Econometric Theory, **29**, 2013, 905–919.
doi:10.1017/S0266466612000795

NONPARAMETRIC IDENTIFICATION OF ACCELERATED FAILURE TIME COMPETING RISKS MODELS

SOKBAE LEE
Seoul National University

ARTHUR LEWBEL
Boston College

We provide new conditions for identification of accelerated failure time competing risks models. These include Roy models and some auction models. In our setup, unknown regression functions and the joint survivor function of latent disturbance terms are all nonparametric. We show that this model is identified given covariates that are independent of latent errors, provided that a certain rank condition is satisfied. We present a simple example in which our rank condition for identification is verified. Our identification strategy does not depend on identification at infinity or near zero, and it does not require exclusion assumptions. Given our identification, we show estimation can be accomplished using sieves.

1. INTRODUCTION

Suppose that there are K competing causes of failure indexed by the integers 1 to K with corresponding logarithms $(T_1, \ldots, T_K)$ of latent failure times. One observes the log duration time to the first failure and the corresponding cause of failure, denoted by $Y = \min_k T_k$ and $D = \arg\min_k T_k$, along with a covariate vector $\mathbf{X}$.

It is well known (Cox, 1962; Tsiatis, 1975) that without some restrictions on this model, the joint distribution of latent failure times is nonparametrically unidentified. Furthermore, the bounds for the latent distribution implied by the observed competing risks are, in general, wide (Peterson, 1976). This negative identification result can be mitigated given independent variation from observed covariates along with some modeling assumptions.

We would like to thank Xiaohong Chen for very helpful discussions at the early stage of this project. We also would like to thank Hidehiko Ichimura, Ivana Komunjer, Rosa Matzkin, Whitney Newey, Jean-Marc Robin, Elie Tamer, and two anonymous referees for helpful comments. This work was supported by the Economic and Social Research Council for the ESRC Centre for Microdata Methods and Practice (RES-589-28-0001) and a small research grant (RES-000-22-2761), by a European Research Council grant (ERC-2009-StG-240910-ROMETA), and by the National Research Foundation of Korea grant funded by the Korean Government (NRF-2011-327-B00073). Address correspondence to Sokbae Lee, Department of Economics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul, 151-742, Republic of Korea; e-mail: sokbae@snu.ac.kr.

© Cambridge University Press 2013

https://doi.org/10.1017/S0266466612000795 Published online by Cambridge University Press

In this paper we show generic identification of nonparametric accelerated failure time competing risks models. Assume that

$$T_k = g_k(\mathbf{X}) + U_k, \tag{1}$$

where each function g_k is unknown, U_k is an unobserved error, and the vector $(U_1, \ldots, U_K)$ is continuously distributed with an unknown joint distribution. The purpose of this paper is to provide conditions under which the functions $g_1, \ldots, g_K$ and the joint survivor function (and hence, also the joint distribution function) of $U_1, \ldots, U_K$ are nonparametrically identified.

The key assumption we make for identification is that $(U_1, \ldots, U_K)$ is independent of $\mathbf{X}$. This conditional independence assumption (that is, errors independent of covariates) is common in empirical practice and is used extensively as a way of achieving identification in a variety of econometric models. See, e.g., Matzkin (2007) for a recent survey. The only other assumptions we require for nonparametric identification are a rank condition and some regularity. We present a simple example in which our conditions for identification are verified. Given our identification, we show that sieve methods can be used for estimation.

There exists a large literature on identification in competing risks models. For example, Heckman and Honoré (1989) show that competing risks versions of mixed proportional hazards and accelerated failure time models can be identified with covariates. Heckman and Honoré (1990) provide identification of the corresponding Roy model. Abbring and Van den Berg (2003) provide weaker conditions than those assumed in Heckman and Honoré (1989) for the mixed proportional hazards competing risks model. Lee (2006) develops an identification result for a competing risks transformation model. Buera (2006) develops nonparametric identification and testable implications of the Roy model using assumptions that are distinct from ours, such as the continuity of regressors. Khan and Tamer (2009) consider a semiparametric version of (1) with a parametric g_k with $K = 2$.

Han and Hausman (1990) provide a flexible parametric estimation method for competing risks models with regressors. Zheng and Klein (1995) obtain an identification result for the latent marginal distributions when the copula between two latent failure times is known. Their identification result can be used to construct a bound if a class of copula functions is restricted. Bond and Shaw (2006) obtain bounds for covariate effects under the assumption that the copula associated with the joint distribution of latent failure times is invariant to the value of covariates. Abbring and Van den Berg (2005) apply the result of Bond and Shaw to bound the treatment effects on duration outcomes. Honoré and Lleras-Muney (2006) derive bounds in an accelerated failure time competing risks model with discrete covariates.

Our new identification result is most closely related to that of Heckman and Honoré (1989, 1990) in terms of the general class of models we identify. The identification results of Heckman and Honoré (1989, 1990) are based on the argument

https://doi.org/10.1017/S0266466612000795 Published online by Cambridge University Press

of letting the index of the duration variable go to zero, thereby implying that corresponding estimation methods would be based only on observations with failure times close to zero. An estimator of Femanian (2003, Sec. 4) is an example. This "identification at (in the neighborhood of) zero" raises the same difficulties as those associated with the more well-known "identification at infinity" problem (see, e.g., Heckman, 1990; Andrews and Schafgans, 1998) and is an example of what Khan and Tamer (2010) call irregular or thin set identification. In contrast, our identification is regular and is not based on identification at infinity or in the neighborhood of zero or other thin set identification.

Our identification also does not depend on exclusion restrictions; that is, we permit all of the elements of $\mathbf{X}$ to appear in all of the $\mathbf{g}(\mathbf{x})$ functions. Exclusion restrictions are usually not imposed in competing risks models (since they are typically difficult to justify in that context) but are commonly used for identification in other types of multiple index models. See, e.g., Ichimura and Lee (1991).

The identification result of Lee (2006) also does not depend on either identification near zero or exclusion assumption. Lee shows that a parametrically specified $\mathbf{g}(\mathbf{x})$ can be identified up to scale and location normalization for a class of transformation models that include accelerated failure time competing risks models as special cases. However, his result applies only to a parametrically specified $\mathbf{g}(\mathbf{x})$ with all elements of $\mathbf{x}$ being continuously distributed. We do not require that all elements of $\mathbf{x}$ be continuously distributed and allow for discrete regressors as well. As far as we know, our theorem is the first identification result in the literature that obtains nonparametric identification of $\mathbf{g}(\mathbf{x})$ in (1) without exclusion assumptions and without identification at zero, while allowing for both continuous and discrete regressors.

Our identification result can be translated directly into identification of a Roy model in which one observes only $Y = \max_k T_k$ and $D = \arg\max_k T_k$ (Heckman and Honoré, 1990). Fox and Gandhi (2009) develop a different identification result for Roy models without the problem of identification at infinity. However, their model and our model in (1) are nonnested, and the two identification results are not directly comparable. Bayer, Khan, and Timmins (2011) consider nonparametric identification and estimation of a generalized Roy model without covariates using conditions that are not comparable to ours. French and Taber (2011) provide a review of the Roy model in the context of identification of common selection models of the labor market. Our result also encompasses random censoring, since random censoring models are equivalent to competing risks models with $K = 2$ (see, e.g., Khan and Tamer, 2009).

