The Identifiability of the Mixed Proportional Hazards Competing Risks Model

Jaap H. Abbring*

Gerard J. van den Berg†

Forthcoming in J. Royal Statistical Society: Series B
May 6, 2003

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 identification given by Heckman and Honoré (1989) can be relaxed. We extend the results to the case in which multiple spells are observed for each subject.

---

*Department of Economics, Free University (Vrije Universiteit), De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands. Email: `jabbring@econ.vu.nl`.

†Department of Economics, Free University Amsterdam, Tinbergen Institute, and CEPR.
Email: `gberg@econ.vu.nl`.

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

We thank Paul Frijters, Jim Heckman, the editor, two anonymous referees, workshop participants at the University of Chicago and the University of Western Ontario and participants in the 2000 Econometric Study Group meeting in Bristol and the 2000 World Congress of the Econometric Society in Seattle for their comments. Jaap Abbring acknowledges financial support by the Royal Netherlands Academy of Arts and Sciences.

1 Introduction

A spell in a state can often end for a number of reasons. Competing risks models specify the observed duration or failure time as the minimum of a number of competing latent failure times and the corresponding cause of failure as 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 jointly continuous nonnegative random failure times T_A and T_B. The extension to more than two risks is trivial and will not be considered in this paper. The observed failure time T equals $\min_{i\in\{A,B\}} T_i$ and the cause of failure I is $\arg\min_{i\in\{A,B\}} T_i$. Jointly, (T, I) is called the *identified minimum* of T_A and T_B.

It is well known that the joint distribution of (T_A, T_B) is not identified from the joint distribution of (T, I) (Cox, 1959, 1962; Tsiatis, 1975). 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. The joint distribution of the latent failure times can only be identified if some additional structure is imposed, for example independence of T_A and T_B.

A particular popular class of competing risks models assumes that the hazard rates of the latent failure times have mixed proportional hazard (MPH) specifications, so that they depend multiplicatively on the elapsed duration, observed regressors and unobserved heterogeneity, or frailty, components (Lancaster, 1990; Van den Berg, 2001). If the unobserved determinants are dependent across the risks then the failure times are dependent given the regressors. In practice there is often ample reason to expect such dependence, in particular if the subject is an individual whose behavior may affect all hazard rates.

Heckman and Honoré (1989) consider a model that nests the MPH competing risks model. They show that the model is identified if there is sufficient variation of the latent failure times with the regressors. Here, identifiability concerns the invertibility of the mapping from the model determinants to the distribution of (T, I) (which summarizes the population data). Identification is nonparametric in the sense that no parametric functional forms are assumed for the model determinants (like the baseline hazards and the frailty distribution in the MPH case). Identifiability is useful because it implies that the estimates of the model specification are not completely driven by parametric functional-form assumptions on the model determinants.

In this paper we show that the conditions of Heckman and Honoré (1989) can be relaxed considerably in the MPH case. In particular, our results require much less variation of the latent failure times with the regressors. As such, this paper provides conditions for identification for the case in which Heckman and Honoré (1989)'s assumption on the covariate effects is not satisfied by the data. This is relevant to empirical work, as in many applications our condition will be satisfied, whereas Heckman and Honoré (1989)'s stronger condition fails to hold.

In this paper we also extend the identification analysis to the case with multiple-spell data, i.e. data on more than one identified minimum for each subject. This extension to multiple spells is quite natural in the MPH framework. Within this framework, multiple-spell data can be viewed as providing multiple independent draws 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, 2001). 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 (Honoré, 1993). 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.

