

# SEMIPARAMETRIC ESTIMATION OF A PANEL DATA PROPORTIONAL HAZARDS MODEL WITH FIXED EFFECTS

---

*Joel Horowitz*  
*Sokbae Lee*

THE INSTITUTE FOR FISCAL STUDIES  
DEPARTMENT OF ECONOMICS, UCL  
**cemmap** working paper CWP21/02

# Semiparametric Estimation of a Panel Data Proportional Hazards Model with Fixed Effects

Joel L. Horowitz  
Department of Economics  
Northwestern University  
Evanston, IL 60208, USA  
joel-horowitz@northwestern.edu

and

Sokbae Lee\*  
Institute for Fiscal Studies  
and  
Department of Economics  
University College London  
London, WC1E 6BT, UK  
l.simon@ucl.ac.uk

April 2003

## Abstract

This paper considers a panel duration model that has a proportional hazards specification with fixed effects. The paper shows how to estimate the baseline and integrated baseline hazard functions without assuming that they belong to known, finite-dimensional families of functions. Existing estimators assume that the baseline hazard function belongs to a known parametric family. Therefore, the estimators presented here are more general than existing ones. This paper also presents a method for estimating the parametric part of the proportional hazards model with dependent right censoring, under which the partial likelihood estimator is inconsistent. The paper presents some Monte Carlo evidence on the small sample performance of the new estimators.

Keywords: Duration analysis, panel data, semiparametric estimation.

JEL Codes: C14, C23, C41

---

\*Corresponding author. Tel: +44-20-7679-5848. Fax: +44-20-7916-2775.

# Semiparametric Estimation of a Panel Data Proportional Hazards Model with Fixed Effects

## 1 Introduction

Much empirical research in economics is concerned with the analysis of duration data. In many applications multiple durations of a given individual are observed together with possible covariates. This paper is concerned with estimating a panel duration model that has a proportional hazards specification with unobserved heterogeneity. The model is formulated in terms of the hazard functions of successive positive random variables  $T_j$  (the durations of interest) conditional on  $d \times 1$  vectors of observed covariates  $X_j$  and an unobserved random variable  $U$  (the unobserved heterogeneity) for  $j = 1, \dots, J$ . The model is

$$\lambda_{T_j}(t_j|x_j, u) = \lambda_0(t_j) \exp(x_j' \beta + u), \quad (1)$$

where  $\lambda_{T_j}$  is the hazard of  $T_j = t_j$  conditional on  $X_j = x_j$  and  $U = u$ ,  $\lambda_0$  is the baseline hazard function, and  $\beta$  is a  $d \times 1$  vector of constant parameters. The random variable  $U$  represents unobserved, permanent attributes of individuals.  $T_1$  and  $T_2$  are assumed to be conditionally independent given  $X_1$ ,  $X_2$ , and  $U$ .<sup>1</sup> The observed covariates  $X_j$  are assumed to be constant within each spell but vary over spells, whereas the unobserved heterogeneity  $U$  is assumed to be constant over spells.<sup>2</sup> Covariates that are constant over spells are not included explicitly. They can be included in  $U$ , and their  $\beta$  coefficients are not identified.  $U$  may be arbitrarily correlated with  $X_j$  and, therefore, is a *fixed effect*. Unlike the random-effects approach, the fixed-effects approach does not require  $X_j$  and  $U$  to be statistically independent of one another or to have any other known statistical relationship.<sup>3</sup> It is assumed throughout most of the paper that  $J = 2$ . The extension to larger  $J$  is discussed briefly in Section 5.3.

---

<sup>1</sup>This requires that covariates be strictly exogenous. This weakness is a general problem of (nonlinear) fixed effects estimators.

<sup>2</sup>There could be another source of heterogeneity that varies over spells. For example, in work history data, there could be job-specific heterogeneity across workers, which varies over spells. In this paper, it is assumed implicitly that this kind of heterogeneity is observed and thus part of  $X_j$ .

<sup>3</sup>If the data are cross-sectional or single-spell, then the fixed-effects approach in this paper cannot be applied. See Horowitz (1999) for estimating the baseline and integrated baseline hazard functions nonparametrically in a cross-sectional proportional hazards model with random effects. Also, see Van der Berg (2001) for comparison between single-spell and multiple-spell models.

This paper presents methods for estimating  $\lambda_0(\cdot)$  and the integrated baseline hazard function  $\Lambda_0(\cdot) \equiv \int_0^\cdot \lambda_0(s)ds$  nonparametrically.<sup>4</sup> That is, this paper shows how to estimate  $\lambda_0$  and  $\Lambda_0$  without assuming that they belong to known, finite-dimensional families of functions. Several existing estimators assume that  $\lambda_0$  belongs to a parametric family. For example, Chamberlain (1985) considers a marginal likelihood approach for models with Weibull, gamma, and lognormal specifications. Ridder and Tunalı (1999) assume that  $\lambda_0$  is piecewise constant. This paper shows how to estimate  $\lambda_0$  and  $\Lambda_0$  nonparametrically when observations of  $T_j$  are uncensored and when they are right-censored.

This paper also considers estimation of  $\beta$  when observations of  $T_j$  are subject to right-censoring. An estimator of  $\beta$  based on a partial likelihood approach already exists for the uncensored and independently censored cases. See Chamberlain (1985), Kalbfleisch and Prentice (1980, 8.1.2), Lancaster (2000), and Ridder and Tunalı (1999) among others. The partial likelihood method cannot be applied to censored panel durations because the standard independent censoring assumption is likely to be violated. In many applications durations are observed over a fixed period. For example, in work history data, the duration of the most recent job of a respondent may be right-censored at the last interview date. Because of the fixed effect, the censoring threshold of  $T_j$  is not independent of  $T_j$  unless  $j = 1$ . Therefore,  $\beta$  cannot be estimated consistently by using the partial likelihood approach. This paper presents a consistent estimator of  $\beta$  under dependent censoring.

The estimation approach developed here consists of two steps. The first step is to express  $\lambda_0$ ,  $\Lambda_0$ , and  $\beta$  as functionals of the population distribution of  $(T_j, X_j)$  by utilizing an identification result of Honoré (1993). The second step is to construct suitable empirical analogs for the unknown population quantities that appear as arguments of these functionals, depending on whether or not observations of  $T_j$  are censored.

Let  $\lambda_{n0}$  and  $\Lambda_{n0}$ , respectively, denote nonparametric estimators of  $\lambda_0$  and  $\Lambda_0$ , where  $n$  is the sample size. It will be shown that  $\lambda_{n0}$  and  $\Lambda_{n0}$  are uniformly consistent, and  $n^{q/(2q+1)}(\lambda_{n0} - \lambda_0)$  and  $n^{1/2}(\Lambda_{n0} - \Lambda_0)$  are asymptotically normal, where  $q$  denotes the

---

<sup>4</sup>A recent working paper by Woutersen (2000) proposes a nonparametric estimator of  $\lambda_0$  for the case of independent censoring. Woutersen (2000) does not provide the asymptotic distribution of his estimator and does not consider estimation of  $\Lambda_0$ .

order of smoothness of  $\lambda_0$ .<sup>5</sup> It will also be shown that the new estimator  $\beta_n$  of  $\beta$  under dependent censoring is consistent, and  $n^{1/2}(\beta_n - \beta)$  is asymptotically normal.

The remainder of the paper is organized as follows. Section 2 provides an informal description of the estimators of  $\lambda_0$ ,  $\Lambda_0$ , and  $\beta$ . Section 3 presents the formal, asymptotic results for the uncensored case. Section 4 provides rule-of-thumb, data-driven methods for choosing bandwidths needed to estimate  $\lambda_0$  and  $\Lambda_0$  for the uncensored case. Extensions of the estimators of  $\lambda_0$  and  $\Lambda_0$  are discussed in Section 5. Section 6 presents the results of some Monte Carlo experiments that illustrate the finite-sample properties of the estimators. Concluding comments are given in Section 7. The proofs of theorems are in Appendix A. Appendix B presents the asymptotic properties of the estimators for the censored case.

## 2 Informal Description of the Estimators

### 2.1 The Uncensored Case

This section provides an informal description of our estimators of  $\lambda_0$  and  $\Lambda_0$  under the assumption that observations of  $T_j$  are uncensored and  $J = 2$ . In this case, an estimator of  $\beta$  is already available (see Section 1).<sup>6</sup> Let  $b_n$  denote the estimator of  $\beta$ .

The estimation approach developed here is based on an identification result of Honoré (1993). When the model (1) is identified,  $\lambda_0$  and  $\Lambda_0$  can be expressed as functionals of the population distribution of  $(T_1, T_2, X_1, X_2)$ . Then estimators of  $\lambda_0$  and  $\Lambda_0$  can be obtained by replacing unknown population quantities with their empirical analogs.

To identify  $\lambda_0$  and  $\Lambda_0$ , observe first that  $T_j$  depends on  $X_j$  only through the index  $Z_j \equiv X_j' \beta$  for  $j = 1, 2$ . Assume conditional on  $(Z_1, Z_2, U)$ ,  $T_1$  and  $T_2$  are independent.

---

<sup>5</sup>The nonparametric estimator of  $\Lambda_0$  can be used to construct a specification test of the model (1). Since  $\Lambda_{n0}$  converges in probability faster than  $\lambda_{n0}$ , a test based on  $\Lambda_{n0}$  would be more powerful than a test based on  $\lambda_{n0}$ . The details of the test are beyond the scope of the paper. Roughly speaking, the specification test consists of testing the distribution of  $\log \Lambda_0(T_1) - \log \Lambda_0(T_2) + (X_1 - X_2)\beta$ , which is distributed as the logistic distribution and independent of  $X_1$  and  $X_2$  under the null hypothesis that the model (1) is correct.

<sup>6</sup>For example, one may use the estimator of Chamberlain (1985). This estimator is based on the fact that the probability of one spell being larger than the other spell, conditional on covariates, is independent of the fixed effects and can be expressed as a logit model.

Then the joint conditional survivor function of  $T_1$  and  $T_2$  is

$$\begin{aligned} S(t_1, t_2 | z_1, z_2) &\equiv \Pr(T_1 > t_1, T_2 > t_2 | Z_1 = z_1, Z_2 = z_2) \\ &= \int \exp[-\Lambda_0(t_1)e^{z_1+u} - \Lambda_0(t_2)e^{z_2+u}] dP_{u|z_1, z_2}, \end{aligned}$$

where  $P_{u|z_1, z_2}$  denotes the distribution of  $U$  conditional on  $(Z_1, Z_2) = (z_1, z_2)$ . By differentiation of  $S(t_1, t_2 | z_1, z_2)$ ,

$$\frac{\partial S(t_1, t_2 | z_1, z_2) / \partial t_1}{\partial S(t_1, t_2 | z_1, z_2) / \partial t_2} = \frac{\lambda_0(t_1)}{\lambda_0(t_2)} \exp(z_1 - z_2). \quad (2)$$

A scale normalization is needed to make identification possible. This is accomplished here by assuming that

$$\int_{S_T} \frac{w_t(t)}{\lambda_0(t)} dt = 1,$$

where  $w_t$  is a scalar-valued function with compact support  $S_T$  that satisfies  $\int_{S_T} w_t(t) dt = 1$  and other conditions in Section 3. This scale normalization is useful for the estimators developed here, as will be seen below.