Our model is also related to some auction models considered in, e.g., Athey and Haile (2002). For example, when one observes both the highest bid in a second-price auction and the identity of the auction winner, then the auction model and observable data are equivalent to our competing risks setup, and so our identification result then provides an alternative to those already existing in the auction literature.

https://doi.org/10.1017/S0266466612000795 Published online by Cambridge University Press

The remainder of the paper is organized as follows. Section 2 gives conditions under which the functions $g_1, \ldots, g_K$ and the joint survivor function of $U_1, \ldots, U_K$ are identified. In Section 3 we consider sieve maximum likelihood estimation of (1) using a random sample of $(Y, D, \mathbf{X})$. Section 4 presents a simple example for which we verify identification conditions. All the proofs that are not given in the main text are in the Appendix.

2. THE IDENTIFICATION RESULT

This section presents the identification result. First we note that for each k, $g_k(\mathbf{x})$ is not identified without location normalization, since a shift in $g_k(\mathbf{x})$ can be offset by a shift in U_k. We therefore fix location by assuming that there exists a point $\mathbf{x}_0$ such that $g_k(\mathbf{x}_0) = 0$ for each $k = 1, \ldots, K$.

Let $P(u_1, \ldots, u_K)$ denote the joint survivor function of $U_1, \ldots, U_K$, so

$$P(u_1, \ldots, u_K) = \Pr(U_1 > u_1, \ldots, U_K > u_K).$$

Let $\mathbf{U}$ be the K vector of elements U_k for $k = 1, \ldots, K$. Define

$$B_k(s \mid \mathbf{x}) := \mathbb{E}[I(Y > s, D = k) \mid \mathbf{X} = \mathbf{x}] \quad \text{and}$$

$$C(s \mid \mathbf{x}) := \int_s^\infty \mathbb{E}[I(Y > t) \mid \mathbf{X} = \mathbf{x}]\, dt.$$

We may equivalently write $C(s \mid \mathbf{x})$ as $C(s \mid \mathbf{x}) = \mathbb{E}[(Y - s)I(Y > s) \mid \mathbf{X} = \mathbf{x}]$. Note that $C(s \mid \mathbf{x})$ exists for any value of $s \in \mathbb{R}$ if $\mathbb{E}[Y \mid \mathbf{X} = \mathbf{x}]$ exists for each $\mathbf{x}$. Let $\mathbf{g}(\mathbf{x})$ be the K vector of elements $g_k(\mathbf{x})$ for $k = 1, \ldots, K$. By construction, the functions $B_k(s \mid \mathbf{x})$, $k = 1, \ldots, K$, and $C(s \mid \mathbf{x})$ can be identified directly from the conditional distribution of Y and D given $\mathbf{X}$.

In what follows, we use the notation $\mathbf{g}$ to denote the unknown true function and use $\mathbf{f}$ or $\mathbf{h}$ to denote a generic element of a function space that includes $\mathbf{g}$ as an element.

The key to our identification result is that we define a function C^* such that the equations $C^*(s, \mathbf{h}) = C(s \mid \mathbf{x})$ and $\partial C^*(s, h_1, \ldots, h_K)/\partial h_k = B_k(s \mid x)$ for $k = 1, \ldots, K$ hold when evaluated at $\mathbf{h} = \mathbf{g}(\mathbf{x})$. Identification of the true $\mathbf{g}(\mathbf{x})$ then follows given the technical conditions required to apply a global implicit function theorem, which ensures that these equations expressing C^* and its derivatives in terms of observable functions C and $B_1, \ldots, B_K$ have a unique solution.

Let $\|\mathbf{f}\|$ denote the L_2 norm of a K-dimensional vector-valued function $\mathbf{f} = (f_1, \ldots, f_K)$. That is, $\|\mathbf{f}\|_{L_2} = \left(\sum_{k=1}^K \int |f_k(\mathbf{x})|^2 d\mathbf{x}\right)^{1/2}$. Let $L_2(\mathcal{X})$ denote the L_2 space for K-dimensional vector-valued functions defined on $\mathcal{X}$, where $\mathcal{X}$ is the support of $\mathbf{X}$. Also, let $\mathcal{G} \subset L_2(\mathcal{X})$ denote the set of possible vector-valued functions that the true $\mathbf{g}$ might be.

For each $\mathbf{h} \equiv (h_1, \ldots, h_K) \in \mathcal{G}$, define

$$C^*[s, \mathbf{h}(\mathbf{x})] := \int_s^\infty P[t - h_1(\mathbf{x}), \ldots, t - h_K(\mathbf{x})]\, dt.$$

https://doi.org/10.1017/S0266466612000795 Published online by Cambridge University Press

Then it is straightforward to show that $C(s \mid \mathbf{x}) = C^*[s, \mathbf{g}(\mathbf{x})]$ for the true function $\mathbf{g}$. Therefore, the identification problem in this paper is to obtain conditions that rule out all solutions $\mathbf{h}(\mathbf{x})$ to $C(s \mid \mathbf{x}) = C^*[s, \mathbf{h}(\mathbf{x})]$ other than the true solution $\mathbf{g}(\mathbf{x})$. To do this, it is necessary at a minimum to consider K different values of s, since $\mathbf{g}$ is an $\mathbb{R}^K$-valued function. Choose K values $s_1, \ldots, s_K$ such that $\int \left\{C^*\left[s_j, \mathbf{h}(\mathbf{x})\right]\right\}^2 d\mathbf{x} < \infty$ for each $j = 1, \ldots, K$ and for any $\mathbf{h} \in \mathcal{G}$. Now define a map $\mathbf{C}^* : \mathcal{G} \mapsto L_2(\mathcal{X})$, where, for any $\mathbf{h} \in \mathcal{G}$, the j element of $\mathbf{C}^*$ satisfies

$$\mathbf{C}^*_j[\mathbf{h}(\mathbf{x})] = C^*\left[s_j, \mathbf{h}(\mathbf{x})\right].$$

The identification problem then reduces to finding conditions that permit inverting $\mathbf{C}^*(\mathbf{h})$ globally as a function of $\mathbf{h}$. Let $\mathbf{C}(\mathbf{x})$ and $\mathbf{C}^*[\mathbf{g}(\mathbf{x})]$ be the K vectors of elements $C(s_j \mid \mathbf{x})$ and $C^*\left[s_j, \mathbf{g}(\mathbf{x})\right]$ for $j = 1, \ldots, K$.