2 The mixed proportional hazards competing risks model

The MPH model is an extension of the Cox (1972) proportional hazard model introduced by Lancaster (1979) and Vaupel, Manton and Stallard (1979). The bivariate MPH model is a convenient framework to model the dependence of the latent failure times T_A and T_B. It traces all such dependence to related observed and unobserved determinants of both durations. More formally, it specifies that T_A and T_B are independent conditional on (x, V_A, V_B), where x is a vector of (observed) regressors and V_A and V_B are unobserved nonnegative random variables that are distributed independently of x such that $\Pr(V_A > 0, V_B > 0) > 0$. The distribution of $(T_A, T_B)|(x, V_A, V_B)$ then factorizes in the marginal distributions of $T_A|(x, V_A, V_B)$ and $T_B|(x, V_A, V_B)$, which are fully characterized by the corresponding hazard rates,

$$\theta_A(t|x, V_A, V_B) = \lambda_A(t)\phi_A(x)V_A \quad \text{and} \quad \theta_B(t|x, V_A, V_B) = \lambda_B(t)\phi_B(x)V_B.$$

The *baseline hazards* $\lambda_A : \mathbb{R}_+ \to (0, \infty)$ and $\lambda_B : \mathbb{R}_+ \to (0, \infty)$ have integrals

$$\Lambda_A(t) := \int_0^t \lambda_A(\tau)d\tau < \infty \quad \text{and} \quad \Lambda_B(t) := \int_0^t \lambda_B(\tau)d\tau < \infty$$

for all $t \in \mathbb{R}_+ := [0, \infty)$. $\phi_A : \mathcal{X} \to (0, \infty)$ and $\phi_B : \mathcal{X} \to (0, \infty)$ are continuous *regressor functions*, with $\mathcal{X}$ the support of x. In applications, these functions are frequently specified as $\phi_A(x) = \exp(x'\beta_A)$ and $\phi_B(x) = \exp(x'\beta_B)$ for some parameter vectors β_A and β_B. We will not make such parametric assumptions. We normalize

$$\Lambda_A(t^*) = \Lambda_B(t^*) = 1 \qquad \text{and} \qquad \phi_A(x^*) = \phi_B(x^*) = 1$$

for some a priori chosen $t^* \in (0, \infty)$ and $x^* \in \mathcal{X}$. These normalizations are innocuous because V_A and V_B can capture the scale of θ_A and θ_B.

Using the conditional independence of T_A and T_B and standard expressions for the marginal survival functions of $T_A|(x, V_A, V_B)$ and $T_B|(x, V_A, V_B)$, we get

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

The joint survival function of $(T_A, T_B)|x$ then follows by taking the expectation over (V_A, V_B) with respect to the distribution G of (V_A, V_B), which gives

$$S(t_A, t_B|x) := \Pr(T_A > t_A, T_B > t_B|x) = \mathcal{L}_G(\Lambda_A(t_A)\phi_A(x), \Lambda_B(t_B)\phi_B(x)). \qquad (1)$$

Here $\mathcal{L}_G$ is the Laplace transform of G, i.e.

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

An interesting feature of the model is that it allows for two different sources of defectiveness of the distribution of $(T_A, T_B)|x$. First, the unobserved heterogeneity components V_A and V_B may have mass points at 0. Second, we allow that $\lim_{t\to\infty} \Lambda_A(t) < \infty$ and $\lim_{t\to\infty} \Lambda_B(t) < \infty$. Abbring (2002) provides discussion and examples of applications.

Heckman and Honoré (1989) do not restrict attention to the class of models captured by equation (1), but consider a somewhat more general specification,

$$S(t_A, t_B|x) = K\left(\exp(-\Lambda_A(t_A)\phi_A(x)), \exp(-\Lambda_B(t_B)\phi_B(x))\right), \qquad (2)$$

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

$$K(x_A, x_B) = \int_0^\infty \int_0^\infty x_A^{v_A} x_B^{v_B} dG(v_A, v_B). \tag{3}$$

3 The main identification result

First, note that the joint distribution of the identified minimum $(T, I)|x$ is fully characterized by the *sub-survival functions* (Tsiatis, 1975)

$$Q_A(t|x) := \Pr(T_A > t, T_B > T_A|x) \quad \text{and} \quad Q_B(t|x) := \Pr(T_B > t, T_A > T_B|x). \tag{4}$$

In the analysis of identification, $Q_A(\cdot|x)$ and $Q_B(\cdot|x)$ are taken to be known for all $x \in \mathcal{X}$. Note that $S(t,t|x) = Q_A(t|x) + Q_B(t|x)$. The sub-survival functions can be characterized explicitly in terms of the corresponding *sub-densities*, which are given by

$$-Q_i'(t|x) = -\lambda_i(t)\phi_i(x) D_i\mathcal{L}_G(\Lambda_A(t)\phi_A(x), \Lambda_B(t)\phi_B(x)), \quad i = A, B, \tag{5}$$

for almost all t. Here, $Q_i'(t|x) := \partial Q_i(t|x)/\partial t$ and $D_i\mathcal{L}_G(s_A, s_B) := \partial\mathcal{L}_G(s_A, s_B)/\partial s_i$.

