Identification of a Competing Risks Model with Unknown Transformations of Latent Failure Times

Sokbae Lee
Centre for Microdata Methods and Practice
Institute for Fiscal Studies
and
Department of Economics
University College London
London, WC1E 6BT, UK.
Email: l.sokbae@ucl.ac.uk

July 2005

Abstract

This paper is concerned with identification of a competing risks model with unknown transformations of latent failure times. The model in this paper includes, as special cases, competing risks versions of log-linear, proportional hazards, mixed proportional hazards, proportion odds rate, and accelerated failure time models. It is shown that important features of covariate effects on latent failure times can be identified without relying on modelling the dependence between latent failure times parametrically nor using an exclusion restriction among covariates. In addition, the paper presents a method for estimating the covariate effects based on the identification result.

Key words: Competing risks model; Identification; Semiparametric estimation; Transformation model.

1 Introduction

This paper is concerned with identification of a competing risks model with unknown transformations of latent failure times. Suppose that there are J competing causes of failure indexed by the integers 1 to J with corresponding latent failure times $(T_1, \ldots T_J)$. One observes the duration to the first failure and the corresponding cause of failure, denoted by $Y = \min_j T_j$ and $\Delta = \arg\min_j T_j$, along with explanatory variables. It is well known (see, for example, Cox (1962) and Tsiatis (1975)) that the distribution of latent failure times is nonparametrically unidentified. Heckman and Honoré (1989) and Abbring and Van den

Berg (2003), among others, demonstrate that one can break this nonidentification result by considering a certain class of models for latent failure times and by exploiting sufficiently independent variations of latent failure times with explanatory variables.

The main purpose of this paper is to provide weak restrictions that are sufficient to identify important features of a competing risks model and to develop a semiparametric estimator based on the identification result. The model and restrictions imposed in this paper are quite different from those of Heckman and Honoré (1989) and Abbring and Van den Berg (2003) and can be viewed as an alternative modelling framework. Specifically, we consider a transformation model for each latent failure time and also get around a difficult problem of identifying the scale factor of covariate effects (or equivalently, the scale factor of a link function). Our identification results are sufficiently weak in a sense that it is not needed to have a parametric form of dependence nor an exclusion restriction on covariates.

It is assumed in this paper that each latent failure time T_j is generated by a linear transformation regression model:

$$H_j(T_j) = X'\beta_j + U_j, \quad j = 1, \ldots, J, \tag{1}$$

where $H_j(\cdot)$ is an unknown, differentiable, strictly increasing function with a derivative $h_j(\cdot)$, X a d-dimensional vector of continuous explanatory variables (not including a constant term), β_j a d-dimensional vector of unknown parameters, and U_j is an unobserved random variable that is independent of X. It is also assumed that the distribution of U_j is unknown and U_j may depend on each other.

The model (1) includes, as special cases, competing risks versions of log-linear, proportional hazards, mixed proportional hazards, proportion odds rate, and accelerated failure time models and may be called a *competing risks transformation model.* For example, the mixed proportional hazards competing risks model can be expressed as a special case of (1) with $U_j = \alpha_j + \varepsilon_j$, where α_j is a cause-specific frailty term, ε_j is an unobserved random variable that has CDF $F_j(\varepsilon) = 1 - \exp(-e^{\varepsilon})$ and $\exp[H_j(t)]$ is the integrated baseline hazard function (Clayton and Cuzick (1985) and Abbring and Van den Berg (2003)) and the accelerated failure time competing risks model may be presented as (1) in which $H_j(t) = \log t$ (Heckman and Honoré (1989)). In survival analysis, transformation models have been studied intensively for single risks data (see, e.g., Cheng, Wei, and Ying (1995) and Horowitz (1996 and 1998, Chap. 5)). Fine (1999) uses a transformation model to analyze the cumulative incidence function for competing risks data.

This paper provides an identification result on β_j, that is a vector of parameters that measure the effects of X on latent failure time T_j. Since H_j and the distribution of U_j are unknown, β_j is identified only up to some location and scale normalization. Also, because H_j can be different from H_k, β_j is not directly comparable to β_k for $j \neq k$. Therefore, only the direction of β_j is identified and ratios between components of β_j give relative importance of components of X.

The paper is organized as follows. Section 2 provides an identification result for β_j. Section 3 presents a sample analog estimator based on the result of Section 2. Section 4 concludes. Asymptotic properties of the estimator of β_j and results of a simulation study are given in the Appendix.

2 Identification

This section provides conditions under which β_j in (1) is identified. In this paper, we assume that $d \geq 2$. Otherwise, there is nothing left to identify since the scale factor has to be normalized. To identify β_j, define $S(t|x) = \text{Prob}(Y > t|X = x)$ and $Q_j(t|x) = \text{Prob}(T_j > t, \Delta = j|X = x)$ for $j = 1, \ldots, J$. Also, let $p(x)$ denote the probability density function of X, S_X the support of X, and x_k and $\beta_j^{(k)}$ the k-th components of x and β_j, respectively. For $k = 1, \ldots, d$, and $j = 1, \ldots, J$, define

$$A_k(t,x) = \frac{\partial S(t|x)}{\partial x_k} p(x),$$
$$B_j(t,x) = -\frac{\partial Q_j(t|x)}{\partial t} p(x),$$

$B(t,x) = [B_1(t,x), \ldots, B_J(t,x)]'$, and $\beta^{(k)} = (\beta_1^{(k)}, \beta_2^{(k)}, \ldots, \beta_J^{(k)})'$ for $k = 1, \ldots, d$. Thus, $\beta^{(k)}$ is a $(J \times 1)$ vector of unknown cause-specific coefficients of the k-th component of X.

To achieve identification of β_j, we make the following assumptions:

Assumption 1 (Identification of β_j). *(a) $(U_1, \ldots, U_J)$ are continuously distributed and independent of X but may be arbitrarily correlated with one another.*
(b) X is a $d(\geq 2)$-dimensional vector of continuous explanatory variables and has a joint probability density function $p(x)$ that is positive on S_X except on the boundary.
(c) $H_j(\cdot)$ is an unknown, differentiable, strictly increasing function with a derivative $h_j(\cdot)$.
(d) For each $j = 1, \ldots, J$,

$$\int [w_T(t)/h_j(t)]dt = 1, \tag{2}$$

where $w_T(t)$ is a weight function with compact support S_T.
(e) Assume that as functions of $x \in S_X$, components of $B(t,x)$ are linearly independent for every $t \in S_T$.