Let $\mathbb{L}[\mathcal{G}, L_2(\mathcal{X})]$ denote the class of all linear and continuous maps from $\mathcal{G}$ to $L_2(\mathcal{X})$. Let $< \cdot, \cdot >$ denote the inner product of $L_2(\mathcal{X})$; that is,

$$< \mathbf{f}, \mathbf{h} > = \left(\sum_{k=1}^{K} \int f_k(\mathbf{x}) h_k(\mathbf{x}) d\mathbf{x} \right),$$

where $\mathbf{f} = (f_1, \ldots, f_K)$ and $\mathbf{h} = (h_1, \ldots, h_K)$. Define a map $\dot{\mathbf{C}}^* : \mathcal{G} \mapsto \mathbb{L}[\mathcal{G}, L_2(\mathcal{X})]$, where the (j,k) element of $\dot{\mathbf{C}}^*$ satisfies

$$\dot{\mathbf{C}}^*_{jk}[\mathbf{h}(\mathbf{x})] = \frac{\partial C^*\left[s_j, h_1(\mathbf{x}), \ldots, h_K(\mathbf{x})\right]}{\partial h_k(\mathbf{x})}$$

$$= -\int_{s_j}^{\infty} P_k[t - h_1(\mathbf{x}), \ldots, t - h_K(\mathbf{x})] dt,$$

where $P_k(u_1, \ldots, u_K) = \partial P(u_1, \ldots, u_K)/\partial u_k$. Note that $\dot{\mathbf{C}}^*$ is the Fréchet-derivative of $\mathbf{C}^*$, passing the derivative through the integral sign in the definition of $\mathbf{C}^*$.

To obtain conditions for identification, we first make the following assumptions.

Assumption 1. Assume that (i) each element of $\mathbf{g}(\mathbf{X})$ is continuously distributed; (ii) $\mathbb{E}[Y \mid \mathbf{X} = \mathbf{x}]$ exists for each $\mathbf{x}$ in the support of $\mathbf{X}$; and (iii) $P(u_1, \ldots, u_K)$ is everywhere continuously differentiable with respect to u_k for all $k = 1, \ldots, K$.

Condition (i) allows the covariate vector $\mathbf{X}$ to contain discrete elements but requires it to have some continuously distributed elements, e.g., it would generally hold if $\mathbf{g}(\mathbf{X})$ is linear with some but not all elements of $\mathbf{X}$ being continuous. Condition (ii) is a weak restriction that ensures that $C(s \mid \mathbf{x})$ is well defined. Condition (iii) imposes some mild smoothness on the underlying distribution function of $\mathbf{U}$.

For each $\mathbf{x}$, let $\mathbf{B}(\mathbf{x})$ denote the $K \times K$ matrix whose (j,k) element is $B_k(s_j|\mathbf{x})$. The following theorem is useful to obtain our identification result.

https://doi.org/10.1017/S0266466612000795 Published online by Cambridge University Press

THEOREM 1. *Let Assumption 1 hold. Then*

$$\left.\frac{\partial C^*[s, h_1(\mathbf{x}), \ldots, h_K(\mathbf{x})]}{\partial h_k(\mathbf{x})}\right|_{\mathbf{h}=\mathbf{g}} = B_k(s \mid \mathbf{x}), \tag{2}$$

equivalently, $\dot{\mathbf{C}}^*(\mathbf{g}) = \mathbf{B}(\mathbf{x})$ *for each* $\mathbf{x}$.

Note that the right-hand side of equation (2) is directly identified from the data, and the left-hand side of equation (2) is a function of $\mathbf{g}(\mathbf{x})$. We now consider identification based on Theorem 1. To do so, we make further assumptions.

Assumption 2. Assume there exists a set of K constants $s_1, \ldots, s_K$ used to define $\mathbf{C}^*$ such that the following conditions hold: (i) $s_j - \mathbf{h}(\mathbf{x}) \in supp(\mathbf{U})$ and $\mathbf{C}^*(\mathbf{h}) \in L_2(\mathcal{X})$ except, possibly, if $\mathbf{x}$ belongs to a set of Lebesgue measure 0; (ii) as a matrix, for each $\mathbf{h} \in \mathcal{G}$, $\dot{\mathbf{C}}^*[\mathbf{h}(\mathbf{x})]$ is invertible except, possibly, if $\mathbf{x}$ belongs to a set of Lebesgue measure 0; (iii) for any $\mathbf{h}_1 \equiv (h_{11}, \ldots, h_{1K}) \in \mathcal{G}$ and $\mathbf{h}_2 \equiv (h_{21}, \ldots, h_{2K}) \in \mathcal{G}$ such that $\mathbf{h}_1 \neq \mathbf{h}_2$, there exist universal constants $\delta > 0$ and $\epsilon > 0$ such that

$$\left\langle \mathbf{C}^*(\mathbf{h}_1) - \mathbf{C}^*(\mathbf{h}_2), \mathbf{h}_1 - \mathbf{h}_2 \right\rangle \geq \delta \|\mathbf{h}_1 - \mathbf{h}_2\|_{L_2}^{1+\epsilon}. \tag{3}$$

The following theorem shows that under Assumption 2, $\mathbf{C}^*(\mathbf{h})$ is globally invertible as a function of $\mathbf{h}$.

THEOREM 2. *Let Assumption 2 hold. Then* $\mathbf{C}^*$ *is invertible, and its inverse map, say* $\mathbf{H}$*, is continuously Fréchet-differentiable. That is,* $\mathbf{g}(\mathbf{x}) = \mathbf{H}[\mathbf{C}(\mathbf{x})]$ *and its Fréchet-derivative, say* $\mathbf{H}'$*, satisfies*

$$\mathbf{H}'[\mathbf{C}(\mathbf{x})] = \left(\dot{\mathbf{C}}^*\{\mathbf{H}[\mathbf{C}(\mathbf{x})]\}\right)^{-1}. \tag{4}$$

Condition (i) of Assumption 2 is a relatively weak assumption, requiring $\mathbf{U}$ to have a nonzero density over a large region. Conditions (ii) and (iii) together comprise the key rank condition that ensures identification of the model. Condition (ii) assumes that the columns of $\dot{\mathbf{C}}^*[\mathbf{h}(\mathbf{x})]$ are linearly independent for each $\mathbf{h} \in \mathcal{G}$. This condition is needed to guarantee the local identification of $\mathbf{g}$. Condition (iii) implies that $\mathbf{C}^*$ is a proper mapping (i.e., the inverse image of a compact set is compact), thereby ensuring that the local identification result extends to global identification. Komunjer (2012) adopts a similar strategy to obtain identification of a class of parametric nonlinear models with a finite number of unconditional moment restrictions. Note that these rank conditions generally require that $\mathbf{X}$ contain at least K continuously distributed elements and also that no one element of $\mathbf{g}(\mathbf{x})$ can be expressed as a function of the other elements of $\mathbf{g}(\mathbf{x})$.