Let  $R(t_1, t_2 | z_1, z_2)$  denote the left-hand side of (2). Under the scale normalization, (2) implies that  $\lambda_0$  has the form

$$\lambda_0(t) = \int_{S_T} w_t(t_2) \exp(z_2 - z_1) R(t, t_2 | z_1, z_2) dt_2$$

for every  $(z_1, z_2)$ . Let  $w_z(\cdot)$  be a scalar-valued function with compact support  $S_Z$  that satisfies  $\int_{S_Z} w_z(z) dz = 1$  and other conditions in Section 3. Also, let  $w(t_2, z_1, z_2) = w_t(t_2)w_z(z_1)w_z(z_2)$ . Then

$$\lambda_0(t) = \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 w(t_2, z_1, z_2) \exp(z_2 - z_1) R(t, t_2 | z_1, z_2). \quad (3)$$

Equation (3) identifies  $\lambda_0$  and is the basis for the estimators of  $\lambda_0$  and  $\Lambda_0$  proposed here.<sup>7</sup>

This completes the first step of our estimation strategy.

---

<sup>7</sup>Observe that  $\lambda_0$  can also be written as

$$\lambda_0(t) = \int_{S_T} dt_1 \int_{S_Z} dz_1 \int_{S_Z} dz_2 w(t_1, z_1, z_2) \exp(z_1 - z_2) R(t_1, t | z_1, z_2)^{-1}. \quad (4)$$

This equation can be the basis for another estimator of  $\lambda_0$ . One can use the arguments of Appendix A to establish asymptotic results for an estimator based on (4). Hence, we just focus on the estimator of  $\lambda_0$  based on (3). Also, one can use a linear combination of these estimators. This will be discussed in detail in Section 5.

In the second step, estimators of  $\lambda_0$  and  $\Lambda_0$  are obtained by replacing the unknown function  $R(t_1, t_2|z_1, z_2)$  in (3) with a uniformly consistent estimator  $R_n(t_1, t_2|z_1, z_2)$ . The resulting estimators of  $\lambda_0$  and  $\Lambda_0$  are

$$\lambda_{n0}(t) = \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 w(t_2, z_1, z_2) \exp(z_2 - z_1) R_n(t, t_2|z_1, z_2) \quad (5)$$

and

$$\Lambda_{n0}(t) = \int_0^t \lambda_{n0}(t_1) dt_1. \quad (6)$$

Section 3 gives conditions under which  $\lambda_{n0}$  and  $\Lambda_{n0}$  are uniformly consistent, and  $n^{q/(2q+1)}(\lambda_{n0} - \lambda_0)$  and  $n^{1/2}(\Lambda_{n0} - \Lambda_0)$  are asymptotically normal, where  $q$  denotes the order of smoothness of  $\lambda_0$ . Intuitively, the rates  $n^{-q/(2q+1)}$  and  $n^{-1/2}$  are possible because integration over  $(t_2, z_1, z_2)$  or  $(t_1, t_2, z_1, z_2)$  in (5)-(6) creates averaging effects that mitigate the curse of dimensionality. Similar averaging effects occur estimation of single index models (e.g., Horowitz and Härdle (1996), Powell, Stock, and Stoker (1989)), partially linear models (e.g., Robinson (1988)), additive models (e.g., Horowitz (2001), Linton and Härdle (1996)), and transformation models (e.g., Horowitz (1996), Horowitz and Gørgens (1999)).

In this paper,  $R$  is estimated with kernels. To describe the estimator, let  $p_{t|z}(t_1, t_2|z_1, z_2)$  denote the probability density function of  $T_1$  and  $T_2$  conditional on  $Z_1 = z_1$  and  $Z_2 = z_2$ . Write

$$R(t_1, t_2|z_1, z_2) = \frac{\int_{t_2}^{\infty} p_{t|z}(t_1, s_2|z_1, z_2) ds_2}{\int_{t_1}^{\infty} p_{t|z}(s_1, t_2|z_1, z_2) ds_1} \equiv \frac{A(t_1, t_2|z_1, z_2)}{B(t_1, t_2|z_1, z_2)}. \quad (7)$$

Let  $\{T_{i1}, T_{i2}, X_{i1}, X_{i2}\}_{i=1}^n$  denote a random sample of  $(T_1, T_2, X_1, X_2)$  in (1). Define  $Z_{ni1} = X'_{i1} b_n$  and  $Z_{ni2} = X'_{i2} b_n$ . Since  $\beta$  is unknown (and therefore,  $Z_{i1}$  and  $Z_{i2}$  are unknown), the estimator is based on  $\{T_{i1}, T_{i2}, Z_{ni1}, Z_{ni2}\}_{i=1}^n$ . Let  $K_T$  and  $K_Z$  be kernel functions of scalar arguments, and let  $\{h_{n1}\}, \{h_{n2}\}$ , and  $\{h_{nz}\}$  ( $n = 1, 2, \dots$ ) be sequences of bandwidths that converge to zero as  $n \rightarrow \infty$ . Conditions that  $K_T$ ,  $K_Z$ ,  $h_{n1}$ ,  $h_{n2}$ , and  $h_{nz}$  need to satisfy are given in Section 3. Let  $p_z(z_1, z_2)$  denote the probability density function of  $Z_1$  and  $Z_2$ . Estimate  $p_z(z_1, z_2)$  by

$$p_{nz}(z_1, z_2) = (nh_{nz}^2)^{-1} \sum_{i=1}^n K_Z \left( \frac{z_1 - Z_{ni1}}{h_{nz}} \right) K_Z \left( \frac{z_2 - Z_{ni2}}{h_{nz}} \right).$$

Let  $1(\cdot)$  be the indicator function. Define

$$A_n(t_1, t_2 | z_1, z_2) = [nh_{n1}h_{nz}^2 p_{nz}(z_1, z_2)]^{-1} \sum_{i=1}^n 1(T_{i2} > t_2) K_T \left( \frac{t_1 - T_{i1}}{h_{n1}} \right) \\ \times K_Z \left( \frac{z_1 - Z_{ni1}}{h_{nz}} \right) K_Z \left( \frac{z_2 - Z_{ni2}}{h_{nz}} \right)$$

and

$$B_n(t_1, t_2 | z_1, z_2) = [nh_{n2}h_{nz}^2 p_{nz}(z_1, z_2)]^{-1} \sum_{i=1}^n 1(T_{i1} > t_1) K_T \left( \frac{t_2 - T_{i2}}{h_{n2}} \right) \\ \times K_Z \left( \frac{z_1 - Z_{ni1}}{h_{nz}} \right) K_Z \left( \frac{z_2 - Z_{ni2}}{h_{nz}} \right).$$

The estimator of  $R(t_1, t_2 | z_1, z_2)$  is obtained by

$$R_n(t_1, t_2 | z_1, z_2) = A_n(t_1, t_2 | z_1, z_2) / B_n(t_1, t_2 | z_1, z_2). \quad (8)$$

A higher-order kernel is needed for  $K_Z$  to insure that certain bias and remainder terms in the asymptotic expansions of  $n^{q/(2q+1)}(\lambda_{n0} - \lambda_0)$  and  $n^{1/2}(\Lambda_{n0} - \Lambda_0)$  vanish as  $n \rightarrow \infty$ . For estimation of  $\lambda_0(t)$ , it is advisable to let  $h_{n2}$  converge to zero faster than  $h_{n1}$  to reduce bias. For estimation of  $\Lambda_0(t)$ , it is necessary to have both  $h_{n1}$  and  $h_{n2}$  converge to zero faster than  $n^{-1/(2q+1)}$ , which is the asymptotically optimal rate for  $\lambda_{n0}(t)$ , to prevent the asymptotic distribution of  $n^{1/2}(\Lambda_{n0} - \Lambda_0)$  from having a non-zero mean.

## 2.2 The Censored Case

This section provides informal descriptions of estimators of  $\beta$ ,  $\lambda_0$ , and  $\Lambda_0$  when  $T_1$  and  $T_2$  are subject to dependent right censoring. There are many possible censoring mechanisms for  $T_1$  and  $T_2$ . In this section, we focus on a pure renewal process in the sense that  $T_1$  and  $T_2$  are the same type of durations and there is no time spent on other states.

We assume that the successive durations,  $T_1$  and  $T_2$ , are observed over a time period of length  $C$ , where  $C$  is random with an unknown probability distribution. It is assumed that  $C$  is observed for every individual and that  $C$  is independent of  $T_1$  and  $T_2$  given  $X_1$  and  $X_2$ .<sup>8</sup> The censoring mechanism here governs the sum of  $T_1$  and  $T_2$ , rather than each

---

<sup>8</sup>This assumption seems reasonable for pure renewal processes, for example, car insurance claim durations analyzed in Abbring, Chiappori, and Pinquet (2003).



separately. In this case, one observes not  $T_j$  but  $Y_j \equiv \min(T_j, C_j)$ , where  $C_1 = C$  and  $C_2 = (C - T_1)1(T_1 \leq C)$ .<sup>9</sup> Observe that  $C_2$  depends on  $T_1$ , and, therefore, on  $T_2$  because of the fixed effect. Hence, the censoring mechanism here violates the standard independence assumption, under which  $C_j$  is independent of  $T_j$  given  $X_j$  for  $j = 1, 2$ .<sup>10</sup> Define indicator variables by  $\Delta_j = 1(T_j \leq C_j)$  for  $j = 1, 2$ . An observed random sample now consists of  $\{(Y_{i1}, Y_{i2}, X_{i1}, X_{i2}, \Delta_{i1}, \Delta_{i2}, C_i) : i = 1, \dots, n\}$ .

### 2.2.1 Estimating $\beta$

This subsection shows how to estimate  $\beta$  under dependent right censoring. As was discussed in Section 1,  $\beta$  cannot be estimated consistently by using the partial likelihood approach. This is because  $\Pr(Y_1 < Y_2 | X_1, X_2, U, \min(T_1, T_2) < \min(C_1, C_2))$  is now dependent on the fixed effect. An approach based on (2), however, can be used to obtain a consistent estimator of  $\beta$ . Abusing notation a bit, let  $S(t_1, t_2 | x_1, x_2) = \Pr(T_1 > t_1, T_2 > t_2 | X_1 = x_1, X_2 = x_2)$ . As in (2),

$$\frac{\partial S(t, t | x_1, x_2) / \partial t_1}{\partial S(t, t | x_1, x_2) / \partial t_2} = \exp[(x_1 - x_2)' \beta] \quad (9)$$

by setting  $t_1 = t_2 = t$ . Let  $R_\beta(t | x_1, x_2)$  denote the left-hand side of (9). Since (9) holds for any  $t$ , write

$$\int_{S_\beta} w_\beta(t) R_\beta(t | x_1, x_2) dt = \exp[(x_1 - x_2)' \beta], \quad (10)$$

where  $w_\beta(\cdot)$  is a scalar-valued function with compact support  $S_\beta$  that satisfies  $\int_{S_\beta} w_\beta(t) dt = 1$  and other conditions in Appendix B.1. This yields

$$\beta = [E(X_1 - X_2)(X_1 - X_2)']^{-1} E \left[ (X_1 - X_2) \log \left( \int_{S_\beta} w_\beta(t) R_\beta(t | X_1, X_2) dt \right) \right] \quad (11)$$

provided that  $E(X_1 - X_2)(X_1 - X_2)'$  is nonsingular. Define  $V = \int_{S_\beta} w_\beta(t) R_\beta(t | X_1, X_2) dt$  and  $\Delta X = X_1 - X_2$ . Equation (11) suggests that  $\beta$  can be estimated by a no-intercept OLS regression of a sample analog of  $\log V$  on  $\Delta X$ .