We need a general result on completely monotone functions.

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 $m : x \in \{\omega \in \mathbb{R}^n : -\omega \in \Omega\} \mapsto -x$.

Note that for $n = 1$ this definition reduces to the familiar definitions in Widder (1946).

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 $f = g$.*

The proof is available from the authors upon request. It exploits two facts that are well-known for functions on $\mathbb{R}$ and that are also true for functions on $\mathbb{R}^n$: (i) completely monotone functions are real analytic and (ii) real analytic functions are uniquely determined by their values on a nonempty open set.

We make the following assumptions on the MPH competing risks model in (1).

Assumption 1. (Variation with observed regressors) $\{(\phi_A(x), \phi_B(x))\,; x \in \mathcal{X}\}$ contains a nonempty open set $\Phi \subset \mathbb{R}^2$.

Assumption 2. (Tail of the frailty distribution) $\mathbb{E}[V_A] < \infty$ and $\mathbb{E}[V_B] < \infty$.

Heckman and Honoré (1989) tighten Assumption 1 by imposing that $\Phi = (0, \infty)^2$. The restriction to MPH competing risks models provides us with the latitude to relax this strong assumption on the regressor effects. With two regressors and $\phi_i(x) = \exp(x'\beta_i)$, it is sufficient for Assumption 1 that $(\beta_A\ \beta_B)$ has full rank and $\mathcal{X}$ contains a non-empty open set in $\mathbb{R}^2$. Note that Assumption 1 is fundamentally weaker than exclusion restrictions of the sort encountered in instrumental variable analysis, which require a regressor that affects one endogenous variable but not the other. Assumption 2 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. *Under Assumptions 1–2, the MPH competing risks model (characterized by the functions ϕ_A, ϕ_B, Λ_A, Λ_B, and $\mathcal{L}_G$) is identified from the distribution of $(T, I)|x$.*

Proof. The proof successively establishes identification of (i) (ϕ_A, ϕ_B), (ii) $\mathcal{L}_G$, by exploiting Proposition 1 and the variation in $(\phi_A(x), \phi_B(x))$, and (iii) (Λ_A, Λ_B), as the unique solution to an initial value problem involving data, $(\phi_A(x), \phi_B(x))$ (for arbitrary $x \in \mathcal{X}$) and $\mathcal{L}_G$.

(i) The regressor functions ϕ_A and ϕ_B.
Pick an arbitrary $x \in \mathcal{X}$. For almost all t, $Q_A(\cdot|x)$ and $Q_A(\cdot|x^*)$ are differentiable and

$$\frac{Q'_A(t|x)}{Q'_A(t|x^*)} = \phi_A(x)\frac{D_A\mathcal{L}_G\left[\Lambda_A(t)\phi_A(x), \Lambda_B(t)\phi_B(x)\right]}{D_A\mathcal{L}_G\left[\Lambda_A(t), \Lambda_B(t)\right]}, \tag{6}$$

where we use $\phi_A(x^*) = \phi_B(x^*) = 1$. As $t \downarrow 0$, (6) reduces to $\phi_A(x)$ because $D_A\mathcal{L}_G(\cdot) \longrightarrow \mathbb{E}[V_A] < \infty$ by Assumption 2. Note that here it is crucial that V_A is independent of x. Since x is arbitrary, this identifies ϕ_A. Identification of ϕ_B is analogous.

(ii) The Laplace transform $\mathcal{L}_G$ of the frailty distribution.
Evaluating equation (1) at $t_A = t_B = t^*$ gives $S(t^*, t^*|x) = \mathcal{L}_G(\phi_A(x), \phi_B(x))$ because $\Lambda_A(t^*) = \Lambda_B(t^*) = 1$. Note that $S(t^*, t^*|x)$ is observed and (ϕ_A, ϕ_B) is identified in *(i)*.