Condition (a) allows for arbitrary correlations among U_j. Location normalization is achieved by excluding an intercept term in X (see condition (b)). Condition (c) is common in analyzing a transformation model. Scale normalization is accomplished by condition (d). This assumption is useful to create averaging effects, so that a sample analog estimator based on our identification result converges in probability at a rate of $n^{-1/2}$, where n is the sample size. The same type of scale normalization is used for similar reasons in Horowitz (2001) and Horowitz and Lee (2004). Condition (e) amounts to assuming that cause-specific sub-densities of latent failure times conditional on $X = x \in S_X$ are linearly independent for every $t \in S_T$. There are no possible values of t and x satisfying this condition if β_j, H_j, α_j, and F_j are identical over $j = 1, \ldots, J$; however, this is not an interesting case to use a competing risks model.

The following theorem gives a constructive identification result for β_j.

Theorem 1. *Let Assumption 1 hold. Then for each $k = 1, \ldots, d$, $\beta^{(k)}$ can be expressed as*

$$\beta^{(k)} = \int w_T(t) E[B(t,X)B(t,X)']^{-1} E[B(t,X)A_k(t,X)]\, dt. \tag{3}$$

Proof of Theorem 1. Let $f(u_1, \ldots, u_J)$ denote the joint probability density function of $(U_1, \ldots, U_J)$. Notice that

$$\begin{aligned} S(t|x) &= \Pr(H_j(T_j) > H_j(t) \text{ for all } j | X = x) \\ &= \Pr(U_j > H_j(t) - x'\beta_j \text{ for all } j) \\ &= \int_{H_1(t)-x'\beta_1}^{\infty} \cdots \int_{H_J(t)-x'\beta_J}^{\infty} f(u_1, \ldots, u_J)\, du_1 \cdots du_J \end{aligned}$$

and

$$\begin{aligned} Q_j(t|x) &= \Pr(H_j(T_j) > H_j(t) \text{ and } H_l(T_l) > H_l(T_j) \text{ for all } l \neq j | X = x) \\ &= \Pr(U_j > H_j(t) - x'\beta_j \text{ and } U_l > H_l(T_j) - x'\beta_l \text{ for all } l \neq j) \\ &= \int_{H_j(t)-x'\beta_j}^{\infty} \underbrace{\int_{H_1[H_j^{-1}(x'\beta_j+u_j)]-x'\beta_1}^{\infty} \cdots \int_{H_J[H_j^{-1}(x'\beta_j+u_j)]-x'\beta_J}^{\infty}}_{J \text{ integrals excluding } j} f(u_1, \ldots, u_J) \underbrace{du_1 \cdots du_J}_{du_j \text{ is excluded}}\, du_j \end{aligned}$$

for $j = 1, \ldots, J$. By differentiation,

$$\frac{\partial Q_j(t|x)}{\partial t} = -h_j(t) \underbrace{\int_{H_1(t)-x'\beta_1}^{\infty} \cdots \int_{H_J(t)-x'\beta_J}^{\infty}}_{J \text{ integrals excluding } j} f(u_1, \ldots, u_{j-1}, H_j(t) - x'\beta_j, u_{j+1}, \ldots, u_J) \underbrace{du_1 \cdots du_J}_{du_j \text{ is excluded}}$$

and

$$\frac{\partial S(t|x)}{\partial x_k} = \sum_{j=1}^{J} \beta_j^{(k)} \underbrace{\int_{H_1(t)-x'\beta_1}^{\infty} \cdots \int_{H_J(t)-x'\beta_J}^{\infty}}_{J \text{ integrals excluding } j} f(u_1, \ldots, u_{j-1}, H_j(t) - x'\beta_j, u_{j+1}, \ldots, u_J) \underbrace{du_1 \cdots du_J}_{du_j \text{ is excluded}},$$

where x_k and $\beta_j^{(k)}$ are the k-th components of x and β_j, respectively. It follows that

$$\frac{\partial S(t|x)}{\partial x_k} = \sum_{j=1}^{J} \frac{-\partial Q_j(t|x)}{\partial t} \frac{\beta_j^{(k)}}{h_j(t)}. \tag{4}$$

Multiplying by $p(x)$ both right and left sides of (4) gives

$$\frac{\partial S(t|x)}{\partial x_k} p(x) = \sum_{j=1}^{J} \frac{-\partial Q_j(t|x)}{\partial t} p(x) \frac{\beta_j^{(k)}}{h_j(t)}. \tag{5}$$

To express identifying relationships compactly, define, for $k = 1, \ldots, d$, let

$$b_k(t) = \left(\frac{\beta_1^{(k)}}{h_1(t)}, \frac{\beta_2^{(k)}}{h_2(t)}, \ldots, \frac{\beta_J^{(k)}}{h_J(t)} \right)'.$$

Then it follows from (5) that

$$A_k(t, x) = B(t, x)' b_k(t).$$

To identify $\beta^{(k)}$, write

$$B(t, x) A_k(t, x) = B(t, x) B(t, x)' b_k(t). \tag{6}$$

To solve for $b_k(t)$, substitute the random vector X for x in (6) and take expectations to obtain

$$E[B(t, X) A_k(t, X)] = E[B(t, X) B(t, X)'] b_k(t). \tag{7}$$

By the assumption that components of $B(t, x)$ are linearly independent for every $t \in S_T$, we have that $E[B(t, X)B(t, X)']$ is nonsingular for every $t \in S_T$. Therefore, under the scale normalization (2), $\beta^{(k)}$ can be expressed as in the equation (3), which proves the theorem. □

It can be seen from the expression of $S(t|x)$ in the proof of Theorem 1 that the expectation of Y conditional on $X = x$ belongs to the class of multiple-index models (see, for example, Ichimura and Lee (1991)). Typically, certain exclusion restrictions (for example, certain components of parameters are zero) are needed for multiple-index models to achieve identification of parameters. As shown by Heckman and Honoré (1989), Abbring and Van den Berg (2003), and equation (3), exclusion restrictions are not required for the identification of semiparametric competing risks models.

It is important to notice that there exists an important difference between identification results of Heckman and Honoré (1989) and Abbring and Van den Berg (2003) and one obtained in Theorem 1. Those of Heckman and Honoré (1989) and Abbring and Van den Berg (2003) are based on the arguments of letting $t \to 0$, thereby implying that corresponding estimation methods would be based on only observations with failure times close to zero. An estimator of Femanian (2003, Section 4) is such an example. This is mainly because the scale factor has to be identified in the setup of Heckman and Honoré (1989) and Abbring and Van den Berg (2003). The difficulty of identifying the scale factor is not specific to competing risks models. A similar problem arises in a single-risk mixed proportional hazard model (see, for example, Horowitz (1999)).