Given the smoothness assumption on $P(u_1, \ldots, u_K)$ from Theorem 1, condition (iii) of Assumption 2 is equivalent to assuming that there exist universal

https://doi.org/10.1017/S0266466612000795 Published online by Cambridge University Press

constants $\delta > 0$ and $\epsilon > 0$ such that

$$\sum_{j=1}^{K}\sum_{k=1}^{K}\int \dot{\mathbf{C}}^{*}_{jk}\left[\tilde{\mathbf{h}}_j(\mathbf{x})\right]\left[h_{1j}(\mathbf{x})-h_{2j}(\mathbf{x})\right]\left[h_{1k}(\mathbf{x})-h_{2k}(\mathbf{x})\right]d\mathbf{x}$$
$$\geq \delta\left\{\sum_{j=1}^{K}\int\left[h_{1j}(\mathbf{x})-h_{2j}(\mathbf{x})\right]^2 d\mathbf{x}\right\}^{(1+\epsilon)/2}, \tag{5}$$

where $\tilde{\mathbf{h}}_j(\mathbf{x})$ is between $\mathbf{h}_1(\mathbf{x})$ and $\mathbf{h}_2(\mathbf{x})$ for each $j = 1, \ldots, K$. Note that the mean value $\tilde{\mathbf{h}}_j(\mathbf{x})$ could be different across j's, since the mean-value theorem is applied componentwise. Abusing notation a bit, for any $(\mathbf{h}_1, \ldots, \mathbf{h}_K) \in \mathcal{G}^K$, let $\dot{C}^*(\mathbf{h}_1, \ldots, \mathbf{h}_K)$ denote the $K \times K$ matrix whose (j,k) element is $\dot{\mathbf{C}}^{*}_{jk}[\mathbf{h}_j(\mathbf{x})]$. A sufficient condition for (5) to hold with $\epsilon = 1$ is the following.

Assumption 3. There exists a universal constant $\delta > 0$ such that $\dot{\mathbf{C}}^*(\mathbf{h}_1, \ldots, \mathbf{h}_K) - \delta I_K$ is a positive-semidefinite matrix for any $(\mathbf{h}_1, \ldots, \mathbf{h}_K) \in \mathcal{G}^K$, where I_K is the K-dimensional identity matrix.

Note that this sufficient condition also implies condition (ii) of Assumption 2, since $\dot{\mathbf{C}}^*(\mathbf{h})$ is a positive-definite matrix for any $\mathbf{h} \in \mathcal{G}$ under Assumption 3. A simple example in Section 4 illustrates a case when conditions (ii) and (iii) of Assumption 2 and also Assumption 3 are satisfied.

Theorems 1 and 2 can be combined to obtain the identification of $\mathbf{g}$. The following theorem provides the main result of this paper. Since its proof is constructive, we present the proof of this theorem in the main text.

THEOREM 3. *Let Assumptions 1 and 2 hold. It then follows from identification of the conditional distribution of Y and D given $\mathbf{X}$ that the vector-valued function $\mathbf{g}(\mathbf{x})$ is identified for all $\mathbf{x} \in supp\,(\mathbf{X})$.*

Proof of Theorem 3. The functions B_k for $k = 1, \ldots, K$ and C are identified from the conditional distribution of Y and D given $\mathbf{X}$. It suffices to show that Theorems 1 and 2 provide a construction of $\mathbf{g}(\mathbf{x})$ given the functions B_k and C, so $\mathbf{g}(\mathbf{x})$ is identified.

To do this, first note that the Fréchet-derivative in (4) is equivalent to the usual derivative, since the right-hand side of (4) is not an integral equation. Define $b_k(s \mid \mathbf{C}(\mathbf{x})) := \mathbb{E}[I(Y > s, D = k) \mid \mathbf{C}(\mathbf{x})]$. Since by Theorem 2, $\mathbf{g}(\mathbf{x}) = \mathbf{H}[\mathbf{C}(\mathbf{x})]$ and $\mathbf{H}$ is one-to-one, we have that

$$\mathbb{E}[I(Y > s, D = k) \mid \mathbf{C}(\mathbf{x})] = \mathbb{E}[I(Y > s, D = k) \mid \mathbf{g}(\mathbf{x})] \tag{6}$$

for each s and $\mathbf{x}$. Also, note that by construction, (Y, D) is a function only of $\mathbf{g}(\mathbf{X})$ and $\mathbf{U}$, and by assumption $\mathbf{U} \perp \mathbf{X}$. Therefore, the distribution of (Y, D) given $\mathbf{X} = \mathbf{x}$ equals the distribution of (Y, D) given $\mathbf{g}(\mathbf{X}) = \mathbf{g}(\mathbf{x})$. This implies that

$$\mathbb{E}[I(Y > s, D = k) \mid \mathbf{g}(\mathbf{x})] = \mathbb{E}[I(Y > s, D = k) \mid \mathbf{x}]. \tag{7}$$

https://doi.org/10.1017/S0266466612000795 Published online by Cambridge University Press

Hence, it follows from (6) and (7) that we have that

$b_k[s \mid \mathbf{C}(\mathbf{x})] = B_k(s \mid \mathbf{x})$.

Let $\mathbf{b}[\mathbf{C}(\mathbf{x})]$ be the matrix of elements $b_k\left[s_j \mid \mathbf{C}(\mathbf{x})\right]$ for $j = 1, \ldots, K$ and $k = 1, \ldots, K$ with previously chosen $s_1, \ldots, s_K$. Then equations (2) and (4) mean that

$$\left.\frac{d\mathbf{H}(\mathbf{c})}{d\mathbf{c}'}\right|_{\mathbf{c}=\mathbf{C}(\mathbf{x})} = \mathbf{b}(\mathbf{c})^{-1}\Big|_{\mathbf{c}=\mathbf{C}(\mathbf{x})},$$

where $d\mathbf{H}(\mathbf{c})/d\mathbf{c}'$ is the Jacobian matrix. Integrating this expression to obtain the function $\mathbf{H}$ then gives $\mathbf{g}(\mathbf{x}) = \mathbf{H}[\mathbf{C}(\mathbf{x})]$ (using the location normalization $\mathbf{H}[\mathbf{C}(\mathbf{x}_0)] = 0$). More specifically, let $\mathbf{L}$ be a K-dimensional smooth curve connecting $\mathbf{C}(\mathbf{x})$ to $\mathbf{C}(\mathbf{x}_0)$. Then the line integral of the kth row of $d\mathbf{H}(\mathbf{c})/d\mathbf{c}'$ (viewed as the gradient of $g_k(\mathbf{x})$) along $\mathbf{L}$ gives $g_k(\mathbf{x})$ by the fundamental theorem for line integrals. This then allows us to solve for $\mathbf{g}(\mathbf{x})$. ■

Note that (7) holds without $\mathbf{g}$ being necessarily one-to-one. This is because by construction, (Y, D) is a function only of $\mathbf{g}(\mathbf{X})$ and $\mathbf{U}$, and by assumption $\mathbf{U} \perp \mathbf{X}$, as explained in the proof.