<sup>9</sup>With minor modifications, arguments in this section apply to standard censoring mechanisms where  $C_j$  is conditionally independent of  $T_j$  given  $X_j$  for  $j = 1, 2$ . We are grateful to an anonymous referee who raised this issue. Under the standard censoring mechanism,  $\beta$  can be estimated by the partial likelihood approach as well.

<sup>10</sup>Lin, Sun, and Ying (1999), Visser (1996), and Wang and Wells (1998) have considered estimation of the joint survivor (or distribution) function of  $T_1$  and  $T_2$  (without covariates) under the same type of dependent censoring.

Carrying out this OLS regression requires an estimator of  $R_\beta(t|x_1, x_2)$ . There may be several methods for estimating  $R_\beta(t|x_1, x_2)$  under dependent right censoring, but we present here a simple estimator based on Burke (1988) and Wang and Wells (1998). An alternative estimator of  $R_\beta(t|x_1, x_2)$  will be described briefly in Appendix B.3.

Define the joint conditional sub-distribution function  $F(t_1, t_2|x_1, x_2) = \Pr(Y_1 \leq t_1, Y_2 \leq t_2, \Delta_1 = \Delta_2 = 1|X_1 = x_1, X_2 = x_2)$  and its density  $f(t_1, t_2|x_1, x_2) = \partial^2 F(t_1, t_2|x_1, x_2)/\partial t_1 \partial t_2$ . Also, let  $G(c|x_1, x_2) = \Pr(C > c|X_1 = x_1, X_2 = x_2)$  denote the survivor function of  $C$  conditional on  $X_1 = x_1$  and  $X_2 = x_2$ . As in equation (7) of Wang and Wells (1998), observe that

$$S(t_1, t_2|x_1, x_2) = \int_{t_1}^{\infty} \int_{t_2}^{\infty} \frac{f(s_1, s_2|x_1, x_2)}{G(s_1 + s_2|x_1, x_2)} ds_1 ds_2. \quad (12)$$

Therefore,  $R_\beta(t|x_1, x_2)$  can be written as

$$R_\beta(t|x_1, x_2) = \frac{\int_t^{\infty} f(t, s_2|x_1, x_2)/G(t + s_2|x_1, x_2) ds_2}{\int_t^{\infty} f(s_1, t|x_1, x_2)/G(s_1 + t|x_1, x_2) ds_1} \equiv \frac{\tilde{A}_\beta(t|x_1, x_2)}{\tilde{B}_\beta(t|x_1, x_2)}. \quad (13)$$

The right-hand side of (13) can be estimated with kernels. For simplicity, assume that the distribution of  $X_1$  and  $X_2$  is absolutely continuous with respect to Lebesgue measure on  $R^{2d}$ . It is straightforward to include discrete covariates. Let  $K_X$  be a kernel function of  $d$ -dimensional arguments,  $\{h_{nx}\}$  ( $n = 1, 2, \dots$ ) be a sequence of bandwidths that converge to zero as  $n \rightarrow \infty$ , and  $p_x(x_1, x_2)$  denote the probability density function of  $X_1$  and  $X_2$ .

Let  $p_{nx}(x_1, x_2)$  and  $G_n(c|x_1, x_2)$  denote the kernel estimators of  $p_x(x_1, x_2)$  and  $G(c|x_1, x_2)$ , that is

$$p_{nx}(x_1, x_2) = \left(nh_{nx}^{2d}\right)^{-1} \sum_{i=1}^n K_X \left(\frac{x_1 - X_{i1}}{h_{nx}}\right) K_X \left(\frac{x_2 - X_{i2}}{h_{nx}}\right)$$

and

$$G_n(c|x_1, x_2) = \left[nh_{nx}^{2d}p_{nx}(x_1, x_2)\right]^{-1} \sum_{i=1}^n 1(C_i > c) K_X \left(\frac{x_1 - X_{i1}}{h_{nx}}\right) K_X \left(\frac{x_2 - X_{i2}}{h_{nx}}\right).$$

Define

$$\begin{aligned} \tilde{A}_{n\beta}(t|x_1, x_2) &= \left[nh_{n1}h_{nx}^{2d}p_{nx}(x_1, x_2)\right]^{-1} \sum_{i=1}^n \frac{\Delta_{i1}\Delta_{i2}1(Y_{i2} > t)}{G_n(Y_{i1} + Y_{i2}|X_{i1}, X_{i2})} K_T \left(\frac{t - Y_{i1}}{h_{n1}}\right) \\ &\quad \times K_X \left(\frac{x_1 - X_{i1}}{h_{nx}}\right) K_X \left(\frac{x_2 - X_{i2}}{h_{nx}}\right) \end{aligned}$$

and

$$\begin{aligned} \tilde{B}_{n\beta}(t|x_1, x_2) &= \left[ nh_{n2}h_{nx}^{2d}p_{nx}(x_1, x_2) \right]^{-1} \sum_{i=1}^n \frac{\Delta_{i1}\Delta_{i2}1(Y_{i1} > t)}{G_n(Y_{i1} + Y_{i2}|X_{i1}, X_{i2})} K_T \left( \frac{t - Y_{i2}}{h_{n2}} \right) \\ &\quad \times K_X \left( \frac{x_1 - X_{i1}}{h_{nx}} \right) K_X \left( \frac{x_2 - X_{i2}}{h_{nx}} \right). \end{aligned}$$

The estimator of  $R_\beta(t|x_1, x_2)$  can be obtained by

$$\tilde{R}_{n\beta}(t|x_1, x_2) = \tilde{A}_{n\beta}(t|x_1, x_2) / \tilde{B}_{n\beta}(t|x_1, x_2). \quad (14)$$

Observe that  $\tilde{R}_{n\beta}(t|x_1, x_2)$  only uses uncensored data ( $\Delta_{i1} = \Delta_{i2} = 1$ ) and is weighted by the inverse of  $G_n$  to take into account the effect of censoring.

Let  $w_x(\cdot)$  be a scalar-valued function with compact support  $S_X$  that satisfies conditions in Appendix B.1. Then the OLS estimator  $\beta_n$  of  $\beta$  is

$$\beta_n = \left( n^{-1} \sum_{i=1}^n w_{xi} \Delta X_i \Delta X_i' \right)^{-1} \left( n^{-1} \sum_{i=1}^n w_{xi} \Delta X_i \log V_{ni} \right), \quad (15)$$

where  $w_{xi} = w_x(X_{i1})w_x(X_{i2})$ ,  $\Delta X_i = X_{i1} - X_{i2}$  and  $V_{ni} = \int_{S_\beta} w_\beta(t) R_{n\beta}(t|X_{i1}, X_{i2}) dt$ . The weight function  $w_x$  is introduced here to estimate  $\beta$  without being overly influenced by the tail behavior of the distributions of  $X_1$  and  $X_2$ .

### 2.2.2 Estimating $\lambda_0$ and $\Lambda_0$

In this subsection, we present modified versions of the estimators of  $\lambda_0$  and  $\Lambda_0$  described in Section 2.1. Observe that (3) holds for the latent variables  $T_1$  and  $T_2$ . Therefore,  $\lambda_0$  and  $\Lambda_0$  can be estimated by using (5) and (6) if a consistent estimator of  $R(t_1, t_2|z_1, z_2)$  is available.

For simplicity, it is assumed in this subsection that  $C$  is independent of  $(T_1, T_2, X_1, X_2)$ . Abusing notation a bit, define  $F(t_1, t_2|z_1, z_2) = \Pr(Y_1 \leq t_1, Y_2 \leq t_2, \Delta_1 = \Delta_2 = 1|Z_1 = z_1, Z_2 = z_2)$ ,  $f(t_1, t_2|z_1, z_2) = \partial^2 F(t_1, t_2|z_1, z_2) / \partial t_1 \partial t_2$ , and  $G(c) = \Pr(C > c)$ . As in Section 2.2.1,  $R(t_1, t_2|z_1, z_2)$  can be written as

$$R(t_1, t_2|z_1, z_2) = \frac{\int_{t_2}^{\infty} f(t_1, s_2|z_1, z_2) / G(t_1 + s_2) ds_2}{\int_{t_1}^{\infty} f(s_1, t_2|z_1, z_2) / G(s_1 + t_2) ds_1} \equiv \frac{\tilde{A}(t_1, t_2|z_1, z_2)}{\tilde{B}(t_1, t_2|z_1, z_2)}. \quad (16)$$

Again the right-hand side of (16) can be estimated with kernels. Estimate  $G(\cdot)$  by the empirical survivor function<sup>11</sup>

$$G_n(c) = n^{-1} \sum_{i=1}^n 1(C_i > c).$$

Define

$$\begin{aligned} \tilde{A}_n(t_1, t_2 | z_1, z_2) &= [nh_{n1}h_{nz}^2 p_{nz}(z_1, z_2)]^{-1} \sum_{i=1}^n \frac{\Delta_{i1}\Delta_{i2}1(Y_{i2} > t_2)}{G_n(Y_{i1} + Y_{i2})} K_T\left(\frac{t_1 - Y_{i1}}{h_{n1}}\right) \\ &\quad \times K_Z\left(\frac{z_1 - Z_{ni1}}{h_{nz}}\right) K_Z\left(\frac{z_2 - Z_{ni2}}{h_{nz}}\right) \end{aligned}$$

and

$$\begin{aligned} \tilde{B}_n(t_1, t_2 | z_1, z_2) &= [nh_{n2}h_{nz}^2 p_{nz}(z_1, z_2)]^{-1} \sum_{i=1}^n \frac{\Delta_{i1}\Delta_{i2}1(Y_{i1} > t_1)}{G_n(Y_{i1} + Y_{i2})} K_T\left(\frac{t_2 - Y_{i2}}{h_{n2}}\right) \\ &\quad \times K_Z\left(\frac{z_1 - Z_{ni1}}{h_{nz}}\right) K_Z\left(\frac{z_2 - Z_{ni2}}{h_{nz}}\right). \end{aligned}$$

The estimator of  $R(t_1, t_2 | z_1, z_2)$  is obtained by

$$\tilde{R}_n(t_1, t_2 | z_1, z_2) = \tilde{A}_n(t_1, t_2 | z_1, z_2) / \tilde{B}_n(t_1, t_2 | z_1, z_2). \quad (17)$$

### 3 Asymptotic Properties of the Estimators

This section establishes the asymptotic properties of  $\lambda_{n0}$  and  $\Lambda_{n0}$  proposed in Section 2.1 under the assumption that complete spells of  $T_1$  and  $T_2$  are available. Appendix B.1 gives conditions under which  $n^{1/2}(\beta_n - \beta)$  is asymptotically normal, and Appendix B.2 presents the asymptotic properties of  $\lambda_{n0}$  and  $\Lambda_{n0}$  for the censored case.

We make the following assumptions:

**Assumption 3.1 (Random Sampling).**  $\{T_{i1}, T_{i2}, X_{i1}, X_{i2} : i = 1, \dots, n\}$  is a random sample of  $(T_1, T_2, X_1, X_2)$  in (1).

**Assumption 3.2 (Conditional Independence).**  $T_1$  and  $T_2$  are conditionally independent given  $X_1, X_2$ , and  $U$ .

---

<sup>11</sup>If only  $\min(C, T_1 + T_2)$  is observed, then the Kaplan-Meier estimator of  $G$  can be used.