It is also important to note that the continuity of the distribution of X is not needed for identification of the covariate effects in the framework of Heckman and Honoré (1989) and Abbring and Van den Berg (2003). This is due to the fact that in their framework, the joint survivor function of latent failure times can be written as an exponential function, which does not necessarily hold in our model (1). Finally, we note that (3) is expressed in terms of a density-weighted form (see also (5)), which would be convenient to construct a resulting sample analog estimator.

When all components of X are discrete, it is unclear whether one can point-identify covariate effects. However, some recent studies show that it is possible to derive bounds for covariate effects in some competing risks models with discrete covariates. For example, Bond and Shaw (2003) 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 (2003) to bound the treatment effects on duration outcomes. Honoré and Lleras-Muney (2004) derive bounds in an accelerated failure time competing risks model with discrete covariates.

In summary, we have achieved identification of covariate effects of a competing risks transformation model that includes semiparametric versions of competing risks models analyzed in Heckman and Honoré (1989) and Abbring and Van den Berg (2003) without relying on the arguments of letting $t \to 0$. This suggests that if an applied researcher is mainly interested in the direction of β_j and relative importance between components of β_j, it is possible to develop an estimation method that uses all observations.

3 Sample Analog Estimation

The equation (3) can be used as the basis for a sample analog estimator of $\beta^{(k)}$. This section describes how one can construct such an estimator based on (3). A sample analog estimator of $\beta^{(k)}$ can be obtained by replacing unknown population quantities in (3) with nonparametric estimators. Specifically, in this paper, estimation is carried out with kernel estimators.

To do so, assume that $\{(Y_i, \Delta_i, X_i) : i = 1, \ldots, n\}$ is a random sample of (Y, Δ, X). Let $K_X(\cdot)$ denote a d-dimensional kernel function with a bandwidth h_{nx} and $K_Y(\cdot)$ denote a one-dimensional kernel function with a bandwidth h_{ny}. Also, let $1(\cdot)$ denote a usual indicator function. For each $X_i \in S_X$, estimate the j-th component $B_j(t, X_i)$ of $B(t, X_i)$ by

$$B_n^{(j)}(t, X_i) = n^{-1} \sum_{l=1}^{n} 1(\Delta_l = j) K_{h_{ny}}(t - Y_l) K_{h_{nx}}(X_i - X_l),$$

where

$$K_{h_{ny}}(t - Y_l) = \frac{1}{h_{ny}} K_Y\left(\frac{t - Y_l}{h_{ny}}\right) \quad \text{and} \quad K_{h_{nx}}(X_i - X_l) = \frac{1}{h_{nx}^d} K_X\left(\frac{X_i - X_l}{h_{nx}}\right).$$

Then $E[B(t, X)B(t, X)']$ can be estimated by a sample average of $B_n(t, X_i)B_n(t, X_i)'$, where $B_n(t, X_i) = [B_n^{(1)}(t, X_i), \ldots, B_n^{(J)}(t, X_i)]'$.

To estimate $E[B(t, X)A_k(t, X)]$, define $S(t, x) = S(t|x)p(x)$, $S^{(k)}(t, x) = \partial S(t, x)/\partial x_k$, and $p^{(k)}(x) = \partial p(x)/\partial x_k$. Notice that

$$E[B(t, X)A_k(t, X)] = E\left[B(t, X)S^{(k)}(t, X)\right] - E\left[B(t, X)1(Y > t)p^{(k)}(X)\right].$$

For each $X_i \in S_X$, estimate $S^{(k)}(t, X_i)$ and $p^{(k)}(X_i)$, respectively, by partial derivatives of kernel estimators of $S(t, X_i)$ and $p(X_i)$. That is,

$$S_n^{(k)}(t, X_i) = n^{-1} \sum_{l=1}^{n} 1(Y_l > t) K_{h_{nx}}^{(k)}(X_i - X_l) \quad \text{and} \quad p_n^{(k)}(X_i) = n^{-1} \sum_{l=1}^{n} K_{h_{nx}}^{(k)}(X_i - X_l),$$

where

$$K_{h_{nx}}^{(k)}(X_i - X_l) = \frac{1}{h_{nx}^{d+1}} K_X^{(k)}\left(\frac{X_i - X_l}{h_{nx}}\right)$$

and $K_X^{(k)}$ is the partial derivative of K_X with respect to the k-th argument of K_X. Then $E[B(t, X)A_k(t, X)]$ can be estimated by sample averages. Therefore, our estimator of $\beta^{(k)}$

is defined as

$$\hat{\beta}_n^{(k)} = \int w_T(t) \left[n^{-1} \sum_{i=1}^{n} \tau(X_i) B_n(t, X_i) B_n(t, X_i)' \right]^{-1} \times \left\{ \left[n^{-1} \sum_{i=1}^{n} \tau(X_i) B_n(t, X_i) S_n^{(k)}(t, X_i) \right] - \left[n^{-1} \sum_{i=1}^{n} \tau(X_i) B_n(t, X_i) 1(Y_i > t) p_n^{(k)}(X_i) \right] \right\} dt, \tag{8}$$

where $\tau(x)$ is a trimming function such that $\tau(x) = 1(x \in \mathcal{X})$ with a compact subset $\mathcal{X}$ of $\mathbf{R}^d$. The trimming function is introduced to prevent the estimator defined in (8) from being overly influenced by the tail behavior of the distribution of X. Under suitable regularity conditions including some smoothness assumptions on the underlying distribution, it is straightforward to obtain the asymptotic properties of the proposed estimator. In fact, it can be shown that $\hat{\beta}_n^{(k)}$ converges in probability at a rate of $n^{-1/2}$ and is asymptotically normally distributed with mean zero. Asymptotic properties of $\hat{\beta}_n^{(k)}$ and results of a simulation study are given in the Appendix.

4 Conclusions

This paper has shown that a transformation model can be used to identify important features of a dependent competing risks model that includes, as special cases, competing risks versions of log-linear, proportional hazards, mixed proportional hazards, proportion odds rate, and accelerated failure time models. It would be interesting to identify and estimate link functions H_j and the joint distribution of error terms U_j. This is a topic for future research.