It can be seen from the identity between $C(t|\mathbf{x})$ and $C^*[s, \mathbf{g}(\mathbf{x})]$ that the model in (1) belongs to the class of multiple-index models (see, for example, Ichimura and Lee, 1991). Often some exclusion restrictions (for example, having some components of parameter vectors be zero) are needed for multiple-index models to achieve identification of parameters. As is common in the literature on identification in competing risks models, exclusion restrictions are not required for the identification of (1), though if present, exclusions could be useful for verifying the required rank condition.

As a corollary, we also obtain the following identification result for the joint survivor distribution of $\mathbf{U}$.

THEOREM 4. *Let Assumptions 1 and 2 hold. Then the joint survivor function of* $\mathbf{U}$ *is identified for every value of* $\mathbf{U}$ *in the support of* $Y - \mathbf{g}(\mathbf{x})$.

3. SIEVE MAXIMUM LIKELIHOOD ESTIMATION

In this section we consider estimation of (1) using a random sample $\{(Y_i, D_i, \mathbf{X}_i) : i = 1, \ldots, n\}$ of $(Y, D, \mathbf{X})$. Our identification results are constructive, so an analog estimator would be possible. For example, estimation could be based on minimizing a quadratic form of the equations in Theorem 2, using estimates of the functions B_k and $C = C^*$. However, this would involve either a somewhat arbitrary choice of values of $s_1, \ldots, s_K$, or minimizing over a set of such chosen values. We instead propose a more standard sieve-type estimator.

To do so, define $q_j(y|\mathbf{x})$ to be a subdensity associated with the jth cause. That is,

$$q_j(y|\mathbf{x}) := -\left.\frac{\partial B_j(s|\mathbf{x})}{\partial s}\right|_{s=y}.$$

https://doi.org/10.1017/S0266466612000795 Published online by Cambridge University Press

Note that

$$q_j(y|\mathbf{x}) = \underbrace{\int_{y-g_1(\mathbf{x})}^{\infty} \cdots \int_{y-g_K(\mathbf{x})}^{\infty}}_{K-1 \text{ integrals excluding } j} f_{\mathbf{U}}(u_1, \ldots, u_{j-1}, y - g_j(\mathbf{x}), u_{j+1}, \ldots, u_K) \underbrace{du_1 \cdots du_K}_{du_j \text{ is excluded}},$$

where $f_{\mathbf{U}}$ is the joint density function of $\mathbf{U}$. Let $\theta_0 := (\mathbf{g}, f_{\mathbf{U}})$ denote unknown, true infinite-dimensional parameters in the model. Then the log likelihood function has the form

$$Q_n(\theta) := n^{-1} \sum_{i=1}^{n} \sum_{j=1}^{K} 1(D_i = j) \log q_j(Y_i|\mathbf{X}_i, \theta),$$

where θ denotes generic elements in the parameter space, and the dependence of q_j on θ is now explicit. The corresponding sieve maximum likelihood estimator (MLE), say $\hat{\theta}_n$, can then be defined as

$$\hat{\theta}_n := \text{argmax}_{\theta \in \Theta_n} Q_n(\theta),$$

where Θ_n is a sieve space that depends on n and approximates the parameter space Θ as $n \to \infty$.

It is rather straightforward to obtain consistency for this sieve MLE using general results available in the literature. See, for example, Ai and Chen (2003), Chen (2007), Chen and Shen (1998), Chernozhukov, Imbens, and Newey (2007), Gallant and Nychka (1987), Newey and Powell (2003), and Shen (1997), among many others. One weakness with the sieve approach is that the identification condition becomes implicit. For example, to prove the consistency of the sieve MLE, we will need to assume that

$$\sum_{j=1}^{K} 1(d = j) \log q_j(y|\mathbf{x}, \theta) \neq \sum_{j=1}^{K} 1(d = j) \log q_j(y|\mathbf{x}, \theta_0) \tag{8}$$

for any $\theta \neq \theta_0$. This is a reasonable assumption given our identification result, but it is a high-level condition relative to Assumptions 1 and 2.

We now formally state assumptions that suffice for consistency of our sieve MLE. The following result is just a slight modification of the general consistency theorem of Newey and Powell (2003, Thm. 4.1), adapted to our specific application. Let $\|\cdot\|$ denote a general norm and Θ a parameter space that is compact with respect to $\|\cdot\|$.

THEOREM 5 (Newey and Powell, 2003). *Assume that (i) for any $\theta \neq \theta_0$, equation (8) holds; (ii) $\mathbb{E}\left[\left|\sum_{j=1}^{K} 1(D = j) \log q_j(Y|\mathbf{X}, \theta)\right|\right]$ is bounded; (iii) $\theta_0 \in \Theta$ and Θ is compact for the norm $\|\theta\|$; and (iv) for any $\theta \in \Theta$, there exists $\theta_J \in \Theta_J$*

https://doi.org/10.1017/S0266466612000795 Published online by Cambridge University Press

such that $\lim_{J\to\infty}\|\theta_J-\theta\|=0$*; (v) for each* $j=1,\ldots,K$*, there exists* $M_j(y,\mathbf{x})$ *such that*

$$|\log q_j(y|\mathbf{x},\theta)-\log q_j(y|\mathbf{x},\tilde{\theta})|\le M_j(y,\mathbf{x})\left\|\theta-\tilde{\theta}\right\|$$

and $\mathbb{E}[M_j(y,\mathbf{x})^2]<\infty$*. Then* $n\to\infty$*,* $\left\|\hat{\theta}_n-\theta_0\right\|\to_p 0$.

As mentioned above, condition (i) assumes identification of the model. Condition (ii) ensures that $Q_n(\theta)$ is well defined with probability approaching one and that the probability limit of $Q_n(\theta)$ has a unique maximum at θ_0. Conditions (iii) and (iv) are satisfied if we use the results of Gallant and Nychka (1987) by considering a Sobolev norm as in Gallant and Nychka and in Newey and Powell (2003). Condition (v) is a readily verified smoothness assumption that ensures the uniform convergence of $Q_n(\theta)$ to its probability limit.

4. AN EXAMPLE

This section considers a simple example of the accelerated failure time competing risks models. As in Clayton and Cuzick (1985) and Heckman and Honoré (1989), we consider a competing risks model with frailty. Specifically, suppose that the model in (1) holds with $K=2$ and $U_k=\sigma_k(\eta+\varepsilon_k)$, where ε_1 are ε_2 are independently and identically distributed from a Type 1 extreme value distribution, η is a frailty term that induces the dependence between U_1 and U_2, and σ_1 and σ_2 are strictly positive scale parameters satisfying $\sigma_1\neq\sigma_2$. Further, assume that η is independent of ε_1 and ε_2. This model is a special case of our model and can also be viewed as a mixed proportional hazards competing risks model with a different Weibull baseline hazard function for each risk. Finally, assume that

$$\mathbb{E}[\exp(-\eta)]<\infty. \tag{9}$$

This finite moment condition is standard in the literature (see, e.g., Elbers and Ridder, 1982; Heckman and Honoré, 1989).