Assumption 3.2 is used to identify  $\lambda_0$  and  $\Lambda_0$ . It precludes the possibility of lagged duration dependence, which is not treated in this paper.<sup>12</sup>

**Assumption 3.3 (Normalization).**  $\int_0^\infty [w_t(t)/\lambda_0(t)] dt = 1$ .

As was explained in Section 2.1, Assumption 3.3 is useful to create averaging effects. The same type of scale normalization is used for a similar reason in Horowitz (2001).

**Assumption 3.4 (Covariates).**  $X_1$  and  $X_2$  have bounded support.<sup>13</sup>

Let  $p(t_1, t_2, z_1, z_2)$  denote the probability density function of  $(T_1, T_2, Z_1, Z_2)$ . In what follows,  $q \geq 2$  and  $r$  are integers such that  $r \geq 4$  for  $\lambda_{n0}$  and  $r \geq 6$  for  $\Lambda_{n0}$ .

**Assumption 3.5 (Smoothness).** *The distribution of  $(T_1, T_2, Z_1, Z_2)$  is absolutely continuous with respect to Lebesgue measure on  $\mathbf{R}^4$ . Furthermore, there are intervals of the real line,  $I_T$  and  $I_Z$ , such that*

- (a)  $I_T = [0, \tau_T)$ , where  $\tau_T \leq \infty$ , and  $I_Z$  is open,
- (b)  $p(t_1, t_2, z_1, z_2)$  is bounded on  $I_T \times I_T \times I_Z \times I_Z$ ,
- (c)  $p(t_1, t_2, z_1, z_2)$  is positive for all  $(t_1, t_2, z_1, z_2) \in \text{int}(I_T \times I_T \times I_Z \times I_Z)$ , and
- (d)  $p(t_1, t_2, z_1, z_2)$  has bounded partial derivatives up to order  $q$  with respect to  $t_j$  and up to order  $r$  with respect to  $z_j$  for  $j = 1, 2$ .

In view of (2) and (7), condition (c) is equivalent to the condition that  $\lambda_0(t) > 0$  for all  $t \in \text{int}(I_T)$  and condition (d) implies that  $\lambda_0$  is  $q$ -times differentiable. Assumption 3.5 also implies that the distribution of  $(Z_1, Z_2)$  is absolutely continuous with respect to Lebesgue measure on  $\mathbf{R}^2$  and  $p_z(z_1, z_2)$  is positive in the interior of the support of the distribution.<sup>14</sup>

**Assumption 3.6 (Weight Functions).** (a) *The weight function  $w_t(\cdot)$  is a bounded, non-negative function with compact support  $S_T \subset I_T$  such that  $\int_{S_T} w_t(t) dt = 1$  and  $w_t$  is  $q$  times continuously differentiable on  $S_T$ .*

---

<sup>12</sup>Honoré (1993) achieves identification of the lagged duration model through an analytic continuation. The resulting identifying relation is very different from (3), and the estimation approach developed here is not applicable to it.

<sup>13</sup>Assumption 3.4 can be relaxed at the expense of more complicated proofs.

<sup>14</sup>Assumption 3.5 is not satisfied if all of the covariates are discrete. However, in that case, the estimators of  $\lambda_0$  and  $\Lambda_0$  can be easily modified and, in fact, are simpler than the estimators presented in Section 2.1.

(b) The weight function  $w_z(\cdot)$  is a bounded, non-negative function with compact support  $S_Z \subset I_Z$  such that  $\int_{S_Z} w_z(z) dz = 1$  and  $w_z$  is  $r$  times continuously differentiable on  $S_Z$ .

**Assumption 3.7 (Estimator of  $\beta$ ).** There is a  $d \times 1$ -vector-valued function  $\Omega(t_1, t_2, x_1, x_2)$  such that

(a)  $E\Omega(T_1, T_2, X_1, X_2) = 0$ ,

(b) the components of  $E[\Omega(T_1, T_2, X_1, X_2)\Omega(T_1, T_2, X_1, X_2)']$  are finite, and

(c) as  $n \rightarrow \infty$ ,

$$b_n - \beta = n^{-1} \sum_{i=1}^n \Omega(T_{i1}, T_{i2}, X_{i1}, X_{i2}) + o_p(n^{-1/2}).$$

Assumption 3.7 is satisfied by the partial likelihood estimator of  $\beta$  mentioned in Section 1.

**Assumption 3.8 (Kernels).** (a)  $K_T$  has support  $[-1, 1]$ , is bounded and symmetrical about 0, has bounded variation, and satisfies

$$\int_{-1}^1 u^j K_T(u) du = \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{if } j = 1 \leq j \leq q - 1, \\ C_T & \text{if } j = q, \end{cases}$$

where  $C_T$  is a positive constant.

(b)  $K_Z$  has support  $[-1, 1]$ , is bounded and symmetrical about 0, has bounded variation, and satisfies

$$\int_{-1}^1 u^j K_Z(u) du = \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{if } j = 1 \leq j \leq r - 1, \\ C_Z & \text{if } j = r, \end{cases}$$

where  $C_Z$  is a positive constant.

(c)  $K_Z$  is everywhere differentiable.  $K'_Z(v) \equiv dK_Z(v)/dv$  is bounded and Lipschitz continuous and has bounded variation.

Assumption 3.8 requires  $K_Z$  to be a higher-order kernel. A higher-order kernel is used to insure that certain bias and remainder terms in the asymptotic expansions of  $n^{q/(2q+1)}(\lambda_{n0} - \lambda_0)$  and  $n^{1/2}(\Lambda_{n0} - \Lambda_0)$  are negligibly small.

**Assumption 3.9 (Bandwidths).** (a) For the estimator  $\lambda_{n0}$ ,  $nh_{n1}^{-1}h_{nz}^6 \rightarrow \infty$ ,  $nh_{n1}^{1+4q} \rightarrow 0$ ,  $nh_{n1}h_{n2}^{2q} \rightarrow 0$ ,  $nh_{n1}h_{nz}^{2r} \rightarrow 0$ ,  $\log n/(nh_{n1}h_{nz}^4)^{1/4} \rightarrow 0$ , and  $\log n/(nh_{n1}^{-1}h_{n2}^2h_{nz}^4)^{1/4} \rightarrow 0$ .

(b) For the estimator  $\Lambda_{n0}$ ,  $nh_{nz}^6 \rightarrow \infty$ ,  $nh_{n1}^{2q} \rightarrow 0$ ,  $nh_{n2}^{2q} \rightarrow 0$ ,  $nh_{nz}^{2r} \rightarrow 0$ ,  $\log n/(nh_{n1}^2 h_{nz}^4)^{1/4} \rightarrow 0$ , and  $\log n/(nh_{n2}^2 h_{nz}^4)^{1/4} \rightarrow 0$ .

Assumptions 3.8 and 3.9 (a) are satisfied, for example, if  $K_T$  is a second-order kernel,  $K_Z$  is a fourth-order kernel,  $h_{n1} \propto n^{-1/5}$ ,  $h_{n2} \propto n^{-\kappa_2}$ , and  $h_{nz} \propto n^{-\kappa_z}$ , where  $1/5 < \kappa_2 < 2/5$ ,  $1/10 < \kappa_z < 1/5$ , and  $\kappa_2 + 2\kappa_z < 3/5$ . Also, Assumptions 3.8 and 3.9 (b) are satisfied, for example, if  $K_T$  is a second-order kernel,  $K_Z$  is a sixth-order kernel,  $h_{n1} \propto n^{-\kappa}$ ,  $h_{n2} \propto n^{-\kappa}$ , and  $h_{nz} \propto n^{-\kappa_z}$ , where  $1/4 < \kappa < 1/3$ ,  $1/12 < \kappa_z < 1/8$ , and  $\kappa + 2\kappa_z < 1/2$ .

Define

$$\begin{aligned}\varphi(t_2, z_1, z_2) &= p_z(z_1, z_2)^{-1} w(t_2, z_1, z_2) \exp(z_2 - z_1), \\ C(t_1, t_2, z_1, z_2) &= B(t_1, t_2 | z_1, z_2)^{-1} \varphi(t_2, z_1, z_2), \\ D(t_1, t_2, z_1, z_2) &= B(t_1, t_2 | z_1, z_2)^{-2} A(t_1, t_2 | z_1, z_2) \varphi(t_2, z_1, z_2), \\ \gamma_t(T_{i1}, T_{i2}, X_{i1}, X_{i2}) &= \left[ \int_{S_T} C(t, t_2, Z_{i1}, Z_{i2}) 1(T_{i2} > t_2) dt_2 \right] \frac{1}{h_{n1}} K_T \left( \frac{t - T_{i1}}{h_{n1}} \right) - \lambda_0(t),\end{aligned}$$

and

$$\begin{aligned}\Gamma_t(T_{i1}, T_{i2}, X_{i1}, X_{i2}) &= \left[ \int_{S_T} C(T_{i1}, t_2, Z_{i1}, Z_{i2}) 1(T_{i2} > t_2) dt_2 \right] 1(0 \leq T_{i1} \leq t) \\ &\quad - \int_0^t D(t_1, T_{i2}, Z_{i1}, Z_{i2}) 1(T_{i1} > t_1) dt_1 \\ &\quad - \Lambda_0(t) \left[ \int_{S_Z} dz_1 \int_{S_Z} dz_2 \frac{w_z(z_1) w_z(z_2)}{p_z(z_1, z_2)} \right] E[X_1 - X_2]' \Omega(T_{i1}, T_{i2}, X_{i1}, X_{i2}).\end{aligned}$$

In addition, define

$$\begin{aligned}B_\lambda(t) &= \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 \left[ \int_{S_T} C(t, s_2, z_1, z_2) 1(t_2 > s_2) ds_2 \right] \frac{\partial^q}{\partial t_1^q} p(t, t_2, z_1, z_2) \\ &\quad \times \frac{1}{q!} \int_{-1}^1 u^q K_T(u) du\end{aligned}$$

and

$$\begin{aligned}V_\lambda(t) &= \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 \left[ \int_{S_T} C(t, s_2, z_1, z_2) 1(t_2 > s_2) ds_2 \right]^2 p(t, t_2, z_1, z_2) \\ &\quad \times \int_{-1}^1 K_T^2(u) du.\end{aligned}$$

The following theorem gives the main result of this section.

**Theorem 3.1.** *Let Assumptions 3.1-3.9 hold. Let  $[0, \tau] \subset I_T$  be a compact interval. Then as  $n \rightarrow \infty$ ,*

$$(a) \quad \lambda_{n0}(t) - \lambda_0(t) = n^{-1} \sum_{i=1}^n \gamma_t(T_{i1}, T_{i2}, X_{i1}, X_{i2}) - E[\gamma_t(T_1, T_2, X_1, X_2)] \\ + h_{n1}^q B_\lambda(t) + o_p \left[ (nh_{n1})^{-1/2} \right] + o_p(h_{n1}^q) \quad \text{and}$$

$$(b) \quad \Lambda_{n0}(t) - \Lambda_0(t) = n^{-1} \sum_{i=1}^n \Gamma_t(T_{i1}, T_{i2}, X_{i1}, X_{i2}) + o_p \left( n^{-1/2} \right)$$

*uniformly over  $t \in [0, \tau]$ .*

Theorem 3.1 implies that the rate of convergence in probability of  $\lambda_{n0}$  to  $\lambda_0$  is maximized at a  $n^{-q/(2q+1)}$  rate by setting  $h_{n1} \propto n^{-1/(2q+1)}$  and that  $\Lambda_{n0}$  converges to  $\Lambda_0$  in probability uniformly at a  $n^{-1/2}$  rate. Let  $\Rightarrow$  denote weak convergence in the space of bounded, real-valued functions on  $[0, \tau]$  equipped with the uniform metric. The following corollary of Theorem 3.1 is easily proved.