5 Acknowledgments

I am grateful to Jaap Abbring, Joel Horowitz, Geert Ridder, Myunghwan Seo, and Gerard Van den Berg for helpful discussions. In addition, I would like to thank the Leverhulme Trust and ESRC through the funding of the Centre for Microdata Methods and Practice and of the research programme *Evidence, Inference and Inquiry.*

Appendix

Appendix A gives regularity conditions under which the estimator of β_j is $n^{-1/2}$-consistent and asymptotically normal, Appendix B provides the proof of asymptotic normality, and Appendix C reports results of a small simulation study.

Appendix A: Asymptotic Results

This section establishes the asymptotic properties of $\hat{\beta}_n^{(k)}$. In particular, this section gives conditions under which $n^{1/2}[\hat{\beta}_n^{(k)} - \beta^{(k)}]$ is asymptotically normal. In addition to Assumption 1, we make the following assumptions:

Assumption 2 (Random Sampling)**.** *$\{(Y_i, \Delta_i, X_i) : i = 1, \ldots, n\}$ is a random sample of (Y, Δ, X), where $Y = \min_j T_j$ and $\Delta = arg\ min_j T_j$, and each latent failure time T_j is generated by model* (1)*.*

Assumption 3 (Trimming Function)**.** *The trimming function $\tau(x) = 1(x \in \mathcal{X})$ has compact support $\mathcal{X}$.*

Assumption 4 (Smoothness)**.** *The functions $S(t|x)$, $\partial S(t|x)/\partial x_k$, $B_j(t, x)$, $\partial B_j(t, x)/\partial x_k$, $p(x)$, $\partial p(x)/\partial x_k$, and $w_T(t)$ are r_y-times continuous differentiable with respect to t on S_T and r_x-times continuous differentiable with respect to x on S_X.*

Assumption 5 (Kernels)**.** *(a) K_Y has support $[-1, 1]$, is bounded and symmetrical about 0, has bounded variation, and satisfies*

$$\int_{-1}^{1} u^j K_Y(u) du = \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{if } j = 1 \leq j \leq r_y - 1, \\ C_Y & \text{if } j = r_y, \end{cases}$$

where C_Y is a positive constant.
(b) K_X is a d-dimensional product of univariate kernel functions K, that is $K_X(u) = \prod_{k=1}^{d} K(u_k)$, where $u = (u_1, \ldots, u_d)$. The function K has support $[-1, 1]$, is bounded and symmetrical about 0, has bounded variation, and satisfies

$$\int_{-1}^{1} u^j K(u) du = \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{if } j = 1 \leq j \leq r_x - 1, \\ C_X & \text{if } j = r_x, \end{cases}$$

where C_X is a positive constant.
(c) K is everywhere differentiable. $K'(v) \equiv dK(v)/dv$ is bounded and Lipschitz continuous and has bounded variation.

Assumption 6 (Bandwidths)**.** *$nh_{ny}h_{nx}^{2+d} \to \infty$, $nh_{ny}^{2r_y} \to 0$, and $nh_{nx}^{2r_x} \to 0$.*

Assumptions 5 and 6 are satisfied, for example, if K_Y is a second-order kernel, K_X is a r_x-order kernel with $r_x > [7(2 + d)]/10$, $h_{ny} \propto n^{-2/7}$, and $h_{nx} \propto n^{-\kappa_x}$ with $1/(2r_x) < \kappa_x < 5/[7(2 + d)]$.

Define

$$
\begin{aligned}
Q(t) &= E\left[\tau(X)B(t,X)B(t,X)'\right], \\
D_1(t) &= E\left[\tau(X)B(t,X)S^{(k)}(t,X)\right], \\
D_2(t) &= E\left[\tau(X)B(t,X)1(Y>t)p^{(k)}(X)\right],
\end{aligned}
$$

and let $e(\Delta_i)$ denote a $(J \times 1)$ vector such that the j-th component of $e(\Delta_i)$ is $1(\Delta_i = j)$. In addition, define

$$
\begin{aligned}
\Gamma_1^{(k)}(Y_i, X_i, \Delta_i) &= \int w(t)Q(t)^{-1}\left[\tau(X_i)B(t,X_i)S^{(k)}(t,X_i) - D_1(t)\right]dt \\
&\quad - \int w(t)Q(t)^{-1}\left[\tau(X_i)B_j(t,X_i)1(Y_i>t)p^{(k)}(X_i) - D_2(t)\right]dt, \\
\Gamma_2^{(k)}(Y_i, X_i, \Delta_i) &= e(\Delta_i)w(Y_i)Q(Y_i)^{-1}\tau(X_i)\left[S^{(k)}(Y_i,X_i)p(X_i) - S(Y_i,X_i)p^{(k)}(X_i)\right], \\
\Gamma_3^{(k)}(Y_i, X_i, \Delta_i) &= -\int w(t)Q(t)^{-1}\tau(X_i)\left[1(Y_i>t)\partial[B(t,X_i)p(X_i)]/\partial x_k - \partial[B(t,X_i)S(t,X_i)]/\partial x_k\right]dt,
\end{aligned}
$$

and $\Gamma^{(k)}(Y_i, X_i, \Delta_i) = \sum_{l=1}^{3}\Gamma_l^{(k)}(Y_i, X_i, \Delta_i)$. Finally, let $\beta = (\beta^{(1)}, \beta^{(2)}, \ldots, \beta^{(d)})'$, $\hat{\beta}_n = (\hat{\beta}_n^{(1)}, \hat{\beta}_n^{(2)}, \ldots, \hat{\beta}_n^{(d)})'$, and $\Gamma(Y_i, X_i, \Delta_i) = [\Gamma^{(1)}(Y_i, X_i, \Delta_i), \Gamma^{(2)}(Y_i, X_i, \Delta_i), \ldots, \Gamma^{(d)}(Y_i, X_i, \Delta_i)]'$. Here, β is a $(dJ) \times 1$ vector of all unknown parameters stacked in a way that the first J components are cause-specific coefficients of the first component of X, the second J components are cause-specific coefficients of the second component of X, and so on; $\hat{\beta}_n$ and $\Gamma(Y_i, X_i, \Delta_i)$ are ordered likewise.

The following theorem gives the main result of this section.

Theorem 2. *Let Assumptions 1-6 hold. Then for each $k = 1, \ldots, d$,*

$$
\hat{\beta}_n^{(k)} - \beta^{(k)} = n^{-1}\sum_{i=1}^{n}\left\{\Gamma^{(k)}(Y_i, X_i, \Delta_i) - E[\Gamma^{(k)}(Y, X, \Delta)]\right\} + o_p(n^{-1/2}).
$$

In particular,

$$
n^{1/2}(\hat{\beta}_n - \beta) \to_d \mathbf{N}(0, V),
$$

where $V = Var[\Gamma(Y, X, \Delta)]$.