So, we can let $(\phi_A(x), \phi_B(x))$ range over the set Φ of Assumption 1 to trace out $\mathcal{L}_G$ on a nonempty open set. As $\mathcal{L}_G$ is completely monotone, this identifies $\mathcal{L}_G$ by Proposition 1.

(iii) The integrated baseline hazards Λ_A and Λ_B.

Pick an arbitrary $x \in \mathcal{X}$. We can rewrite equation (5) as a system of differential equations in the sense of Carathéodory (1918), i.e. for almost all $t \in (0, \infty)$

$$\Lambda'(t) = f\left(t, \Lambda(t)\right), \quad \text{with initial conditions } \; \Lambda_A(t^*) = \Lambda_B(t^*) = 1. \tag{7}$$

Here, $f_i(t, \Lambda(t)) := Q_i'\left(t|x\right)\left[\phi_i(x) D_i \mathcal{L}_G\left(\phi_A\left(x\right)\Lambda_A(t), \phi_B\left(x\right)\Lambda_B(t)\right)\right]^{-1}$, $i = A, B$, and $\Lambda := (\Lambda_A, \Lambda_B)$ and $f := (f_A, f_B)$. The function f is known, as we observe the functions Q_A' and Q_B' and have identified the numbers $\phi_A(x)$, $\phi_B(x)$ and the function $\mathcal{L}_G$ in *(i)* and *(ii)*. Standard theory implies that (7) has a unique solution Λ on $[0, \infty)$ in terms of f, and the remainder of the proof demonstrates this. Write $f_i(t, \Lambda) = -Q_i'\left(t|x\right) r_i(\Lambda)$, with $r_i(\Lambda) := -\left[\phi_i(x) D_i \mathcal{L}_G(\phi_A\left(x\right)\Lambda_A, \phi_B\left(x\right)\Lambda_B)\right]^{-1}$, $i = A, B$. Note that $r := (r_A, r_B)$ is continuously differentiable and, by implication, satisfies a Lipschitz condition on compact sets $K \subset (0, \infty)^2$. Because $|Q_A'(\cdot|x) + Q_B'(\cdot|x)|$ is integrable on compact sets $J \subset (0, \infty)$ and

$$||f(t, \Lambda) - f(t, \tilde{\Lambda})|| \leq |Q_A'(t|x) + Q_B'(t|x)| \cdot ||r(\Lambda) - r(\tilde{\Lambda})|| \tag{8}$$

for all $(t, \Lambda), (t, \tilde{\Lambda}) \in J \times K$, this implies that f satisfies a generalized Lipschitz condition with respect to Λ on $J \times K$ for all compact J and K (Walter, 1998, Section 10, Supplement II). By Walter (1998), Theorem 10.XX(b), this implies that (7) has a unique solution Λ on $(0, \infty)$. With $\Lambda_A(0) = 0$ and $\Lambda_B(0) = 0$, this uniquely determines Λ on $[0, \infty)$. $\square$

Note that $\mathcal{L}_G$ in turn identifies G by the uniqueness of the bivariate Laplace transform. Also, note that step *(iii)* of the proof can be repeated for all $x \in \mathcal{X}$. This would give a range of unique solutions Λ to (7). Obviously, all these solutions should be the same, which provides overidentifying restrictions similar to those discussed by Melino and Sueyoshi (1990) for the single-risk MPH model.

To break the non-identification result of Cox (1959, 1962) and Tsiatis (1975) we exploit that we can independently vary $\phi_A(x)$ and $\phi_B(x)$. Some intuition can be derived from the (observed) *crude* hazard rate

$$\frac{-Q_A'(t|x)}{S(t, t|x)} = \lambda_A\left(t\right)\phi_A\left(x\right)\frac{-D_A \mathcal{L}_G\left(\Lambda_A\left(t\right)\phi_A\left(x\right), \Lambda_B\left(t\right)\phi_B\left(x\right)\right)}{\mathcal{L}_G\left(\Lambda_A\left(t\right)\phi_A\left(x\right), \Lambda_B\left(t\right)\phi_B\left(x\right)\right)}.$$