**Corollary 3.2.** *Let the assumptions of Theorem 3.1 hold.*

(a) *Assume  $h_{n1} \propto n^{-1/(2q+1)}$ . For  $t \in [0, \tau]$ ,*

$$n^{q/(2q+1)}[\lambda_{n0}(t) - \lambda_0(t)] \rightarrow_d \mathbf{N}(B_\lambda(t), V_\lambda(t)).$$

(b) *On  $[0, \tau]$ ,*

$$n^{1/2}[\Lambda_{n0}(t) - \Lambda_0(t)] \Rightarrow \chi_\Lambda(t),$$

*where  $\chi_\Lambda(t)$  is a tight Gaussian process with mean 0 and covariance function  $E[\chi_\Lambda(t)\chi_\Lambda(t')] = E[\Gamma_t(T_1, T_2, X_1, X_2)\Gamma_{t'}(T_1, T_2, X_1, X_2)]$ .*

Under the assumptions of Corollary 3.2, the asymptotic distribution of  $n^{q/(2q+1)}(\lambda_{n0} - \lambda_0)$  is not centered at zero. The asymptotic bias  $B_\lambda$  can be removed by undersmoothing  $\lambda_{n0}$  (equivalently, by letting  $h_{n1}$  converge faster than  $n^{-1/(2q+1)}$ ) at the expense of the reduced rate of convergence. The asymptotic variance  $V_\lambda$  of  $\lambda_{n0}$  and the covariance function of  $\chi_\Lambda$  can be estimated consistently by replacing unknown quantities with sample analogs. See Appendix A.2 for details.



## 4 Bandwidth Selection

This section describes rule-of-thumb, data-driven methods for choosing the values of the bandwidths  $h_{n1}$ ,  $h_{n2}$ , and  $h_{nz}$  for the uncensored case.

We first consider the choice of  $h_{n1}$ . An asymptotically optimal bandwidth  $h_{n1}^*$  in estimation of  $\lambda_0$  can be defined as a minimizer of the weighted asymptotic integrated mean-square error of  $\lambda_{n0}$ . It follows from Section 3 that  $h_{n1}^* = c_* n^{-1/(2q+1)}$ , where

$$c_* = \left[ \frac{\int w(t) V_\lambda(t) dt}{2q \int w(t) B_\lambda^2(t) dt} \right]^{1/(2q+1)}$$

and  $w(\cdot)$  is a weight function. A feasible bandwidth requires an estimate of the constant factor  $c_*$ . To develop a rule of thumb for choosing  $h_{n1}$ , assume that  $\varepsilon \equiv e^U$  has a gamma distribution with mean 1 and unknown variance  $\theta$  and is independent of  $X_j$ . Also, assume that  $\lambda_0$  belongs to a known parametric family. In the Monte Carlo experiments reported in Section 5, we use the following form

$$\lambda_0(t, \alpha) = \alpha_1 t^{\alpha_1 - 1} + \alpha_3 \alpha_2 t^{\alpha_2 - 1},$$

where  $\alpha \equiv (\alpha_1, \alpha_2, \alpha_3)$  is a vector of unknown positive constants. This form can be viewed as a mixture of Weibull hazards and is flexible enough to exhibit non-monotone hazards. Under the parametric specification of  $\lambda_0$ , it is straightforward to show that the probability density function of  $T_1$  and  $T_2$  conditional on  $Z_1 = z_1$  and  $Z_2 = z_2$  has the form

$$p_{t|z}(t_1, t_2 | z_1, z_2) = \frac{(1 + \theta) \lambda_0(t_1, \alpha) \lambda_0(t_2, \alpha) e^{z_1 + z_2}}{[\theta \Lambda_0(t_1, \alpha) e^{z_1} + \theta \Lambda_0(t_2, \alpha) e^{z_2} + 1]^{2+1/\theta}}. \quad (18)$$

This suggests that  $\theta$  and  $\alpha$  can be estimated by maximizing the log-likelihood function obtained from  $p_{t|z}$ . Once  $\theta$  and  $\alpha$  are estimated, then  $c_*$  can be evaluated numerically with an additional assumption about the distribution of  $Z_1$  and  $Z_2$ . In the Monte Carlo experiments, we use

$$p_z(z_1, z_2) = \frac{1}{s_1 s_2} \phi\left(\frac{z_1 - m_1}{s_1}\right) \phi\left(\frac{z_2 - m_2}{s_2}\right),$$

where  $\phi$  is the probability density function of the standard normal distribution, and  $m_j$  and  $s_j$  are the sample mean and standard deviation of  $Z_{nj}$  for each  $j = 1, 2$ . Let  $\hat{c}_*$  denote the resulting constant factor.

Now consider  $h_{n2}$  and  $h_{nz}$  in estimation of  $\lambda_0$ . Unlike  $h_{n1}$ ,  $h_{n2}$  and  $h_{nz}$  do not affect the asymptotic distribution of  $\lambda_{n0}$  if Assumption 3.9 is satisfied. Therefore, the values of  $h_{n2}$  and  $h_{nz}$  are less critical than the value of  $h_{n1}$ . If  $K_T$  is a second-order kernel and  $K_Z$  is a fourth-order kernel, then the following rule of thumb can be used:  $h_{n2} = \hat{c}_* n^{-2/9}$  and  $h_{nz} = s_* \hat{c}_* n^{-1/9}$ , where  $s_* = (s_1 + s_2)/2$ . This rule satisfies Assumption 3.9 and the Monte Carlo experiments in Section 5 indicate that it performs well. Similarly, one can choose the values of bandwidths in estimation of  $\Lambda_0$ . If  $K_T$  is a second-order kernel and  $K_Z$  is a sixth-order kernel, then one can use the following rule:  $h_{n1} = h_{n2} = \hat{c}_* n^{-2/7}$  and  $h_{nz} = s_* \hat{c}_* n^{-1/11}$ .

A similar, data-based method could be developed to choose the values of the bandwidths for the censored case, although details for the censored case would be quite different from those for the uncensored case. The rule-of-thumb bandwidths presented here can be used as pilot bandwidths for more sophisticated plug-in methods.

## 5 Extensions

### 5.1 Time-varying Covariates

This section outlines an extension of the model (1) that allows for time-varying covariates, provided that the time-varying covariates have the same known time paths for all individuals. The model has the form

$$\lambda_{T_j} \left( t_j \mid x_j, \{x_{vj}(s_j)\}_0^{t_j}, u \right) = \lambda_0(t_j) \exp \left( x_j' \beta + x_{vj}(t_j) \beta_v + u \right),$$

where  $X_{vj}$  is an (additional) real-valued, time-varying explanatory variable,  $\beta_v$  is an unknown coefficient of  $X_{vj}$ , and  $\{x_{vj}(s_j)\}_0^{t_j}$  denotes the time path of  $X_{vj}$  up to  $t_j$  for  $j = 1, 2$ . Moreover, assume that  $X_{vj}(t_j)$ 's have the same time path for all individuals and are constant on intervals, for example  $X_{vj}(t_j) = 1(t_j > \tau_j)$  for some known  $\tau_j$  satisfying  $\tau_1 \neq \tau_2$ .

First consider the uncensored case. The partial likelihood approach of Chamberlain (1985) and Ridder and Tunalı(1999) allows for time-varying covariates and thus estimators of  $\beta$  and  $\beta_v$  are available. Hence, as in Section 2.1, we only consider estimation of  $\lambda_0$  and  $\Lambda_0$ . Let  $Z_{vj}(t_j) = X_{vj}(t_j) \beta_v$  for  $j = 1, 2$  and let  $S(t_1, t_2 \mid z_1, z_2, \{z_{v1}(s_1)\}_0^{t_1}, \{z_{v2}(s_2)\}_0^{t_2})$  denote

the joint survivor function of  $T_1$  and  $T_2$  conditional on  $Z_1 = z_1$ ,  $Z_2 = z_2$ ,  $\{Z_{v1}(s_1)\}_0^{t_1} = \{z_{v1}(s_1)\}_0^{t_1}$ , and  $\{Z_{v2}(s_2)\}_0^{t_2} = \{z_{v2}(s_2)\}_0^{t_2}$ . It is straightforward to show that

$$\frac{\lambda_0(t_1)}{\lambda_0(t_2)} = \frac{\partial S(t_1, t_2 | z_1, z_2, \{z_{v1}(s_1)\}_0^{t_1}, \{z_{v2}(s_2)\}_0^{t_2}) / \partial t_1}{\partial S(t_1, t_2 | z_1, z_2, \{z_{v1}(s_1)\}_0^{t_1}, \{z_{v2}(s_2)\}_0^{t_2}) / \partial t_2} \exp(-[z_1 - z_2] - [z_{v1}(t_1) - z_{v2}(t_2)]).$$

Then estimators of  $\lambda_0$  and  $\Lambda_0$  can be obtained by methods identical to those in Section 2.1 except that the averaging is now done interval by interval.<sup>15</sup>

Now consider the censored case. Abusing notation a bit, let  $S(t_1, t_2 | x_1, x_2, \{x_{v1}(s_1)\}_0^{t_1}, \{x_{v2}(s_2)\}_0^{t_2})$  denote the joint survivor function of  $T_1$  and  $T_2$  conditional on  $X_1 = x_1$ ,  $X_2 = x_2$ ,  $\{X_{v1}(s_1)\}_0^{t_1} = \{x_{v1}(s_1)\}_0^{t_1}$ , and  $\{X_{v2}(s_2)\}_0^{t_2} = \{x_{v2}(s_2)\}_0^{t_2}$ . By setting  $t_1 = t_2 = t$ , we have

$$\log \left[ \frac{\partial S(t, t | x_1, x_2, \{x_{v1}(s_1)\}_0^t, \{x_{v2}(s_2)\}_0^t) / \partial t_1}{\partial S(t, t | x_1, x_2, \{x_{v1}(s_1)\}_0^t, \{x_{v2}(s_2)\}_0^t) / \partial t_2} \right] = [x_1 - x_2]' \beta + [x_{v1}(t) - x_{v2}(t)] \beta_v.$$

By integrating out over  $t$ , we have

$$\begin{aligned} & \int_{S_\beta} w_\beta(t) \log \left[ \frac{\partial S(t, t | x_1, x_2, \{x_{v1}(s_1)\}_0^t, \{x_{v2}(s_2)\}_0^t) / \partial t_1}{\partial S(t, t | x_1, x_2, \{x_{v1}(s_1)\}_0^t, \{x_{v2}(s_2)\}_0^t) / \partial t_2} \right] dt \\ &= [x_1 - x_2]' \beta + \left[ \int_{S_\beta} w_\beta(t) [x_{v1}(t) - x_{v2}(t)] dt \right] \beta_v. \end{aligned}$$

The estimation methods in Section 2.2 now can be adapted to develop estimators of  $\beta$ ,  $\beta_v$ ,  $\lambda_0$  and  $\Lambda_0$  for the censored case.