The asymptotic variance V can be estimated consistently by replacing unknown quantities with sample analogs. Alternatively, it is expected that usual inference can be carried out by the nonparametric bootstrap. See, for example, Chen, Linton, and Van Keilegom (2003) for general results regarding the consistency of the nonparametric bootstrap applied to semiparametric estimators. Details are not worked out, however.

Appendix B: Proof of Theorem 2

To establish the asymptotic properties of the estimator, it is useful to apply a uniform result for degenerate U-processes of Sherman (1994a, b). To do so, let $Z_1, \ldots, Z_n$ be independent random vectors from a distribution P on a set $\mathcal{S}$. Let k be a positive integer and for each $n \geq 1$, let $f_n(\cdot, \theta)$ denote a real-valued function, indexed by $\theta \in \Theta$, on a product space $\mathcal{S}^k = \mathcal{S} \otimes \cdots \otimes \mathcal{S}$. As in Sherman (1994a,b), define a U-process of order k

$$U_n^k f_n = (n)_k^{-1} \sum_{\mathbf{i}_k} f_n(Z_{i_1}, \ldots, Z_{i_k}, \theta),$$

where $(n)_k = n(n-1)\cdots(n-k+1)$, and $\mathbf{i}_k = (i_1, \ldots, i_k)$ ranges over the $(n)_k$ ordered k-tuples of distinct integers from the set $\{1, \ldots, n\}$. The function $f_n(\cdot, \theta)$ is called a P-degenerate function if $E[f_n(s_1, \ldots, s_{i-1}, \cdot, s_{i+1}, \ldots, s_k, \theta)] \equiv 0$ for each $i = 1, \ldots, k$ and each $s^k \equiv (s_1, \ldots, s_k) \in \mathcal{S}^k$. The following lemma is an immediate consequence of the Main Corollary of Sherman (1994a, Section 6).

Lemma 1. *For $n \geq 1$, let $\mathcal{F}_n = \{f_n(\cdot, \theta) : \theta \in \Theta\}$ be a class of P-degenerate functions on $\mathcal{S}^k$, $k \geq 1$. If (a) $\mathcal{F}_n$ is Euclidean for an envelope F satisfying $E[F^2] < \infty$, (b) $E[\sup_{\mathcal{F}_n} f_n(\cdot, \theta)^2] = O(\delta_n^2)$, then*

$$\sup_{\mathcal{F}_n} |U_n^k f_n(\cdot, \theta)| = O_p\left(\delta_n^\alpha / n^{k/2}\right),$$

where $0 < \alpha < 1$.

Proof. As in Corollary 4 of Sherman (1994a) and Theorem 3 of Sherman (1994b), we apply the Main Corollary of Sherman (1994a, Section 6) with $p = 1$ to obtain

$$E\left[n^{k/2} \sup_{\mathcal{F}_n} |U_n^k f_n(\cdot, \theta)|\right] \leq \Lambda \left(E\left[\sup_{\mathcal{F}_n} f_n(\cdot, \theta)^2\right]\right)^{\alpha/2},$$

where Λ is a universal constant and $0 < \alpha < 1$. Then the lemma follows from the assumption (b) and Chebyshev's inequality. $\square$

Let $\tau_i = \tau(X_i)$. Define

$$\hat{D}_{n1}(t) = n^{-1} \sum_{i=1}^n \tau_i B_n(t, X_i) S_n^{(k)}(t, X_i)$$

and

$$\hat{D}_{n2}(t) = n^{-1} \sum_{i=1}^n \tau_i B_n(t, X_i) 1(Y_i > t) p_n^{(k)}(X_i).$$

Notice that the j-th components of $\hat{D}_{n1}(t)$ and $\hat{D}_{n2}(t)$ can be written as

$$\hat{D}_{n1}^{(j)}(t) = \frac{1}{n^3} \sum_{i=1}^n \sum_{l=1}^n \sum_{m=1}^n \tau_i 1(\Delta_l = j) K_{h_{ny}}(t - Y_l) K_{h_{nx}}(X_i - X_l) 1(Y_m > t) K_{h_{nx}}^{(k)}(X_i - X_m)$$

and

$$\hat{D}_{n2}^{(j)}(t) = \frac{1}{n^3} \sum_{i=1}^n \sum_{l=1}^n \sum_{m=1}^n \tau_i 1(\Delta_l = j) K_{h_{ny}}(t - Y_l) K_{h_{nx}}(X_i - X_l) 1(Y_i > t) K_{h_{nx}}^{(k)}(X_i - X_m).$$

Then $\hat{D}_{n1}^{(j)}(t)$ and $\hat{D}_{n2}^{(j)}(t)$ can be viewed as V-processes of order 3 in view of Von Mises (1947). Define $D_1^{(j)}(t) = E\left[\tau(X) B_j(t, X) S^{(k)}(t, X)\right]$ and $D_2^{(j)}(t) = E\left[\tau(X) B_j(t, X) 1(Y > t) p^{(k)}(X)\right]$. The following lemma gives an asymptotic expansion of $\tilde{D}_{n1}^{(j)}(t) - \tilde{D}_{n2}^{(j)}(t)$.

Lemma 2. *The following holds uniformly over $t \in S_T$: for each j,*

$$
\begin{aligned}
&[\tilde{D}_{n1}^{(j)}(t) - \tilde{D}_{n2}^{(j)}(t)] - [D_1^{(j)}(t) - D_2^{(j)}(t)] \\
&= n^{-1} \sum_{i=1}^{n} \tau_i B_j(t, X_i) S^{(k)}(t, X_i) - D_1^{(j)}(t) \\
&- n^{-1} \sum_{i=1}^{n} \tau_i B_j(t, X_i) 1(Y_i > t) p^{(k)}(X_i) + D_2^{(j)}(t) \\
&+ n^{-1} \sum_{i=1}^{n} \tau_i 1(\Delta_i = j) K_{h_{ny}}(t - Y_i) \left[S^{(k)}(t, X_i) p(X_i) - S(t, X_i) p^{(k)}(X_i) \right] \\
&- E\left[\tau(X) 1(\Delta = j) K_{h_{ny}}(t - Y) \left\{ S^{(k)}(t, X) p(X) - S(t, X) p^{(k)}(X) \right\} \right] \\
&- n^{-1} \sum_{i=1}^{n} \tau_i \left[1(Y_i > t) \partial [B_j(t, X_i) p(X_i)] / \partial x_k - \partial [B_j(t, X_i) S(t, X_i)] / \partial x_k \right] \\
&+ E\left[\tau(X) \left\{ 1(Y > t) \partial [B_j(t, X) p(X)] / \partial x_k - \partial [B_j(t, X) S(t, X)] / \partial x_k \right\} \right] \\
&+ o_p(n^{-1/2}).
\end{aligned}
$$