When $h_j(\mathbf{x})=g_j(\mathbf{x})$, the joint survivor function of T_1 and T_2 then has the form

$$P[t-h_1(\mathbf{x}),t-h_2(\mathbf{x})]=\int\exp\left[-\exp\left\{\frac{1}{\sigma_1}[t-h_1(\mathbf{x})]-\eta\right\}\right]\times\exp\left[-\exp\left\{\frac{1}{\sigma_2}[t-h_2(\mathbf{x})]-\eta\right\}\right]dG_\eta(\eta), \tag{10}$$

where $G_\eta(\cdot)$ is the distribution function of η. Now we have

$$-P_1[t-h_1(\mathbf{x}),t-h_2(\mathbf{x})]=\frac{1}{\sigma_1}\exp\left\{\frac{1}{\sigma_1}[t-h_1(\mathbf{x})]\right\}\Phi(t,\mathbf{h}(\mathbf{x})),$$

$$-P_2[t-h_1(\mathbf{x}),t-h_2(\mathbf{x})]=\frac{1}{\sigma_2}\exp\left\{\frac{1}{\sigma_2}[t-h_2(\mathbf{x})]\right\}\Phi(t,\mathbf{h}(\mathbf{x})),$$

https://doi.org/10.1017/S0266466612000795 Published online by Cambridge University Press

where

$$\Phi(t,\mathbf{h}(\mathbf{x})) := \int \exp(-\eta)\exp\left[-\exp\left\{\frac{1}{\sigma_1}[t-h_1(\mathbf{x})]-\eta\right\}\right] \times \exp\left[-\exp\left\{\frac{1}{\sigma_2}[t-h_2(\mathbf{x})]-\eta\right\}\right] dG_\eta(\eta).$$

It follows that for $j,k=1,2$,

$$\dot{\mathbf{C}}^*_{jk}[\mathbf{h}(\mathbf{x})] = \frac{1}{\sigma_k}\exp\left(\frac{-h_k(\mathbf{x})}{\sigma_k}\right)\Psi(s_j,\sigma_k,\mathbf{h}(\mathbf{x})),$$

where $\Psi(s,\sigma,\mathbf{h}(\mathbf{x})) := \Psi_1(s,\sigma)\Psi_2(\mathbf{h}(\mathbf{x}))$,

$$\Psi_1(s,\sigma) := \int_s^\infty \exp\left(\frac{t}{\sigma}\right)\exp\left[-\exp\left(\frac{t}{\sigma_1}\right)\right]\exp\left[-\exp\left(\frac{t}{\sigma_2}\right)\right]dt,$$

$$\Psi_2(\mathbf{h}(\mathbf{x})) := \int \exp(-\eta)\exp\left[-\exp\left\{\frac{-h_1(\mathbf{x})}{\sigma_1}-\eta\right\}\right] \times \exp\left[-\exp\left\{\frac{-h_2(\mathbf{x})}{\sigma_2}-\eta\right\}\right] dG_\eta(\eta).$$

Note that the dependence of $\Psi_1(s,\sigma)$ and $\Psi_2(\mathbf{h}(\mathbf{x}))$ on (σ_1,σ_2) is implicit to ease the complexity of the notation. The model in this example has the feature that $\partial P(t-h_1,t-h_2)/\partial h_j$ equals $\partial P(t-h_1,t-h_2)/\partial t$ times a function that does not depend on t. This feature yields a simple closed-form expression for $\dot{\mathbf{C}}^*_{jk}$ here, since construction of $\dot{\mathbf{C}}^*_{jk}$ involves integrating $\partial P(t-h_1,t-h_2)/\partial h_j$ over t. In particular, this feature along with the simple form of the joint survivor function in (10) is what makes $\Psi(s,\sigma,\mathbf{h}(\mathbf{x}))$ in $\dot{\mathbf{C}}^*_{jk}[\mathbf{h}(\mathbf{x})]$ factor into the product of an integral over t (evaluated at s) and an integral over η.

Now note that $\dot{\mathbf{C}}^*_{jj}[\mathbf{h}(\mathbf{x})]$ is strictly positive for each $j=1,2$ for all $\mathbf{h}(\mathbf{x})$ and $\Psi_2(\mathbf{h}(\mathbf{x}))$ is bounded uniformly in $\mathbf{h}(\mathbf{x})$ in view of (9) and the fact that $\exp[-\exp\{\sigma_k^{-1}[t-h_k(\mathbf{x})]-\eta\}]$ is a survivor function for each $k=1,2$. Note also that

$$\begin{aligned}
&\dot{\mathbf{C}}^*_{11}[\mathbf{h}(\mathbf{x})]\dot{\mathbf{C}}^*_{22}[\mathbf{h}(\mathbf{x})] - \dot{\mathbf{C}}^*_{12}[\mathbf{h}(\mathbf{x})]\dot{\mathbf{C}}^*_{21}[\mathbf{h}(\mathbf{x})] \\
&\quad= \frac{1}{\sigma_1}\exp\left(\frac{-h_1(\mathbf{x})}{\sigma_1}\right)\Psi(s_1,\sigma_1,\mathbf{h}(\mathbf{x}))\frac{1}{\sigma_2}\exp\left(\frac{-h_2(\mathbf{x})}{\sigma_2}\right)\Psi(s_2,\sigma_2,\mathbf{h}(\mathbf{x})) \\
&\qquad- \frac{1}{\sigma_1}\exp\left(\frac{-h_1(\mathbf{x})}{\sigma_1}\right)\Psi(s_2,\sigma_1,\mathbf{h}(\mathbf{x}))\frac{1}{\sigma_2}\exp\left(\frac{-h_2(\mathbf{x})}{\sigma_2}\right)\Psi(s_1,\sigma_2,\mathbf{h}(\mathbf{x})) \\
&\quad= \frac{1}{\sigma_1\sigma_2}\exp\left(\frac{-h_1(\mathbf{x})}{\sigma_1}+\frac{-h_2(\mathbf{x})}{\sigma_2}\right)[\Psi_2(\mathbf{h}(\mathbf{x}))]^2\,\Delta(s_1,s_2,\sigma_1,\sigma_2),
\end{aligned}$$

where

$$\Delta(s_1,s_2,\sigma_1,\sigma_2) := \Psi_1(s_1,\sigma_1)\Psi_1(s_2,\sigma_2) - \Psi_1(s_2,\sigma_1)\Psi_1(s_1,\sigma_2).$$

https://doi.org/10.1017/S0266466612000795 Published online by Cambridge University Press

We can find a pair of s_1 and s_2 such that $\Delta(s_1, s_2, \sigma_1, \sigma_2) \neq 0$ as long as $\sigma_1 \neq \sigma_2$. Thus, there exist universal constants s_1 and s_2 such that the determinant of $\dot{\mathbf{C}}^*[\mathbf{h}(\mathbf{x})]$ is nonzero for all $\mathbf{h}(\mathbf{x})$ as long as $\sigma_1 \neq \sigma_2$. The condition that $\sigma_1 \neq \sigma_2$ is important. If $\sigma_1 = \sigma_2$, the determinant of $\dot{\mathbf{C}}^*[\mathbf{h}(\mathbf{x})]$ is zero for any $\mathbf{h}(\mathbf{x})$ and for any s_1, s_2. In our example, different σ_k's mean that different causes of failure affect the latent baseline hazard functions differently. Given $\sigma_1 \neq \sigma_2$, we have now shown that condition (ii) of Assumption 2 is satisfied.