## 5.2 Combination of Possible Estimators

This section presents a method for combining possible estimators of  $\lambda$  and  $\Lambda_0$ . As was noted in Section 2,  $\lambda_0$  can be expressed as (3) or (4). Combining these expressions yields

$$\begin{aligned} \lambda_0(t) &= \alpha(t) \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 w(t_2, z_1, z_2) \exp(z_2 - z_1) R(t, t_2 | z_1, z_2) \\ &\quad + (1 - \alpha(t)) \int_{S_T} dt_1 \int_{S_Z} dz_1 \int_{S_Z} dz_2 w(t_1, z_1, z_2) \exp(z_1 - z_2) R(t_1, t | z_1, z_2)^{-1} \end{aligned} \tag{19}$$

for any  $\alpha(t)$  such that  $0 \leq \alpha(t) \leq 1$  for all  $t$ . This suggests that  $\lambda_0$  can be estimated by (19) with  $R$  replaced by its consistent estimator  $R_n$ . Let  $\hat{\lambda}_{n0}$  denote the resulting estimator of  $\lambda_0$ .

---

<sup>15</sup>We are grateful to an anonymous referee who pointed this out.

For simplicity, we consider only uncensored case and assume that  $h_{n1} = h_{n2} \equiv h_n$ . Under the assumptions of Theorem 3.1, it can be shown that as  $n \rightarrow \infty$ ,

$$\begin{aligned} \hat{\lambda}_{n0} &= \frac{\alpha(t)}{nh_n} \sum_{i=1}^n \left[ \int_{S_T} C(t, t_2, Z_{i1}, Z_{i2}) 1(T_{i2} > t_2) dt_2 \right] K_T \left( \frac{t - T_{i1}}{h_n} \right) \\ &\quad + \frac{1 - \alpha(t)}{nh_n} \sum_{i=1}^n \left[ \int_{S_T} \tilde{C}(t_1, t, Z_{i1}, Z_{i2}) 1(T_{i1} > t_1) dt_1 \right] K_T \left( \frac{t - T_{i2}}{h_n} \right) \\ &\quad - \lambda_0(t) + o_p \left[ (nh_n)^{-1/2} \right] \end{aligned}$$

uniformly over  $t \in [0, \tau]$ , where

$$\tilde{C}(t_1, t_2, z_1, z_2) = [A(t_1, t_2 | z_1, z_2) p_z(z_1, z_2)]^{-1} w(t_1, z_1, z_2) \exp(z_1 - z_2).$$

The weight function  $\alpha(t)$  can be chosen to minimize the mean squared error of  $\hat{\lambda}_{n0}(t)$  for each  $t \in [0, \tau]$ .

Similarly,  $\Lambda_0$  can be expressed as

$$\begin{aligned} \Lambda_0(t) &= \alpha(t) \int_0^t dt_1 \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 w(t_2, z_1, z_2) \exp(z_2 - z_1) R(t_1, t_2 | z_1, z_2) \\ &\quad (1 - \alpha(t)) \int_0^t dt_2 \int_{S_T} dt_1 \int_{S_Z} dz_1 \int_{S_Z} dz_2 w(t_1, z_1, z_2) \exp(z_1 - z_2) R(t_1, t_2 | z_1, z_2)^{-1}. \end{aligned}$$

A new estimator of  $\Lambda_0$  can be obtained by replacing  $R$  in the equation above with  $R_n$ .

### 5.3 Estimation with Longer Panels

The estimation approach described in this paper extends easily to the case of longer panels. First consider the case when observations of  $T_j$  are uncensored. Observations of any pair of the set  $\{1, \dots, J\}$  can be used to construct nonparametric estimators of  $\lambda_0$  and  $\Lambda_0$  as in Section 2.1 (or as in Section 5.1). This gives  $J(J-1)/2$  different estimators, and these can be linearly combined to construct a more efficient estimator. It may be an interesting question what linear combination yields the smallest integrated mean square error among all linear combinations possible, but it is beyond the scope of this paper. Chamberlain (1985) discusses estimation of  $\beta$  when  $J$  completed spells are available for each individual.

For the censored case, we assume that  $C_1 = C$  and  $C_j = (C - \sum_{k=1}^{j-1} T_k) 1(T_{j-1} \leq C_{j-1})$  for  $j = 2, \dots, J$ . Here,  $C$  is conditionally independent of  $T_j$  given  $X_j$ . As in Section 2.2,

observe that  $C$  censors the sum of  $T_j$ , not each separately, and that  $\Delta_j = 1$  for  $j < J$  if  $\Delta_J = 1$ .

To describe an estimator of  $\beta$ , let  $(t_l, t_k)$  be a pair such that  $l \neq k$ . Define the joint survivor function  $S(t_l, t_k|x_l, x_k) = \Pr(T_l > t_l, T_k > t_k|X_l = x_l, X_k = x_k)$ , the joint conditional sub-distribution function  $F(t_1, \dots, t_J|x_l, x_k) = \Pr(Y_1 \leq t_1, \dots, Y_J \leq t_J, \Delta_J = 1|X_l = x_l, X_k = x_k)$ , its corresponding density  $f(t_1, \dots, t_J|x_l, x_k) = \partial^J F(t_1, \dots, t_J|x_l, x_k)/\partial t_1 \dots \partial t_J$ , and the conditional survivor function of the censoring threshold  $G(c|x_1, x_2) = \Pr(C > c|X_l = x_l, X_k = x_k)$ .

As in the equation (12),

$$S(t_l, t_k|x_l, x_k) = \int_{t_l}^{\infty} \int_{t_k}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \frac{f(s_1, \dots, s_J|x_l, x_k)}{G(s_1 + \dots + s_J|x_l, x_k)} ds_{-lk} ds_k ds_l, \quad (20)$$

where  $t_{-lk}$  denotes a vector containing all components of  $(t_1, \dots, t_J)$  except  $t_l$  and  $t_k$ . By differentiating  $S$  with respect to  $t_l$  and  $t_k$  and then setting  $t_l = t_k = t$ ,

$$\frac{\partial S(t_l, t_k|x_l, x_k)/\partial t_l}{\partial S(t_l, t_k|x_l, x_k)/\partial t_k} \Big|_{t_l=t_k=t} = \exp[(x_l - x_k)'\beta]. \quad (21)$$

Now  $\beta$  can be estimated by using a procedure similar to the one described in Section 2.2.1. Estimators of  $\lambda_0$  and  $\Lambda_0$  can also be developed analogously.

## 6 Monte Carlo Experiments

This section presents the results of a small set of Monte Carlo experiments that illustrate the numerical performance of the estimators of  $\lambda_0$ ,  $\Lambda_0$ , and  $\beta$ . Samples were generated by simulation from model (1) with  $J = 2$ . In the experiments,  $\beta = 1$ ,  $X_1 \sim \mathbf{N}(0, 1)$ ,  $X_2 \sim \mathbf{N}(0, 1)$ , and  $X_1$  and  $X_2$  are independent. The fixed effect was generated by  $U = (X_1 + X_2)/2$ . Experiments were carried out with two baseline hazard functions, which are taken from Horowitz (1999). One is  $\lambda_0(t) = 0.087t$ , which makes (1) a Weibull proportional hazard model with unobserved heterogeneity. The other baseline hazard function is  $\lambda_0(t) = 0.05(t/5)^{-2/3} + 0.57(t/5)^5$ , which is U-shaped.

Experiments were also carried out for both the uncensored and censored cases. The censoring threshold  $C$  was generated from the exponential distribution with mean 20. Recall that  $C_1 = C$  and  $C_2 = (C - T_1)1(T_1 \leq C)$ . Under this censoring mechanism, the means of

$\Delta_1$  and  $\Delta_2$  are about 0.78 and 0.64, respectively, for the Weibull hazard model and about 0.87 and 0.76, respectively, for the U-shaped hazard model.

The experiments used sample sizes of  $n = 100$  and  $500$ . There were 100 Monte Carlo replications per experiment, and the experiments were carried out in GAUSS using GAUSS pseudo-random number generators.

We first focus on the finite sample performance of the estimators of  $\lambda_0$  and  $\Lambda_0$  for the uncensored case. The partial likelihood estimator was used to estimate  $\beta$ . The kernel functions used in estimation of  $\lambda_0$  are

$$K_T(u) = (15/16)(1 - u^2)^2 1(|u| \leq 1) \quad (22)$$

and

$$K_Z(u) = (105/64)(1 - 5u^2 + 7u^4 - 3u^6) 1(|u| \leq 1). \quad (23)$$

These are second-order and fourth-order kernels. The following sixth-order kernel along with (22) is used in estimation of  $\Lambda_0$ :

$$K_Z(u) = (315/2048)(15 - 140u^2 + 378u^4 - 396u^6 + 143u^8) 1(|u| \leq 1). \quad (24)$$

All the kernel functions are taken from Müller (1984). The bandwidths were chosen by the data-based methods described in Section 4. The weight functions and the means of the values of bandwidths used in the experiments are shown in Table 1.<sup>16</sup> It is not difficult to compute  $\lambda_{n0}$  and  $\Lambda_{n0}$ . The triple integral in (5) was evaluated numerically using the Gauss-Legendre quadrature method. The quadruple integral in (6) was first evaluated analytically with respect to  $t_1$  and the remaining triple integral was evaluated numerically. See Horowitz and Gørgens (1999, 2.4) for details how the integral in (6) can be evaluated analytically with respect to  $t_1$ .

The results of the experiments are summarized graphically in Figure 1 for the Weibull model and Figure 2 for the U-shaped hazard model. The left-hand panels of the figures show the means of 100 estimates of  $\lambda_0$  and  $\Lambda_0$  (solid lines) and the true  $\lambda_0$  and  $\Lambda_0$  (dashed lines).

---

<sup>16</sup>The weight function  $w_t(\cdot)$  does not satisfy the differentiability requirement of Assumption 3.6. This does not matter in a finite sample because there are no observations of  $T_2$  at discontinuous points.

The right-hand panels show five individual estimates of  $\lambda_0$  and  $\Lambda_0$  (solid lines) and the true  $\lambda_0$  and  $\Lambda_0$  (dashed lines). The baseline hazard functions used in the experiments do not satisfy the scale normalization; hence, the estimates were normalized by dividing them by  $\int_0^\infty w_t(t)/\lambda_0(t)dt$ . It can be seen that the true functions and the means of the estimates are quite close to one another, especially when  $n = 500$ . It is not surprising that the estimates of  $\lambda_0$  are more variable than those of  $\Lambda_0$  given the rates of convergence of the estimators obtained in Section 3. Most of the individual estimates are reasonable approximations to the functions they estimate.

In order to investigate whether there is an advantage to using a combined estimator of  $\lambda_0$  described in Section 5, we computed  $\hat{\lambda}_{n0}$  using equal weight for each  $t$  ( $\alpha(t) = 0.5$ ) with the same bandwidths used in  $\lambda_{n0}$ . Figure 3 shows the means of 100 estimates of  $\lambda_0$  and five individual estimates. It can be seen that the biases of  $\hat{\lambda}_{n0}$  remain virtually the same as those of  $\lambda_{n0}$  but the variances of  $\hat{\lambda}_{n0}$  are somewhat smaller than those of  $\lambda_{n0}$ . This is not surprising given the fact that  $\hat{\lambda}_{n0}$  is just a weighted average of consistent estimators.

We now turn to investigate the small sample performance of the estimators for the censored case. The parameter  $\beta$  was estimated by the method described in Section 2.2.1. The regularity conditions established in Appendix B.1 require  $K_T$  to be a higher-order kernel in order to prevent  $\beta_n$  from having the asymptotic bias. As is well known, however, kernel estimates with second-order kernels often outperform those with higher-order kernels for small sample sizes.<sup>17</sup> Due to this reason, the experiments were carried out using both the second-order and fourth-order kernels (22) - (23) for  $K_T$ .<sup>18</sup> The second-order kernel (22) was used for  $K_X$ . The single integral in  $V_{ni}$  in Section 2.2.1 was evaluated numerically using the quadrature method. As in the uncensored case, the kernels (22) and (23) were used in estimation of  $\lambda_0$ ; the kernels (22) and (24) were used for  $\Lambda_{n0}$ . Estimates of  $\beta$  with the fourth-order kernel were used as  $\beta_n$  in estimation of  $\lambda_0$  and  $\Lambda_0$ . The weight functions and the values of bandwidths used for the censored case are shown in Table 2. The bandwidths were chosen to roughly minimize the (integrated) mean square errors of the estimators.