This is the rate of failure due to cause A at time t conditional on x and survival up to time t. The ratio in the right-hand side equals $\mathbb{E}\left[V_A|x, T_A > t, T_B > t\right]$. Suppose we know ϕ_A and ϕ_B. By Assumption 1 we can vary $\phi_B(x)$ for fixed $\phi_A(x)$ by appropriately varying x. First, suppose that V_A and V_B are independent, so that $\mathcal{L}_G$ factorizes in the Laplace transforms $\mathcal{L}_{G_A}$ and $\mathcal{L}_{G_B}$ of the marginal distributions G_A of V_A and G_B of V_B. Then,

$$\frac{-Q'_A(t|x)}{S(t,t|x)} = \lambda_A(t)\,\phi_A(x)\,\frac{-\mathcal{L}'_{G_A}\left(\Lambda_A(t)\,\phi_A(x)\right)}{\mathcal{L}_{G_A}\left(\Lambda_A(t)\,\phi_A(x)\right)} = \lambda_A(t)\,\phi_A(x)\,\mathbb{E}\left[V_A|x, T_A > t\right]$$

is clearly not affected by a change in $\phi_B(x)$ that leaves $\phi_A(x)$ unchanged. After all, $\mathbb{E}\left[V_A|x, T_A > t, T_B > t\right] = \mathbb{E}\left[V_A|x, T_A > t\right]$ only depends on x through $\phi_A(x)$. However, if V_A and V_B are dependent, $\mathbb{E}\left[V_A|x, T_A > t, T_B > t\right]$ generally depends on x through $\phi_B(x)$ as well and $-Q'_A(t|x)/S(t,t|x)$ changes. This is due to the well-known fact that V_B and $\phi_B(x)$ are dependent conditional on survival $T_B > t > 0$ even if V_B and x are independent unconditionally. So, conditional on $T_B > t$, $\phi_B(x)$ affects V_A indirectly through V_B. In conclusion, the variation in the crude hazard $-Q'_A(t|x)/S(t,t|x)$ with $\phi_B(x)$ for given $\phi_A(x)$ is informative on the dependence of V_A and V_B. An analogous argument holds for the crude hazard corresponding to cause B, $-Q'_B(t|x)/S(t,t|x)$.

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). Instead, suppose we observe two spells in a *stratum* that is characterized by a single realization of (V_A, V_B). The stratum could either correspond to a single physical unit, like an individual, for which we observe two spells in exactly the same state or consist of single spells corresponding to two similar physical units, for example twins. For each stratum, we observe two identified minima (T_1, I_1) and (T_2, I_2), with $T_k = \min_{i\in\{A,B\}} T_{i,k}$ and $I_k = \arg\min_{i\in\{A,B\}} T_{i,k}$ for some latent failure times $(T_{A,k}, T_{B,k})$, $k = 1, 2$. We first suppress regressors x. The main result does not rely on regressor variation and we can think of the analysis as being conditional on x. In particular, we allow (V_A, V_B) to be dependent on x (this was not allowed in Sections 2 and 3).

We assume that the pairs of latent failure times $(T_{A,1}, T_{B,1})$ and $(T_{A,2}, T_{B,2})$ are independent conditional on (V_A, V_B). In other words, multiple spells within a stratum are only dependent through the unobserved determinants. If we also again assume that the latent failure times are independent conditional on (V_A, V_B), the joint distribution of $(T_{A,1}, T_{B,1}, T_{A,2}, T_{B,2})|(V_A, V_B)$ factorizes in the marginal distributions of $T_{i,k}|(V_A, V_B)$, $i = A, B$ and $k = 1, 2$. In turn, these are characterized by the corresponding hazard rates $\theta_{i,k}(t|V_A, V_B) = \lambda_{i,k}(t)V_i$, where the baseline hazards $\lambda_{i,k} : \mathbb{R}_+ \to (0, \infty)$ have integrals

$$\Lambda_{i,k}(t) := \int_0^t \lambda_{i,k}(\tau)d\tau < \infty$$

for all $t \in \mathbb{R}_+$, $i = A, B$ and $k = 1, 2$. We normalize $\Lambda_{A,1}(t^*) = \Lambda_{B,1}(t^*) = 1$ for some a priori chosen $t^* \in (0, \infty)$. These normalizations are again innocuous because V_A and V_B can capture the scale of the first-spell hazards $\theta_{A,1}$ and $\theta_{B,1}$. The joint survival function of $(T_{A,1}, T_{B,1}, T_{A,2}, T_{B,2})|(V_A, V_B)$ easily follows as

