mathcal{L}_G$) is non-parametrically identified from the distribution of $(T_1, I_1, T_2, I_2)|x$.*

Proof. Again, $Z_{A,1}$, $Z_{B,1}$, $Z_{A,2}$, $Z_{B,2}$ are identified as in Proposition 3. Identification of ϕ_A, ϕ_B and G* follows from the obvious multiple spell equivalent to the first two steps of the proof of the single spell Proposition 2, where $(\phi_A(x), \phi_B(x))$ is replaced by $\zeta^*(t_1,t_2|x)$, and Φ by Φ^*. cl

Without Assumption 6, much of the strength of the multiple spell argument is lost. Even in this case, however, we are still able to allow for general nonproportionality between duration and regressor effects in the conditional hazard rates.

We end this section by concluding that with multiple spell data, the integrated baseline hazards and regressor effects are identified without most of the assumptions used for the single spell result. In particular, we do not need finite means of the frailties, or independence between (V_A, V_B) and x. Also, x may enter in an arbitrary nonproportional manner in the conditional hazard rates, and we do not even need variation of x. These results are in line with Honore (1993), who derives identification results for single risk MPH models if multiple spell data are available. However, we cannot fully identify the mixing distribution under the same weak conditions as in Honore (1993). The competing risks nature of the data complicates tracing the bivariate Laplace transform of this distribution.

Proposition 4 shows that an additional minor and testable restriction on the integrated baseline hazards is sufficient to identify G without further assumptions on the role of x. If this condition does not hold, we can still allow for general non-proportionalities between duration and regressor effects in the conditional hazard rates. This result, as stated in Proposition 5, does however rely on regressor variation, finite means of the unobservables, and independence of the unobservables and the regressors.

5 Conclusion

In this paper we show that the conditions for non-parametric identification of the dependent competing risks model with regressors, as given by Heckman and Honore (1989), can be relaxed. In particular, Heckman and Honore (1989) require the hazard rates corresponding to the latent failure times to cover all values in $(0,\infty)^2$ by varying the regressor values over their support. Instead, we only need these hazard rates to vary over a nonempty open subset of $(0,\infty)^2$ by varying the regressor values. In applications, the latter condition is much more likely to be satisfied.

With multiple spell data, the integrated baseline hazards and regressor effects are identified without most of the assumptions used for the single spell result. In particular, we do not need to assume independence between the observed and unobserved regressors. Also, the observed regressors may enter in an arbitrary nonproportional manner in the conditional hazard rates, and we do not even need variation of observed regressors.

The multiple spell results are quite general and can be derived without most of the assumptions used for the single spell result. It should be noted, however, that these results are particularly useful if we have "ideal" data, containing a complete set of multiple durations for each subject. In practice, censoring of multiple spell data may be more problematic than censoring of single spell data, and this somewhat reduces the relevance of multiple spell identification results. See Visser (1996) and Ridder and Tunalı (1999) for discussions of these problems in the context of single risk duration models.

Acknowledgements

Gerard J. van den Berg is also affiliated to the Tinbergen Institute and the CEPR. We thank Paul Frijters, Jim Heckman, workshop participants at the University of Chicago and the University of Western Ontario, and participants at the Econometric Study Group meeting 2000 in Bristol, for their comments.

Appendix A Proof of Proposition 1

We need the following definition of real analyticity, adapted from Narasimhan (1971).

Definition 2. Let Ω be a nonempty open set in $\mathbb{R}^n$. The function $f : \Omega \to \mathbb{R}$ is *real analytic* if to each $\omega \in \Omega$ corresponds a power series in x − w that converges to f(x) for all x in some neighborhood $U \subset \Omega$ of w.

The following lemma is proven in Widder (1946) for the special case of $n = 1$ (Theorem 3a in Chapter IV). This lemma with $n = 1$ is sometimes called S. Bernstein's Theorem (e.g., Krantz and Parks, 1992, Theorem 2.4.1). Here we prove it for general n.

Lemma 1. *Let Ω be a nonempty open set in $\mathbb{R}^n$. If $f : \Omega \to \mathbb{R}$ is absolutely monotone, then f is real analytic.*

Proof. Let $w \in \Omega$, and let $\rho > 0$ be such that $\omega + h \in \Omega$ for $h \in U_n(\rho) := \{\eta \in \mathbb{R}^n : (\eta'\eta)^{1/2} < \rho\}$. For functions $f : \mathbb{R}^n \to \mathbb{R}$ define

$$D_i f(x) := \frac{\partial}{\partial x_i} f(x),$$

where $x := (x_1, \ldots, x_n)$. Let D be the $n \times 1$-vector $(D_1, \ldots, D_n)$, so that $Df(x) = \partial f(x)/\partial x$. By Taylor's Theorem with exact remainder (e.g., Widder, 1961), we have that

$$f(w+h) = \sum_{j=0}^{k} \frac{1}{j!} (h'D)^j f(w) + R_k(\omega, h),$$

with

$$R_k(\omega, h) = \int^1 \frac{(1-t)^k}{k!} (h'D)^{k+1} f(w + th)dt,$$

for $h \in U_n(\rho)$.