<sup>17</sup>For example, see Efromovich (2001) for theoretical arguments why the higher-order kernels perform poorly in small samples.

<sup>18</sup>When the fourth-order kernel is used,  $V_{ni}$  in (15) can be negative for finite samples. To deal with this problem, we set  $w_{xi} = 0$  when  $V_{ni}$  is not strictly positive.

The results for the censored case are summarized in Table 3 and Figures 4-5. Table 3 reports the results of the experiments for  $\beta_n$ . It is not surprising that the estimates of  $\beta$  exhibit some biases when the second-order kernel is used, given the fact that a higher-order kernel is needed to remove the bias. On the other hand, the use of the higher-order kernel reduces the biases at the expense of increased variances. In order to compare the censored estimator of  $\beta$  to the uncensored estimator, we computed the root mean square error (RMSE) of the partial likelihood estimator without censoring. The resulting RMSE's were 0.228 and 0.098, respectively, for sample sizes of  $n = 100$  and  $500$ .<sup>19</sup> Thus, the RMSE of the censored estimator is quite larger than that of the uncensored estimator roughly by a factor of 2. Figures 4 and 5 show the means of 100 estimates of  $\lambda_0$  and  $\Lambda_0$  and five individual estimates, as was shown in Figures 1 and 2. It can be seen that as in the uncensored case, the true functions and the means of the estimates are quite close to one another and the individual estimates are reasonable approximations to the functions they estimate.

## 7 Conclusions

This paper has presented nonparametric estimators of the baseline and integrated baseline hazard functions in a panel data proportional hazards model with fixed effects. The paper has also shown how the parametric part of the model can be estimated consistently with dependent right censoring, under which the partial likelihood estimator is inconsistent. Although our censored estimator is a  $n^{-1/2}$ -consistent estimator, it seems to have quite large variance as compared to the uncensored counterpart. Therefore, it may be an interesting problem to develop a more efficient estimator than one proposed here. Furthermore, it may also be interesting to find the semiparametric efficiency bound for the parametric part of the model by extending the result of Hahn (1994). These are topics for future research.

## 8 Acknowledgements

We thank John Geweke, George Neumann, Forrest Nelson, Gene Savin, seminar participants at Iowa, Maryland, Northwestern, Princeton, Queen's, Rutgers, Toronto, UCL, UCLA,

---

<sup>19</sup>The functional form of the baseline hazard function is irrelevant for the partial likelihood estimator.



Western Ontario, Yale, the 2002 North American Summer Meetings of Econometric Society, and the 10th International Conference on Panel Data, the editor, and three anonymous referees for helpful comments and suggestions on an earlier draft of this paper. This research was supported in part by NSF Grant SES-9910925.

## A Appendix: Uncensored Case

### A.1 Proofs of Theorems

This subsection of Appendix A presents the proofs of Theorem 3.1 and Corollary 3.2. Define a Euclidean class of functions as in Pakes and Pollard (1989). Define  $A(t_1, t_2, z_1, z_2) = A(t_1, t_2|z_1, z_2)p_z(z_1, z_2)$ ,  $B(t_1, t_2, z_1, z_2) = B(t_1, t_2|z_1, z_2)p_z(z_1, z_2)$ ,  $A_n(t_1, t_2, z_1, z_2) = A_n(t_1, t_2|z_1, z_2)p_{nz}(z_1, z_2)$ , and  $B_n(t_1, t_2, z_1, z_2) = B_n(t_1, t_2|z_1, z_2)p_{nz}(z_1, z_2)$ . Equation (8) can be rewritten as

$$R_n(t_1, t_2|z_1, z_2) = A_n(t_1, t_2, z_1, z_2)/B_n(t_1, t_2, z_1, z_2). \quad (25)$$

In order to prove Theorem 3.1 and Corollary 3.2, it is more convenient to use (25) than (8).

Before we prove Theorem 3.1, it is useful to prove some lemmas that establish asymptotic linear approximations of  $A_n(t_1, t_2, z_1, z_2)$  and  $B_n(t_1, t_2, z_1, z_2)$ . Define

$$\begin{aligned} A_n^{(1)}(t_1, t_2, z_1, z_2) &= \frac{1}{nh_{n1}h_{nz}^2} \sum_{i=1}^n 1(T_{i2} > t_2) K_T \left( \frac{t_1 - T_{i1}}{h_{n1}} \right) K_Z \left( \frac{z_1 - Z_{i1}}{h_{nz}} \right) K_Z \left( \frac{z_2 - Z_{i2}}{h_{nz}} \right), \\ A_n^{(2)}(t_1, t_2, z_1, z_2) &= -\frac{1}{nh_{n1}h_{nz}^3} \sum_{i=1}^n 1(T_{i2} > t_2) K_T \left( \frac{t_1 - T_{i1}}{h_{n1}} \right) \\ &\quad \times \left\{ K'_Z \left( \frac{z_1 - Z_{i1}}{h_{nz}} \right) K_Z \left( \frac{z_2 - Z_{i2}}{h_{nz}} \right) X_{i1} + K_Z \left( \frac{z_1 - Z_{i1}}{h_{nz}} \right) K'_Z \left( \frac{z_2 - Z_{i2}}{h_{nz}} \right) X_{i2} \right\}, \\ A^{(2)}(t_1, t_2, z_1, z_2) &= -\frac{\partial}{\partial z_1} A(t_1, t_2|z_1, z_2) EX_1 - \frac{\partial}{\partial z_2} A(t_1, t_2|z_1, z_2) EX_2, \end{aligned}$$

and

$$S_\Omega = n^{-1} \sum_{i=1}^n \Omega(T_{i1}, T_{i2}, X_{i1}, X_{i2}).$$

**Lemma A.1.** *As  $n \rightarrow \infty$ , the following holds uniformly over  $(t_1, t_2, z_1, z_2) \in [0, \tau] \times S_T \times S_Z \times S_Z$ :*

- (a)  $A_n(t_1, t_2, z_1, z_2) = A_n^{(1)}(t_1, t_2, z_1, z_2) + A_n^{(2)}(t_1, t_2, z_1, z_2)' S_\Omega + O_p(n^{-1}h_{nz}^{-3}) + o_p\left[\log n/(n^3h_{n1}h_{nz}^7)^{1/2}\right] + o_p(n^{-1/2})$ .
- (b)  $A_n^{(1)}(t_1, t_2, z_1, z_2) = A(t_1, t_2, z_1, z_2) + O(h_{n1}^q) + O(h_{nz}^r) + o\left[\log n/(nh_{n1}h_{nz}^2)^{1/2}\right]$  *a.s.*
- (c)  $A_n^{(2)}(t_1, t_2, z_1, z_2) = A^{(2)}(t_1, t_2, z_1, z_2) + O(h_{n1}^q) + O(h_{nz}^r) + o\left[\log n/(nh_{n1}h_{nz}^4)^{1/2}\right]$  *a.s.*

*Proof. Part (a):* By a Taylor series expansion, write

$$A_n(t_1, t_2, z_1, z_2) = A_n^{(1)}(t_1, t_2, z_1, z_2) + A_n^{(2)}(t_1, t_2, z_1, z_2)'(b_n - \beta) + R_n^{(A)}(t_1, t_2, z_1, z_2), \quad (26)$$

where  $R_n^{(A)}(t_1, t_2, z_1, z_2)$  is a remainder term such that

$$\begin{aligned} R_n^{(A)}(t_1, t_2, z_1, z_2) &= \left[ R_n^{(A1)}(t_1, t_2, z_1, z_2) + R_n^{(A2)}(t_1, t_2, z_1, z_2) \right]' (b_n - \beta), \\ R_n^{(A1)}(t_1, t_2, z_1, z_2) &= \frac{1}{nh_{n1}h_{nz}^3} \sum_{i=1}^n X_{i1} 1(T_{i2} > t_2) K_T \left( \frac{t_1 - T_{i1}}{h_{n1}} \right) \\ &\quad \times \left\{ K'_Z \left( \frac{z_1 - Z_{i1}}{h_{nz}} \right) K_Z \left( \frac{z_2 - Z_{i2}}{h_{nz}} \right) - K'_Z \left( \frac{z_1 - \tilde{Z}_{ni1}}{h_{nz}} \right) K_Z \left( \frac{z_2 - \tilde{Z}_{ni2}}{h_{nz}} \right) \right\}, \\ R_n^{(A2)}(t_1, t_2, z_1, z_2) &= \frac{1}{nh_{n1}h_{nz}^3} \sum_{i=1}^n X_{i2} 1(T_{i2} > t_2) K_T \left( \frac{t_1 - T_{i1}}{h_{n1}} \right) \\ &\quad \times \left\{ K_Z \left( \frac{z_1 - Z_{i1}}{h_{nz}} \right) K'_Z \left( \frac{z_2 - Z_{i2}}{h_{nz}} \right) - K_Z \left( \frac{z_1 - \tilde{Z}_{ni1}}{h_{nz}} \right) K'_Z \left( \frac{z_2 - \tilde{Z}_{ni2}}{h_{nz}} \right) \right\}, \end{aligned}$$

$\tilde{Z}_{nij} = X'_{ij} \tilde{b}_n$  for  $j = 1, 2$ , and  $\tilde{b}_n$  is between  $b_n$  and  $\beta$ .

Further, write

$$R_n^{(A)}(t_1, t_2, z_1, z_2) = H_n^{(A1)}(t_1, t_2, z_1, z_2) + H_n^{(A2)}(t_1, t_2, z_1, z_2) + H_n^{(A3)}(t_1, t_2, z_1, z_2),$$

where

$$\begin{aligned} H_n^{(A1)}(t_1, t_2, z_1, z_2) &= \frac{1}{nh_{n1}h_{nz}^3} \sum_{i=1}^n X_{i1} 1(T_{i2} > t_2) K_T \left( \frac{t_1 - T_{i1}}{h_{n1}} \right) \\ &\quad \times K'_Z \left( \frac{z_1 - Z_{i1}}{h_{nz}} \right) \left[ K_Z \left( \frac{z_2 - Z_{i2}}{h_{nz}} \right) - K_Z \left( \frac{z_2 - \tilde{Z}_{ni2}}{h_{nz}} \right) \right], \\ H_n^{(A2)}(t_1, t_2, z_1, z_2) &= \frac{1}{nh_{n1}h_{nz}^3} \sum_{i=1}^n X_{i1} 1(T_{i2} > t_2) K_T \left( \frac{t_1 - T_{i1}}{h_{n1}} \right) \\ &\quad \times K_Z \left( \frac{z_2 - Z_{i2}}{h_{nz}} \right) \left[ K'_Z \left( \frac{z_1 - Z_{i1}}{h_{nz}} \right) - K'_Z \left( \frac{z_1 - \tilde{Z}_{ni1}}{h_{nz}} \right) \right], \end{aligned}$$

and

$$\begin{aligned} H_n^{(A3)}(t_1, t_2, z_1, z_2) &= \frac{1}{nh_{n1}h_{nz}^3} \sum_{i=1}^n X_{i1} 1(T_{i2} > t_2) K_T \left( \frac{t_1 - T_{i1}}{h_{n1}} \right) \\ &\quad \times \left[ K_Z \left( \frac{z_2 - \tilde{Z}_{ni2}}{h_{nz}} \right) - K_Z \left( \frac{z_2 - Z_{i2}}{h_{nz}} \right) \right] \left[ K'_Z \left( \frac{z_1 - Z_{i1}}{h_{nz}} \right) - K'_Z \left( \frac{z_1 - \tilde{Z}_{ni1}}{h_{nz}} \right) \right]. \end{aligned}$$