We now turn to verification of Assumption 3. It suffices to show that there exist universal constants $\delta(> 0), s_1, s_2$ such that

$$\dot{\mathbf{C}}^*_{jj}[\mathbf{h}_j(\mathbf{x})] > \delta \tag{11}$$

for each $j = 1, 2$, and that

$$\left\{\dot{\mathbf{C}}^*_{11}[\mathbf{h}_1(\mathbf{x})] - \delta\right\}\left\{\dot{\mathbf{C}}^*_{22}[\mathbf{h}_2(\mathbf{x})] - \delta\right\} - \dot{\mathbf{C}}^*_{12}[\mathbf{h}_1(\mathbf{x})]\dot{\mathbf{C}}^*_{21}[\mathbf{h}_2(\mathbf{x})] > 0 \tag{12}$$

for all $\mathbf{h}_1(\mathbf{x})$ and $\mathbf{h}_2(\mathbf{x})$. One sufficient condition for (11) to hold is to assume that all elements of $\mathcal{G}$ are bounded functions, so that $\mathbf{h}_1$ and $\mathbf{h}_2$ are bounded. In other words, condition (11) can be satisfied if the range of regression functions $g_k(\mathbf{x})$ is bounded (or the support of $\mathcal{X}$ is compact and $g_k(\mathbf{x})$ is continuous). It is difficult to construct a lower-level sufficient condition than (12) without making stringent assumptions. For example, doing so would likely require that there exists a sufficiently large difference between σ_1 and σ_2, in addition to the boundedness of $\mathbf{h}_1$ and $\mathbf{h}_2$.

5. CONCLUSIONS

We have shown general identification of a competing risks or Roy model, where each risk is modeled as a nonparametric regression. The conditional mean functions of each risk and the joint survivor (distribution) function of the latent risk errors are nonparametrically identified and can be estimated using a sieve MLE.

Our results do not depend on identification at infinity or in the neighborhood of zero, and they do not depend on exclusion restrictions. However, they do assume that the errors in the risks are independent of the covariates (which is a common modeling assumption), and they depend on a rank condition, which requires at least as many continuous covariates as there are risks. The necessity of these covariate conditions should not be surprising, since identification essentially requires that it be possible for each risk to vary as a function of observables, holding the other risks fixed.

We conclude by noting one small extension. Many empirical applications of competing risks models include censoring, for example, if the risks correspond to different ways one could exit unemployment (such as finding a job or dropping out of the labor force); then, in a data set with attrition, incomplete unemployment spells will be observed for those people who are unemployed when they drop out of the survey. Such censoring is immediately encompassed by our result, since it

https://doi.org/10.1017/S0266466612000795 Published online by Cambridge University Press

can be treated as just another risk, as if exiting the survey were another way to exit unemployment. In the sieve estimator, this risk may be given a simplified model, with a conditional mean function that depends only on covariates that affect the probability of attrition (such as time in the survey), and possibly an error that is independent of the latent errors in the other risks. Indeed, since random censoring can be intepreted as a special case of a risk, our result with $K = 2$ also provides identification for a general class of ordinary random censoring models.

REFERENCES

Abbring, J.H. & G.J. Van den Berg (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 (2005) Social Experiments and Instrumental Variables with Duration Outcomes. Tinbergen Institute discussion paper TI 2005-047/3.

Ai, C. & X. Chen (2003) Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica* 71, 1795–1843.

Andrews, D.W.K. & M.M.A. Schafgans (1998) Semiparametric estimation of the intercept of a sample selection model. *Review of Economic Studies* 65, 497–517.

Athey, S. & P.A. Haile (2002) Identification of standard auction models. *Econometrica* 70(6), 2107–2140.

Bayer, P.J., S. Khan, & C. Timmins (2011) Nonparametric identification and estimation in a Roy model with common nonpecuniary returns. *Journal of Business and Economic Statistics* 29(2), 201–215.

Berger, M.S. (1977) *Nonlinearity and Functional Analysis: Lectures on Nonlinear Problems in Mathematical Analysis*. Academic Press.

Bond, S.J. & J.E.H. Shaw (2006) Bounds on the covariate-time transformation for competing risks survival analysis. *Lifetime Data Analysis* 12, 285–303.

Buera, F.J. (2006) Non-parametric Identification and Testable Implications of the Roy Model. Working paper, Northwestern University and UCLA.

Chen, X. (2007) Large sample sieve estimation of semi-nonparametric models. In J.J. Heckman & E.E. Leamer (eds.), *Handbook of Econometrics*, vol. 6B, chap. 76, 5549–5632.

Chen, X. & X. Shen (1998) Sieve extremum estimates for weakly dependent data. *Econometrica* 66, 289–314.

Chernozhukov, V., G. Imbens, & W. Newey (2007) Instrumental variable identification and estimation of nonseparable models via quantile conditions. *Journal of Econometrics* 139, 4–14.

Clayton, D. & J. Cuzick (1985) Multivariate generalization of the proportional hazard model. *Journal of the Royal Statistical Society, Series A* 148, 82–117.

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

Elbers, C. & G. Ridder (1982) True and spurious duration dependence: The identifiability of the proportional hazard model. *Review of Economic Studies* 49, 403–409.

Femanian, J.-D. (2003) Nonparametric estimation of competing risks models with covariates. *Journal of Multivariate Analysis* 85, 156–191.

Fox, J. T. & A. Gandhi (2009) Identifying Heterogeneity in Economic Choice and Selection Models Using Mixtures. Working paper, University of Chicago and University of Wisconsin.

French, E. & C. Taber (2011) Identification of models of the labor market. In O. Ashenfelter & D. Card (eds.), *Handbook of Labor Economics*, vol. 4A, chap. 6, 537–617.

Gallant, A.R. & D. Nychka (1987) Semi-non-parametric maximum likelihood estimation. *Econometrica* 55, 363–390.

Han, A. & J.A. Hausman (1990) Flexible parametric estimation of duration and competing risks models. *Journal of Applied Econometrics* 5, 1–28.