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

Finally, taking expectations with respect to the unobservables (V_A, V_B) gives

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

It is intuitively clear that multiple-spell data facilitate identification. The analogies with linear panel-data models with fixed effects and the models for paired duration data of Holt and Prentice (1974) and Holt (1978) suggest that we can deal with unobserved heterogeneity in multiple-spell data by exploiting within-stratum variation. Indeed, we have the following result.

Proposition 3. *(i) The functions $\Lambda_{A,1}$, $\Lambda_{B,1}$, $\Lambda_{A,2}$, and $\Lambda_{B,2}$ are identified from the distribution of (T_1, I_1, T_2, I_2).*
(ii) $\mathcal{L}_G$ is identified if $\left\{(\Lambda_{A,1}(t_1) + \Lambda_{A,2}(t_2), \Lambda_{B,1}(t_1) + \Lambda_{B,2}(t_2))\, ; (t_1, t_2) \in \mathbb{R}^2_+\right\}$ contains a nonempty open set in $\mathbb{R}^2$.

Proof. (i) The distribution of (T_1, I_1, T_2, I_2) provides the probabilities of all (sub-)survival events like $(T_{A,1} > t_1, T_{B,1} > T_{A,1}, T_{A,2} > t_2, T_{B,2} > t_2)$, etcetera. So, analogously to (5) we can compute the sub-density

$$\begin{aligned}
&-\frac{\partial \Pr\left(T_{A,1} > t_1, T_{B,1} > T_{A,1}, T_{A,2} > t_2, T_{B,2} > t_2\right)}{\partial t_1} \\
&\qquad = -\lambda_{A,1}(t_1) D_A \mathcal{L}_G(\Lambda_{A,1}(t_1) + \Lambda_{A,2}(t_2), \Lambda_{B,1}(t_1) + \Lambda_{B,2}(t_2))
\end{aligned}$$

for almost all t_1 and all t_2 and the sub-density

$$
\begin{aligned}
-\frac{\partial \Pr\left(T_{A,1} > t_1, T_{B,1} > t_1, T_{A,2} > t_2, T_{B,2} > T_{A,2}\right)}{\partial t_2}& \\
&= -\lambda_{A,2}(t_2) D_A \mathcal{L}_G(\Lambda_{A,1}(t_1) + \Lambda_{A,2}(t_2), \Lambda_{B,1}(t_1) + \Lambda_{B,2}(t_2))
\end{aligned}
$$

for almost all t_2 and all t_1. With the normalization $\Lambda_{A,1}(t^*) = 1$, this implies that

$$
\Lambda_{A,2}(t) = \int_0^t \left[\int_0^{t^*} \frac{\partial \Pr\left(T_{A,1} > \tau_1, T_{B,1} > T_{A,1}, T_{A,2} > \tau_2, T_{B,2} > \tau_2\right)/\partial \tau_1}{\partial \Pr\left(T_{A,1} > \tau_1, T_{B,1} > \tau_1, T_{A,2} > \tau_2, T_{B,2} > T_{A,2}\right)/\partial \tau_2} d\tau_1 \right]^{-1} d\tau_2.
$$

Similar computations give

$$
\frac{\Lambda_{A,1}(t)}{\Lambda_{A,2}(t_2)} = \int_0^t \left[\int_0^{t_2} \frac{\partial \Pr\left(T_{A,1} > \tau_1, T_{B,1} > \tau_1, T_{A,2} > \tau_2, T_{B,2} > T_{A,2}\right)/\partial \tau_2}{\partial \Pr\left(T_{A,1} > \tau_1, T_{B,1} > T_{A,1}, T_{A,2} > \tau_2, T_{B,2} > \tau_2\right)/\partial \tau_1} d\tau_2 \right]^{-1} d\tau_1,
$$

which identifies $\Lambda_{A,1}$ for arbitrary $t_2 \in (0, \infty)$. $\Lambda_{B,1}$ and $\Lambda_{B,2}$ can be identified analogously.