Proof. The proof consists of several steps. First, we will establish asymptotic equivalence between $\hat{D}_{n1}^{(j)}(t)$ and $\hat{D}_{n2}^{(j)}(t)$ and corresponding U-processes. Second, combining standard decomposition of U-processes with Lemma 1 gives leading terms in asymptotic expansions of $\hat{D}_{n1}^{(j)}(t)$ and $\hat{D}_{n2}^{(j)}(t)$. Third, the leading terms are approximated using standard arguments in kernel estimation.

Step 1. Asymptotic equivalence of V-processes and U-processes

Write $\hat{D}_{n1}^{(j)}(t)$ and $\hat{D}_{n2}^{(j)}(t)$ as

$$
\begin{aligned}
\hat{D}_{n1}^{(j)}(t) &= \tilde{D}_{n1}^{(j)}(t) + R_{n1}(t) \quad \text{and} \\
\hat{D}_{n2}^{(j)}(t) &= \tilde{D}_{n2}^{(j)}(t) + R_{n2}(t),
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{D}_{n1}^{(j)}(t) &= (n)_3^{-1} \sum_{i \neq l \neq m \neq i} \tau_i 1(\Delta_l = j) K_{h_{ny}}(t - Y_l) K_{h_{nx}}(X_i - X_l) 1(Y_m > t) K_{h_{nx}}^{(k)}(X_i - X_m), \\
\tilde{D}_{n2}^{(j)}(t) &= (n)_3^{-1} \sum_{i \neq l \neq m \neq i} \tau_i 1(\Delta_l = j) K_{h_{ny}}(t - Y_l) K_{h_{nx}}(X_i - X_l) 1(Y_i > t) K_{h_{nx}}^{(k)}(X_i - X_m),
\end{aligned}
$$

and $\sum_{i \neq l \neq m \neq i}$ refers to summation over distinct integers from the set $\{1, \ldots, n\}$. It is straightforward to show that $R_{n1}(t) = o_p(n^{-1/2})$ and $R_{n2}(t) = o_p(n^{-1/2})$ uniformly over t provided that $n^{1/2} h_{nx}^{d+1} \to \infty$ and $n h_{ny} h_{nx}^d / (\log n) \to \infty$. Then $\tilde{D}_{n2}^{(j)}(t)$ and $\tilde{D}_{n2}^{(j)}(t)$ are U-processes of order 3, indexed by (t, h_{ny}, h_{nx}).

Step 2. Decomposition of U-processes

Define

$$
\begin{aligned}
B_h^{(j)}(t,x) &= E\left[1(\Delta=j)K_{h_{ny}}(t-Y)\,K_{h_{nx}}(x-X)\right],\\
S_h^{(k)}(t,x) &= E\left[1(Y>t)K_{h_{nx}}^{(k)}(x-X)\right],\\
p_h^{(k)}(x) &= E\left[K_{h_{nx}}^{(k)}(x-X)\right],\\
D_{h1}^{(j)}(t) &= E\left[\tau(X)B_h^{(j)}(t,X)S_h^{(k)}(t,X)\right],\\
D_{h2}^{(j)}(t) &= E\left[\tau(X)B_h^{(j)}(t,X)1(Y>t)p_h^{(k)}(X)\right],\\
P_{h1}(t,x) &= E\left[\tau(X)K_{h_{nx}}(X-x)\,S_h^{(k)}(t,X)\right],\\
P_{h2}(t,x) &= E\left[\tau(X)B_h^{(j)}(t,X)K_{h_{nx}}^{(k)}(X-x)\right],\\
P_{h3}(t,x) &= E\left[\tau(X)K_{h_{nx}}(X-x)\,1(Y>t)p_h^{(k)}(X)\right], \quad \text{and}\\
P_{h4}(t,x) &= E\left[\tau(X)B_h^{(j)}(t,X)1(Y>t)K_{h_{nx}}^{(k)}(X-x)\right]
\end{aligned}
$$

Applying a standard decomposition of a U-process (see, for example, equation (6) of Sherman (1994a, p.449)) gives

$$
\begin{aligned}
&\tilde{D}_{n1}^{(j)}(t) - D_{h1}^{(j)}(t)\\
&= n^{-1}\sum_{i=1}^{n}\tau_i B_h^{(j)}(t,X_i)S_h^{(k)}(t,X_i) - D_{h1}^{(j)}(t)\\
&+ n^{-1}\sum_{i=1}^{n}1(\Delta_i=j)K_{h_{ny}}(t-Y_i)\,P_{h1}(t,X_i) - E\left[1(\Delta=j)K_{h_{ny}}(t-Y)\,P_{h1}(t,X)\right]\\
&+ n^{-1}\sum_{i=1}^{n}1(Y_i>t)P_{h2}(t,X_i) - E\left[1(Y>t)P_{h2}(t,X)\right]\\
&+ U_{n12}(t) + U_{n13}(t),
\end{aligned}
$$

and similarly,

$$
\begin{aligned}
&\tilde{D}_{n2}^{(j)}(t) - D_{h2}^{(j)}(t)\\
&= n^{-1}\sum_{i=1}^{n}\tau_i B_h^{(j)}(t,X_i)1(Y_i>t)p_h^{(k)}(X_i) - D_{h2}^{(j)}(t)\\
&+ n^{-1}\sum_{i=1}^{n}1(\Delta_i=j)K_{h_{ny}}(t-Y_i)\,P_{h3}(t,X_i) - E\left[1(\Delta=j)K_{h_{ny}}(t-Y)\,P_{h3}(t,X)\right]\\
&+ n^{-1}\sum_{i=1}^{n}P_{h4}(t,X_i) - E\left[P_{h4}(t,X)\right]\\
&+ U_{n22}(t) + U_{n23}(t),
\end{aligned}
$$

where $U_{nlk}(t)$ is a P-degenerate U-process of order k on $\mathcal{S}^k$ for each $l=1,2$ and $k=2,3$.