Now, take any $h := (h_1, \ldots, h_n) \in U_n(n^{-1/2}\rho)$. Define $a := \max\{|h_1| \ldots, |h_n|\}$, and denote the $n \times 1$-unit vector by e_n. Note that $0 \le a < n^{-1/2}\rho$, which implies that $ae_n \in U_n(\rho)$. Take any $b \in \mathbb{R}$ such that $a < b < n^{-1/2}\rho$. Then,

$$\begin{aligned}
0 \le |R_k(\omega, h)| &\le \int^1 \frac{(1-t)^k}{k!} (|h|'D)^{k+1} f(w + th)dt \\
&\le a^{k+1} \int^1 \frac{(1-t)^k}{k!} (e_n'D)^{k+1} f(\omega + th)dt \\
&\le \left(\frac{a}{b}\right)^{k+1} \int_0^1 \frac{(1-t)^k}{k!} (be_n'D)^{k+1} f(w + tbe_n)dt \\
&= \left(\frac{a}{b}\right)^{k+1} R_k(\omega, be_n) \\
&\le \left(\frac{a}{b}\right)^{k+1} f(w + be_n) \longrightarrow 0 \quad \text{as } k \to \infty.
\end{aligned}$$

cl

Obviously, Lemma 1 implies that if a function $f : \Omega \to \mathbb{R}$ is completely monotone on a nonempty open set Ω in $\mathbb{R}^n$, then f is real analytic.

Narasimhan (1971) shows that if $f : \Psi \to \mathbb{R}$ is real analytic on a nonempty open connected set Ψ in $\mathbb{R}^n$, and f vanishes on a nonempty open subset of Ψ, then $f \equiv 0$ (Narasimhan, 1971, Proposition 1 in Chapter 1 and Remark 2 on page 4). Proposition 1 now follows immediately, as the difference of two real analytic functions is real analytic.

References

Andersen, P.K., 0. Borgan, R.D. Gill, and N. Keiding (1993), ***Statistical Models Based on Counting Processes,*** Springer, New York.

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

Cox, D.R. (1972), "Regression models and life-tables (with **discussion)",** ***Journal of the Royal Statistical Society Series B, 34,*** 187–202.

Crowder, M. (1996), "On assessing independence of competing risks when failure times are discrete", ***Lifetime Data Analysis, 2,*** **195-209.**

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.

Hahn, J. (1994), "The efficiency bound of the mixed proportional hazard model", ***Review of Economic Studies,*** **61, 607-629.**

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

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. (1993), "Identification results for duration models with multiple spells", ***Review of Economic Studies, 60,*** **241-246.**

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

Klaassen, C.A.J. and A.J. Lenstra (1998), "The information for the treatment effect in the mixed proportional hazards model vanishes", **Working paper,** University of Amsterdam, Amsterdam.

Klein, J.P. and M.L. Moeschberger (1999), ***Survival Analysis,*** Springer, New York.

Krantz, S.G. and H.R. Parks (1992), ***A Primer of Real Analytic Functions,*** Birkhäuser Verlag, Basel.

Lancaster, T. (1979), "Econometric methods for the duration of unemployment", ***Econometrica, 47, 939-956.***

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

Melino, A. and G.T. Sueyoshi (1990), "A simple approach to the identifiability of the proportional hazards model", ***Economics Letters, 33, 63-68.***

Moeschberger, M.L. and J.P. Klein (1995), "Statistical methods for dependent competing risks", ***Lifetime Data Analysis,*** 1, **195-204.**

Narasimhan, R. (1971), ***Several Complex Variables,*** The University of Chicago Press, Chicago/London.

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

Ridder, G. and I. Tunalı (1999), "Stratified partial likelihood estimation", ***Journal of Econometrics,*** **92, 193-232.**

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. (2000), "Duration models: Specification, identification, and multiple durations", in J.J. Heckman and E. Learner, editors, ***Handbook of Econometrics, Volume V,*** North Holland, Amsterdam.

Van den Berg, G.J., M. Lindeboom, and G. Ridder (MM), "Attrition in longitudinal panel data, and the empirical analysis of dynamic labour market behaviour" , ***Journal*** *of* ***Applied Econometrics, 9,*** **421-435.**

Vaupel, J.W., K.G. Manton, and E. Stallard (1979), "The impact of heterogeneity in individual frailty on the dynamics of mortality", ***Demography,*** **16, 439-454.**

Visser, M. (1996), "Nonparametric estimation of the bivariate survival function with an application to vertically transmitted AIDS", ***Biometrika, 83,*** **507-518.**

Walter, W. (1998), ***Ordinary Differential Equations,*** volume 182 of ***Graduate Texts in Mathematics,*** Springer-Verlag, Heidelberg.

Widder, D.V. (1946), ***The Laplace Transform,*** Princeton University Press, Princeton.

Widder, D.V. (1961), ***Advanced Calculus,*** Prentice Hall, Englewood Cliffs, NJ, second edition.

Yamaguchi, K. (1991), ***Event History Analysis,*** Sage, Newbury Park.