(ii) The distribution of (T_1, I_1, T_2, I_2) provides data on $S(t_1, t_1, t_2, t_2)$ for $(t_1, t_2) \in \mathbb{R}_+^2$. By equation (9), $S(t_1, t_1, t_2, t_2) = \mathcal{L}_G(\Lambda_{A,1}(t_1) + \Lambda_{A,2}(t_2), \Lambda_{B,1}(t_1) + \Lambda_{B,2}(t_2))$. So, because $\Lambda_{A,1}$, $\Lambda_{B,1}$, $\Lambda_{A,2}$, and $\Lambda_{B,2}$ are identified by *(i)*, we can trace $\mathcal{L}_G$ on a nonempty open set. By Proposition 1, this identifies $\mathcal{L}_G$. □

This result does not require regressor variation. Rather, we implicitly allow for conditioning on regressors x. In particular, we can think of the baseline hazards as being specific to a particular value of x and thus allow for general interactions between elapsed duration t and x. A problem seems to be that the normalizations exclude variation of the first-spell baseline hazards with x at time t^*. However, this is again innocuous because we allow for dependence of (V_A, V_B) and x, so that V_A and V_B can capture the dependence on x of the first-spell hazards at t^*. Thus, the normalizations only matter if a physical interpretation is given to the actual frailty variables, which is usually not the case.

If the condition in Proposition 3*(ii)* is not satisfied, the identifiability of $\mathcal{L}_G$ is not guaranteed. For example, if all 4 latent durations are exponential and spells are identically distributed within strata, i.e. if $\Lambda_{i,k}(t) \equiv t$ for $i = A, B$ and $k = 1, 2$, then we can only trace $\mathcal{L}_G$ on a 45-degree line through the origin and *(ii)* breaks down. In this case, we can resort to regressor variation. Suppose that we again have a vector of regressors x that is independent of (V_A, V_B). We specify $\theta_{i,k}(t|x, V_A, V_B) = \pi_{i,k}(t|x)V_i$, where the functions $\pi_{i,k}(\cdot|x) : \mathbb{R}_+ \to (0, \infty)$ have integrals

$$
\Pi_{i,k}(t|x) := \int_0^t \pi_{i,k}(\tau|x) d\tau < \infty
$$

for all $t \in \mathbb{R}_+$ and for given $x \in \mathcal{X}$, $i = A, B$ and $k = 1, 2$. We assume that $\Pi_{i,k}$ is continuous on $\mathbb{R}_+ \times \mathcal{X}$, $i = A, B$ and $k = 1, 2$, and normalize $\Pi_{A,1}(t^*|x^*) = \Pi_{B,1}(t^*|x^*) = 1$ for some a priori chosen $t^* \in (0, \infty)$ and $x^* \in \mathcal{X}$. Note that $\Pi_{A,1}(t^*|x)$ and $\Pi_{B,1}(t^*|x)$ are allowed to vary with x. We simply extend the condition in Proposition 3*(ii)* into

Assumption 3. (Variation with observed regressors)
$\left\{(\Pi_{A,1}(t_1|x) + \Pi_{A,2}(t_2|x), \Pi_{B,1}(t_1|x) + \Pi_{B,2}(t_2|x))\,; (t_1, t_2, x) \in \mathbb{R}_+^2 \times \mathcal{X}\right\}$ contains a nonempty open set in $\mathbb{R}^2$.