Using Lemma 1, we will first show that $U_{n12}(t)$, $U_{n13}(t)$, $U_{n22}(t)$, and $U_{n23}(t)$ are of order $o_p(n^{-1/2})$ uniformly over t. To do so, define $f_{n1}^{(j)}(t,h_{ny},h_{nx}) = h_{ny}h_{nx}^{2d+1}\tilde{D}_{n1}^{(j)}(t)$ and $f_{n2}^{(j)}(t,h_{ny},h_{nx}) =$

$h_{ny}h_{nx}^{2d+1}\tilde{D}_{n2}^{(j)}(t)$. Using results of Nolan and Pollard (1987) and Pakes and Pollard (1989), one can show that collections $\mathcal{F}_{n1} = \{f_{n1}^{(j)}(t, h_{ny}, h_{nx})\}$ and $\mathcal{F}_{n2} = \{f_{n2}^{(j)}(t, h_{ny}, h_{nx})\}$ indexed by (t, h_{ny}, h_{nx}) are Euclidean for constant envelopes. Furthermore, $E[\sup_{\mathcal{F}_{n1}} f_{n1}^{(j)}(t, h_{ny}, h_{nx})^2] = O(h_{ny}h_{nx}^{2d})$ and $E[\sup_{\mathcal{F}_{n2}} f_{n2}^{(j)}(t, h_{ny}, h_{nx})^2] = O(h_{ny}h_{nx}^{2d})$. Then applying Lemma 1 with $\delta_n = h_{ny}^{1/2}h_{nx}^{d}$ gives

$$\begin{aligned}
h_{ny}h_{nx}^{2d+1}U_{n1k}(t) &= O_p\left[(h_{ny}^{1/2}h_{nx}^{d})^{\alpha}/n^{k/2}\right] \quad \text{and} \\
h_{ny}h_{nx}^{2d+1}U_{n2k}(t) &= O_p\left[(h_{ny}^{1/2}h_{nx}^{d})^{\alpha}/n^{k/2}\right]
\end{aligned}$$

for $0 < \alpha < 1$ and $k = 2, 3$. By taking α sufficiently close to one,

$$U_{nlk}(t) = o_p(n^{-1/2})$$

for $l = 1, 2$ and $k = 2, 3$, provided that $nh_{ny}h_{nx}^{2+d} \to \infty$.

Step 3. Approximation of leading terms of U-processes

Using standard arguments in kernel estimation with Assumptions 5 and 6, one can show that

$$\begin{aligned}
B_h^{(j)}(t,x) &= B_j(t,x) + o_p(n^{-1/2}), \\
S_h^{(k)}(t,x) &= S^{(k)}(t,x) + o_p(n^{-1/2}), \\
p_h^{(k)}(x) &= p^{(k)}(x) + o_p(n^{-1/2}), \\
D_{h1}^{(j)}(t) &= D_1^{(j)}(t) + o_p(n^{-1/2}), \\
D_{h2}^{(j)}(t) &= D_2^{(j)}(t) + o_p(n^{-1/2}), \\
P_{h1}(t,x) &= \tau(x)S^{(k)}(t,x)p(x) + o_p(n^{-1/2}), \\
P_{h2}(t,x) &= -\tau(x)\partial[B_j(t,x)p(x)]/\partial x_k + o_p(n^{-1/2}), \\
P_{h3}(t,x) &= \tau(x)S(t,x)p^{(k)}(x) + o_p(n^{-1/2}), \quad \text{and} \\
P_{h4}(t,x) &= -\tau(x)\partial[B_j(t,x)S(t,x)]/\partial x_k + o_p(n^{-1/2})
\end{aligned}$$

uniformly over (t, x). Then we have

$$\begin{aligned}
&\tilde{D}_{n1}^{(j)}(t) - D_1^{(j)}(t) \\
&= n^{-1}\sum_{i=1}^{n} \tau_i B_j(t, X_i)S^{(k)}(t, X_i) - D_1^{(j)}(t) \\
&+ n^{-1}\sum_{i=1}^{n} \tau_i 1(\Delta_i = j)K_{h_{ny}}(t - Y_i)\, S^{(k)}(t, X_i)p(X_i) \\
&- E\left[\tau(X)1(\Delta = j)K_{h_{ny}}(t - Y)\, S^{(k)}(t, X)p(X)\right] \\
&- n^{-1}\sum_{i=1}^{n} \tau_i 1(Y_i > t)\partial[B_j(t, X_i)p(X_i)]/\partial x_k + E\left[\tau(X)1(Y > t)\partial[B_j(t, X)p(X)]/\partial x_k\right] \\
&+ o_p(n^{-1/2})
\end{aligned}$$

and

$$
\begin{aligned}
&\tilde{D}_{n2}^{(j)}(t) - D_2^{(j)}(t) \\
&= n^{-1}\sum_{i=1}^{n}\tau_i B_j(t,X_i)1(Y_i > t)p^{(k)}(X_i) - D_2^{(j)}(t) \\
&+ n^{-1}\sum_{i=1}^{n}\tau_i 1(\Delta_i = j)K_{h_{ny}}(t - Y_i)\, S(t,X_i)p^{(k)}(X_i) \\
&- E\left[\tau(X)1(\Delta = j)K_{h_{ny}}(t - Y)\, S(t,X)p^{(k)}(X)\right] \\
&- n^{-1}\sum_{i=1}^{n}\tau_i \partial[B_j(t,X_i)S(t,X_i)]/\partial x_k + E\left[\tau(X)\partial[B_j(t,X)S(t,X)]/\partial x_k\right] \\
&+ o_p(n^{-1/2})
\end{aligned}
$$

uniformly over t. Combining the results above gives the desired result. □

Lemma 3. *The following holds uniformly over $t \in S_T$:*

$$n^{-1}\sum_{i=1}^{n} B_n(t,X_i)B_n(t,X_i)' \to_p E[B(t,X)B(t,X)'].$$

Proof. The proof is omitted because it is straightforward to obtain this lemma using standard arguments in kernel estimation. □

Proof of Theorem 2. The theorem can be proved by Lemmas (2) and (3) and the definition of the estimator (8). □

Appendix C: Simulation Study

This section presents the results of a small simulation study that illustrates the finite-sample performance of the estimator. In each simulation, a sample was generated from the following model with $J = 2$:

$$
\begin{aligned}
H_1(T_1) &= X_1\beta_{11} + X_2\beta_{12} + U_1, \\
H_2(T_2) &= X_1\beta_{21} + X_2\beta_{22} + U_2,
\end{aligned}
$$

where both H_1 and H_2 are the natural log function, X_1 and X_2 were drawn from a bivariate normal distribution with zero means, variances of 5, and correlation of 0.3. U_1 and U_2 were generated, independently of X_1 and X_2, from a bivariate normal distribution with zero means, unit variances, and correlation of 0.8.