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

https://doi.org/10.1017/S0266466612000795 Published online by Cambridge University Press

Heckman, J.J. & B.E. Honoré (1989) The identifiability of the competing risks model. *Biometrika* 76, 325–30.
Heckman, J.J. & B.E. Honoré (1990) The empirical content of the Roy model. *Econometrica* 58, 1121–1149.
Honoré, B.E. & A. Lleras-Muney (2006) Bounds in competing risks models and the war on cancer. *Econometrica* 74, 1675–1698.
Ichimura, H. & L.-F. Lee (1991) Semiparametric least squares estimation of multiple index models: single equation estimation. In W.A. Barnett, J. Powell, & G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, pp. 3–49. Cambridge University Press.
Khan, S. & E. Tamer (2009) Inference on endogenously censored regression models using conditional moment inequalities. *Journal of Econometrics* 152, 104–119.
Khan, S. & E. Tamer (2010) Irregular identification, support conditions, and inverse weight estimation. *Econometrica* 78, 2021–2042.
Komunjer, I. (2012) Global identification in nonlinear models with moment restrictions. *Econometric Theory* 28, 719–729.
Lee, S. (2006) Identification of a competing risks model with unknown transformations of latent failure times. *Biometrika* 93, 996–1002.
Matzkin, R.L. (2007) Nonparametric identification. In J.J. Heckman & E.E. Leamer (eds.), *Handbook of Econometrics*, vol. 6, p. 2, chap. 73, pp. 5307–5368. Elsevier.
Newey, W.K. & J.L. Powell (2003) Instrumental variable estimation of nonparametric models. *Econometrica* 71, 1565–1578.
Peterson, A.V. (1976) Bounds for a joint distribution with fixed sub-distribution functions: Application to competing risks. *Proceedings of the National Academy of Sciences* 73, 11–13.
Shen, X. (1997) On methods of sieves and penalization. *Annals of Statistics* 25, 2555–2591.
Tsiatis, A.A. (1975) A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences* 72, 20–22.
Zheng, M. & J.P. Klein (1995) Estimates of marginal survival for dependent competing risks based on an assumed copula. *Biometrika* 82, 127–138.
Zeidler, E. (1986) *Nonlinear Functional Analysis and Its Applications I: Fixed-Point Theorems*. Springer Verlag.

APPENDIX: Proofs of Theorems

Proof of Theorem 1. Define $M(t_1,\ldots,t_K \mid \mathbf{x}) = \Pr(T_k > t_k, k = 1,\ldots,K \mid \mathbf{X} = \mathbf{x})$ and $M_k(t_1,\ldots,t_K \mid \mathbf{x}) = \partial M(t_1,\ldots,t_K \mid \mathbf{x})/\partial t_k$. Observe that B_k and C are related by

$$B_k(s \mid \mathbf{x}) = -\int_s^\infty M_k(t,\ldots,t \mid \mathbf{x})dt \quad \text{and} \quad C(s \mid \mathbf{x}) = \int_s^\infty M(t,\ldots,t \mid \mathbf{x})dt.$$

Define

$$P(u_1,\ldots,u_K) := \Pr(U_k > u_k, k = 1,\ldots,K) \quad \text{and}$$
$$P_k(u_1,\ldots,u_K) := \partial P(u_1,\ldots,u_K)/\partial u_k.$$

Then

$$M(t_1,\ldots,t_K \mid \mathbf{x}) = P\left[t_1 - g_1(\mathbf{x}),\ldots,t_K - g_K(\mathbf{x})\right],$$

and

$$\begin{aligned} M_k(t_1,\ldots,t_K \mid \mathbf{x}) &= P_k\left[t_1 - g_1(\mathbf{x}),\ldots,t_K - g_K(\mathbf{x})\right] \\ &= -\partial P\left[t_1 - g_1(\mathbf{x}),\ldots,t_K - g_K(\mathbf{x})\right]/\partial g_k(\mathbf{x}). \end{aligned}$$

https://doi.org/10.1017/S0266466612000795 Published online by Cambridge University Press

Therefore,

$$\begin{aligned} C(s \mid \mathbf{x}) &= \int_s^\infty M(t, \ldots, t \mid \mathbf{x}) dt \\ &= \int_s^\infty P\left[t - g_1(\mathbf{x}), \ldots, t - g_K(\mathbf{x})\right] dt \\ &= C^*[s, g_1(\mathbf{x}), \ldots, g_K(\mathbf{x})], \end{aligned}$$

and

$$\begin{aligned} \left.\frac{\partial C^*[s, h_1(\mathbf{x}), \ldots, h_K(\mathbf{x})]}{\partial h_k(\mathbf{x})}\right|_{\mathbf{h}=\mathbf{g}} &= \frac{\partial \int_s^\infty P\left[t - g_1(\mathbf{x}), \ldots, t - g_K(\mathbf{x})\right] dt}{\partial h_k(\mathbf{x})} \\ &= -\int_s^\infty P_k\left[t - g_1(\mathbf{x}), \ldots, t - g_K(\mathbf{x})\right] dt \\ &= -\int_s^\infty M_k(t, \ldots, t \mid \mathbf{x}) dt = B_k(s \mid \mathbf{x}). \end{aligned}$$

■

Proof of Theorem 2. For each $\mathbf{h}_1$, note that $\dot{\mathbf{C}}^*(\mathbf{h}_1)\mathbf{h}_2$ is bijective (as a linear function of $\mathbf{h}_2$ from $L_2(\mathcal{X})$ to itself, while $\mathbf{h}_1$ is fixed), since $\dot{\mathbf{C}}^*[\mathbf{h}(\mathbf{x})]$ is invertible for all $\mathbf{h} \in \mathcal{G}$ (condition (ii)). Then $\mathbf{C}^*(\mathbf{h})$ is a local C^1-diffeomorphism at each $\mathbf{h} \in \mathcal{G}$ by the (local) inverse mapping theorem (see, e.g., Zeidler, 1986, Thm. 4.F, p. 172). Furthermore, (4) is a direct consequence of that (see, also, Zeidler, Cor. 4.37, p. 172). Note that $\mathbf{C}^*$ is a continuous mapping and $L_2(\mathcal{X})$ is a reflexive Banach space with its dual $L_2(\mathcal{X})^*$ being isometrically isomorphic to $L_2(\mathcal{X})$. Now inequality (3) verifies condition (†) in (5.1.6) of Berger (1977, p. 223) by taking a positive function $\eta(r)$ in Berger's condition (using his notation) to be $\eta(r) = \delta r^\epsilon$. This shows that $\mathbf{C}^*$ is a (global) C^1-diffeomorphism, or equivalently that $\mathbf{C}^*$ is proper in view of Theorem 4.G of Zeidler (p. 174). ■

Proof of Theorem 4. For identifying the distribution of $\mathbf{U}$ given the identification of $\mathbf{g}(\mathbf{x})$, let $A(s \mid \mathbf{g}) = \mathbb{E}[I(Y > s) \mid \mathbf{g}(\mathbf{x}) = \mathbf{g}]$ for any s in the support of Y. Then $A(s \mid \mathbf{g}) = P(s - g_1, \ldots, s - g_K)$. It follows that $A(s \mid \mathbf{g})$ identifies this joint survivor distribution over the range of values of $u_1 = s - g_1(\mathbf{x}), \ldots, u_K = s - g_K(\mathbf{x})$. ■

Proof of Theorem 5. This can be proved as in the proof of Theorem 4.1 of Newey and Powell (2003) with minor modifications. ■

https://doi.org/10.1017/S0266466612000795 Published online by Cambridge University Press