A sufficient condition for Assumption 3 is that $\pi_{i,k}$ is proportional in a baseline hazard $\lambda_{i,k}$ and a regressor function $\phi_{i,k}$ as in the single-spell case, $i = A, B$ and $k = 1, 2$, and that $(\phi_{A,1} + \phi_{A,2}, \phi_{B,1} + \phi_{B,2})$ satisfies Assumption 1. We have

Proposition 4. *If Assumptions 2 and 3 are satisfied, then the multiple-spell MPH competing risks model (which is characterized by the functions* $\Pi_{A,1}$, $\Pi_{B,1}$, $\Pi_{A,2}$, $\Pi_{B,2}$, *and* $\mathcal{L}_G$*) is identified from the distribution of* $(T_1, I_1, T_2, I_2)|x$.

Proof. For given $x \in \mathcal{X}$, the model fits the framework of Proposition 3 with baselines $\Lambda_{i,k}$ and unobserved factors $\tilde{V}_i$ such that $\Lambda_{i,k}(t) = \Pi_{i,k}(t|x)/\Pi_{i,1}(t^*|x)$ for all $t \in \mathbb{R}_+$ and $\tilde{V}_i = \Pi_{i,1}(t^*|x)V_i$, $i = A, B$ and $k = 1, 2$. So, $\Pi_{i,k}(\cdot|x)/\Pi_{i,1}(t^*|x)$, $i = A, B$, is identified by Proposition 3. Identification of $\Pi_{i,1}(t^*|\cdot)$, $i = A, B$, and $\mathcal{L}_G$ follows from the obvious multiple-spell equivalent to the first two steps of the proof of Proposition 2. □

The results in this section are akin to the multiple-spell results of Honoré (1993) for single-risk models. However, the competing-risks nature of the data complicates matters. If the condition in Proposition 3*(ii)* is not satisfied, 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 regressors.

5 Conclusion

The main result of this paper is that the dependent MPH competing risks model with regressors is identified under milder conditions than those in Heckman and Honoré (1989). In particular, Heckman and Honoré (1989) assume a large support, $(0, \infty)^2$, of the proportional regressor effects on the hazards of the latent failure times. We only need that this support includes a nonempty open set in $\mathbb{R}^2$. In applications, the latter condition is much more likely to be satisfied. We extend the results to the multiple-spells case.

In applied work, the interest is sometimes restricted to the covariate effects (although in econometrics the duration dependence functions are often of independent interest; see e.g. Machin and Manning, 1999). In this context it is interesting to note that the identification proofs for the single-spell case in this paper and in Heckman and Honoré (1989) are constructive, i.e., they express the model determinants in terms of observables. This suggests an estimation method for covariate effects based on these expressions. From the proofs it is clear that the support of the covariates x does not play a role in such an estimation method; it can be applied even when x assumes only two values. However, it is also clear that such an estimation method is very unattractive from a practical point of view, since it is only based on observations with durations close to zero. It may be more promising to explore the following estimation approach: first, estimate all model determinants along the lines of our constructive identification proof, and, second, re-estimate the covariate effects using data on all durations, exploiting the estimates of the first step. The study of the properties of such estimators may be a fruitful topic for further research.

For the multiple-spell case, the stratified partial likelihood estimator provides estimates of proportional covariate effects under weak identifying conditions (Holt and Prentice, 1974; Holt, 1978; see Van den Berg, 2001, for discussion). We provide constructive identification proofs for a multiple-spell competing risks model that allows for general interactions of covariate effects and durations, and for different covariate-duration effects between spells in a stratum. As in the single-spell case, these may be used for the development of estimators.

References

Abbring, J.H. (2002), "Stayers versus defecting movers: A note on the identification of defective duration models", *Economics Letters*, 74, 327–331.

Carathéodory, C. (1918), *Vorlesungen über Reelle Funktionen*, Teubner, Leipzig.

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

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

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

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.

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

Holt, J.D. (1978), "Competing risk analyses with special reference to matched pair experiments", *Biometrika*, 65, 159–165.

Holt, J.D. and R.L. Prentice (1974), "Survival analysis in twin studies and matched pair experiments", *Biometrika*, 61, 17–30.

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

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.

Machin, S. and A. Manning (1999), "The causes and consequences of long-term unemployment in Europe", in: O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics, Volume III*, North-Holland, Amsterdam.

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

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

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 J.J. Heckman and E. Leamer, editors, *Handbook of Econometrics, Volume V*, North Holland, Amsterdam.

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.

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.