The true parameters are $(\beta_{11}, \beta_{12}, \beta_{21}, \beta_{22}) = (1, 1, 1, -1)$. The simulations used sample sizes of $n = 250$ and 1000, and all the simulations were carried out in GAUSS using GAUSS pseudo-random number generators. The number of replications was 100.

Table 1: Results of a Simulation study

Sample Size	β_{11}/β_{12} $C_h = 6$	$C_h = 8$	$C_h = 10$	β_{21}/β_{22} $C_h = 6$	$C_h = 8$	$C_h = 10$
$n = 250$						
Mean Bias	0.096	0.112	0.096	0.165	0.230	0.273
SD	0.298	0.256	0.255	0.273	0.252	0.265
RMSE	0.312	0.278	0.271	0.318	0.340	0.379
$n = 1000$						
Mean Bias	0.040	0.063	0.083	0.075	0.127	0.198
SD	0.167	0.142	0.128	0.120	0.118	0.119
RMSE	0.171	0.155	0.152	0.141	0.172	0.231

The kernel functions used in estimation are

$$K_Y(u) = (15/16)(1 - u^2)^2 1(|u| \leq 1)$$

and

$$K_X(u_1, u_2) = \prod_{k=1}^{2} (105/64)(1 - 5u_k^2 + 7u_k^4 - 3u_k^6) 1(|u_k| \leq 1).$$

These are second-order and fourth-order kernels. The asymptotic theory in Appendix A provides only qualitative restrictions on h_{ny} and h_{nx} in terms of asymptotic rates, so we use an *ad hoc* choice, namely $h_{ny} = C_h n^{-2/7}$ and $h_{nx} = C_h n^{-1/7}$. Given the choice of kernel functions, Assumption 6 is satisfied with this choice of bandwidths. We report simulation results below for three different cases with $C_h = 6, 8, 10$.

The trimming function $\tau(x_1, x_2)$ was $\tau(x_1, x_2) = \prod_{k=1} 1(|x_k| \leq 2s_x)$ with $s_x = \sqrt{5}$ (standard deviation of X_1 and X_2). The weight function $w_T(t)$ was $w_T(t) = \hat{p}_Y(t) 1(0.01 \leq t \leq 1.18)$, where $\hat{p}_Y(t)$ is the kernel density estimator of Y with a bandwidth $h_p = C_h n^{-1/5}$. The integral in (8) was evaluated numerically using the composite trapezoidal rule with 40 equal grid points.

Table 1 reports the mean bias, standard deviation (SD), and root mean squared error (RMSE) for β_{11}/β_{12} and β_{21}/β_{22}. It can be seen that the bias decreases as the sample size increases from $n = 250$ to $n = 1000$, although it is not negligible. The RMSE decreases about by a half, which is roughly consistent with the asymptotic theory in Appendix A. The simulation results did not change much as C_h varies. Overall, the simulation results indicate that the new estimator performs relatively well in finite samples.

References

Abbring, J.H. and G.J. Van den Berg (2003) The identifiability of the mixed proportional hazards competing risks model. *J. R. Statist. Soc. B*, 65, 701-710.

Abbring, J.H. and G.J. Van den Berg (2005) Social experiments and IV with duration outcomes, Working paper, Department of Economics, Free University, Amsterdam.

Bond, S.J. and J.E.H. Shaw (2003) Bounds on the covariate-time transformation for competing risks survival analysis, Working paper, Department of Statistics, University of Warwick, Coventry.

Chen, X., O. Linton, and I. Van Keilegom (2003) Estimation of semiparametric models when the criterion function is not smooth, *Econometrica* 71, 1591-1608.

Cheng, S.C., L.J. Wei, and Z. Ying (1995) Analysis of transformation models with censored data, *Biometrika*, 82, 835-845.

Clayton, D. and J. Cuzick (1985) Multivariate generalization of the proportional hazard model, *J. R. Statist. Soc. A*, 148, 82-117.

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

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

Fine, J.P. (1999) Analyzing competing risks data with transformation models, *J. R. Statist. Soc. B*, 61, 817-830.

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

B.E. Honoré and A. Lleras-Muney (2004) Bounds in competing risks models and the war on cancer, Working paper, Department of Economics, Princeton University, Princeton.

Horowitz, J.L. (1996) Semiparametric estimation of a regression model with an unknown transformation of the dependent variable, *Econometrica*, 64, 103-137.

Horowitz, J.L. (1998) *Semiparametric methods in econometrics*, New York: Springer-Verlag.

Horowitz, J.L., 1999, Semiparametric estimation of a proportional hazard model with unobserved heterogeneity,*Econometrica*, 67, 1001-1028.

Horowitz, J.L. (2001) Nonparametric estimation of a generalized additive model with an unknown link function, *Econometrica*, 69, 499-513.

Horowitz, J.L. and S. Lee (2004) Semiparametric estimation of a panel data proportional hazards model with fixed effects, *J. Econometrics*, 119, 155-198.

Ichimura, H. and L.-F.Lee (1991) Semiparametric least squares estimation of multiple index models: single equation estimation. In: Barnett W A, Powell J, and Tauchen G (eds) *Nonparametric and Semiparametric Methods in Econometrics and Statistics.* Cambridge University Press, Cambridge, pp. 3-49.

Nolan, D. and D. Pollard (1987) U-processes: rates of convergence, *Ann. Statist.* 15, 780-799.

Pakes, A. and D. Pollard (1989) Simulation and the asymptotics of optimization estimators, *Econometrica* 57, 1027-1057.

Sherman, R.P. (1994a) Maximal inequalities for degenerate U-processes with applications to optimization estimators, *Ann. Statist.* 22, 439-459.

Sherman, R.P. (1994b) U-processes in the analysis of a generalized semiparametric regression estimator, *Econometric Theory*, 10, 372-395.

Tsiatis, A. (1975) A nonidentifiability aspect of the problem of competing risks, *Proc. Nat. Acad. Sci.*, 72, 20-22.

Von Mises, R. (1947) On the asymptotic distribution of differentiable statistical functions, *Ann. Math. Statist.*, 18, 309-348.