Assumption 3.8 implies that  $K_Z$  is Lipschitz continuous, so for some  $M_1 < \infty$ ,

$$\begin{aligned} \|H_n^{(A1)}(t_1, t_2, z_1, z_2)\| &\leq \frac{M_1 \|\tilde{b}_n - \beta\|}{nh_{n1}h_{nz}^4} \sum_{i=1}^n \left| K_T \left( \frac{t_1 - T_{i1}}{h_{n1}} \right) 1(T_{i2} > t_2) \right| \left| K'_Z \left( \frac{z_1 - Z_{i1}}{h_{nz}} \right) \right| \|X_{i1}\| \|X_{i2}\| \\ &\equiv M_1 \|\tilde{b}_n - \beta\| \tilde{H}_n^{(A1)}(t_1, t_2, z_1). \end{aligned}$$

It is not difficult to show that the summand in  $\tilde{H}_n^{(A1)}(t_1, t_2, z_1)$  belongs to a Euclidean class. By Theorem 2.37 of Pollard (1984),

$$\sup_{t_1, t_2, z_1} \left| \tilde{H}_n^{(A1)}(t_1, t_2, z_1) - E\tilde{H}_n^{(A1)}(t_1, t_2, z_1) \right| = o \left[ \log n / (nh_{n1}h_{nz}^7)^{1/2} \right] \quad \text{a.s.}$$

In addition, a change of variables gives  $E\tilde{H}_n^{(A1)}(t_1, t_2, z_1) = O(h_{nz}^{-3})$  uniformly over  $(t_1, t_2, z_1) \in [0, \tau] \times S_T \times S_Z$ . Hence, since  $\|b_n - \beta\| = o_p(n^{-1/2})$ ,

$$\|H_n^{(A1)}(t_1, t_2, z_1, z_2)\| = o_p \left[ \log n / (n^2 h_{n1} h_{nz}^7)^{1/2} \right] + O_p \left( n^{-1/2} h_{nz}^{-3} \right) \quad (27)$$

uniformly over  $(t_1, t_2, z_1, z_2) \in [0, \tau] \times S_T \times S_Z \times S_Z$ . By the same arguments,

$$\|H_n^{(A2)}(t_1, t_2, z_1, z_2)\| = o_p \left[ \log n / (n^2 h_{n1} h_{nz}^7)^{1/2} \right] + O_p \left( n^{-1/2} h_{nz}^{-3} \right) \quad (28)$$

uniformly over  $(t_1, t_2, z_1, z_2) \in [0, \tau] \times S_T \times S_Z \times S_Z$ . In addition, for some  $M_2 < \infty$ ,

$$\begin{aligned} \|H_n^{(A3)}(t_1, t_2, z_1, z_2)\| &\leq \frac{M_2 \|\tilde{b}_n - \beta\|^2}{nh_{n1}h_{nz}^5} \sum_{i=1}^n \left| K_T \left( \frac{t_1 - T_{i1}}{h_{n1}} \right) 1(T_{i2} > t_2) \right| \|X_{i1}\|^2 \|X_{i2}\| \\ &\equiv M_2 \|\tilde{b}_n - \beta\|^2 \tilde{H}_n^{(A3)}(t_1, t_2). \end{aligned}$$

Again, by Theorem 2.37 of Pollard (1984),

$$\sup_{t_1, t_2} \left| \tilde{H}_n^{(A3)}(t_1, t_2) - E\tilde{H}_n^{(A3)}(t_1, t_2) \right| = o \left[ \log n / (nh_{n1}h_{nz}^{10})^{1/2} \right] \quad \text{a.s.}$$

Moreover, a change of variables gives  $E\tilde{H}_n^{(A3)}(t_1, t_2) = O(h_{nz}^{-5})$  uniformly over  $(t_1, t_2, z_1) \in [0, \tau] \times S_T \times S_Z$ . Hence,

$$\|H_n^{(A3)}(t_1, t_2, z_1, z_2)\| = o_p \left[ (\log n) / (n^3 h_{n1} h_{nz}^{10})^{1/2} \right] + O_p \left( n^{-1} h_{nz}^{-5} \right) \quad (29)$$

uniformly over  $(t_1, t_2, z_1, z_2) \in [0, \tau] \times S_T \times S_Z \times S_Z$ . It follows from (27), (28), and (29) that under the assumption that  $nh_{nz} \rightarrow \infty$ ,

$$\|R_n^{(A1)}(t_1, t_2, z_1, z_2)\| = o_p \left[ \log n / (n^2 h_{n1} h_{nz}^7)^{1/2} \right] + O_p \left( n^{-1/2} h_{nz}^{-3} \right)$$

uniformly over  $(t_1, t_2, z_1, z_2) \in [0, \tau] \times S_T \times S_Z \times S_Z$ . Similarly,

$$\|R_n^{(A2)}(t_1, t_2, z_1, z_2)\| = o_p \left[ \log n / (n^2 h_{n1} h_{nz}^7)^{1/2} \right] + O_p \left( n^{-1/2} h_{nz}^{-3} \right)$$

uniformly over  $(t_1, t_2, z_1, z_2) \in [0, \tau] \times S_T \times S_Z \times S_Z$ . Therefore,

$$\|R_n^{(A)}(t_1, t_2, z_1, z_2)\| = o_p \left[ \log n / (n^3 h_{n1} h_{nz}^7)^{1/2} \right] + O_p \left( n^{-1} h_{nz}^{-3} \right) \quad (30)$$

uniformly over  $(t_1, t_2, z_1, z_2) \in [0, \tau] \times S_T \times S_Z \times S_Z$ . Part (a) follows by combining (30) with the fact that  $b_n - \beta = S_\Omega + o_p(n^{-1/2})$ .

**Part (b):** Another application of Theorem 2.37 of Pollard (1984) yields

$$\sup_{t_1, t_2, z_1, z_2} \left| A_n^{(1)}(t_1, t_2, z_1, z_2) - EA_n^{(1)}(t_1, t_2, z_1, z_2) \right| = o \left[ \log n / (nh_{n1}h_{nz}^2)^{1/2} \right] \quad \text{a.s.}$$

In addition,

$$\begin{aligned} EA_n^{(1)}(t_1, t_2, z_1, z_2) &= (h_{n1}h_{nz}^2)^{-1} \int 1(s_2 > t_2) K_T \left( \frac{t_1 - s_1}{h_{n1}} \right) K_Z \left( \frac{z_1 - w_1}{h_{nz}} \right) \\ &\quad \times K_Z \left( \frac{z_2 - w_2}{h_{nz}} \right) p(s_1, s_2, w_1, w_2) ds_1 ds_2 dw_1 dw_2 \\ &= \int 1(\psi_2 > t_2) K_T(\psi_1) K_Z(\xi_1) K_Z(\xi_2) \\ &\quad \times p(t_1 - h_{n1}\psi_1, \psi_2, z_1 - h_{nz}\xi_1, z_2 - h_{nz}\xi_2) d\psi_1 d\psi_2 d\xi_1 d\xi_2 \\ &= A(t_1, t_2, z_1, z_2) + O(h_{n1}^q) + O(h_{nz}^r) \end{aligned}$$

uniformly over  $(t_1, t_2, z_1, z_2) \in [0, \tau] \times S_T \times S_Z \times S_Z$ . Thus, this proves part (b).

**Part (c):** This can be proved by using the similar arguments as in part (b).  $\square$

Define

$$\begin{aligned} B_n^{(1)}(t_1, t_2, z_1, z_2) &= \frac{1}{nh_{n2}h_{nz}^2} \sum_{i=1}^n 1(T_{i1} > t_1) K_T \left( \frac{t_2 - T_{i2}}{h_{n2}} \right) K_Z \left( \frac{z_1 - Z_{i1}}{h_{nz}} \right) K_Z \left( \frac{z_2 - Z_{i2}}{h_{nz}} \right), \\ B_n^{(2)}(t_1, t_2, z_1, z_2) &= -\frac{1}{nh_{n2}h_{nz}^3} \sum_{i=1}^n 1(T_{i1} > t_1) K_T \left( \frac{t_2 - T_{i2}}{h_{n2}} \right) \\ &\quad \times \left\{ K_Z' \left( \frac{z_1 - Z_{i1}}{h_{nz}} \right) K_Z \left( \frac{z_2 - Z_{i2}}{h_{nz}} \right) X_{i1} + K_Z \left( \frac{z_1 - Z_{i1}}{h_{nz}} \right) K_Z' \left( \frac{z_2 - Z_{i2}}{h_{nz}} \right) X_{i2} \right\}, \end{aligned}$$

and

$$B^{(2)}(t_1, t_2, z_1, z_2) = -\frac{\partial}{\partial z_1} B(t_1, t_2 | z_1, z_2) EX_1 - \frac{\partial}{\partial z_2} B(t_1, t_2 | z_1, z_2) EX_2.$$

**Lemma A.2.** *As  $n \rightarrow \infty$ , the following holds uniformly over  $(t_1, t_2, z_1, z_2) \in [0, \tau] \times S_T \times S_Z \times S_Z$ :*

- (a)  $B_n(t_1, t_2, z_1, z_2) = B_n^{(1)}(t_1, t_2, z_1, z_2) + B_n^{(2)}(t_1, t_2, z_1, z_2)' S_\Omega$   
 $+ O_p(n^{-1}h_{nz}^{-3}) + o_p \left[ \log n / (n^3 h_{n2} h_{nz}^7)^{1/2} \right] + o_p(n^{-1/2}).$
- (b)  $B_n^{(1)}(t_1, t_2, z_1, z_2) = B(t_1, t_2, z_1, z_2) + O(h_{n2}^q) + O(h_{nz}^r) + o \left[ \log n / (nh_{n2}h_{nz}^2)^{1/2} \right] \quad \text{a.s.}$
- (c)  $B_n^{(2)}(t_1, t_2, z_1, z_2) = B^{(2)}(t_1, t_2, z_1, z_2) + O(h_{n2}^q) + O(h_{nz}^r) + o \left[ \log n / (nh_{n2}h_{nz}^4)^{1/2} \right] \quad \text{a.s.}$

*Proof.* The lemma follows by repeating the same arguments as in the proof of Lemma 1.  $\square$

*Proof of Theorem 3.1. Part (a):* By the definition and a Taylor series expansion,

$$\begin{aligned}\lambda_{n0}(t) - \lambda_0(t) &= \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 w(t_2, z_1, z_2) \exp(z_2 - z_1) \left[ \frac{A_n(t_1, t_2, z_1, z_2)}{B_n(t_1, t_2, z_1, z_2)} - \frac{A(t_1, t_2, z_1, z_2)}{B(t_1, t_2, z_1, z_2)} \right] \\ &= \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 \\ &\quad \times \left[ C(t_1, t_2, z_1, z_2) A_n(t_1, t_2, z_1, z_2) - D(t_1, t_2, z_1, z_2) B_n(t_1, t_2, z_1, z_2) + R_n^{(\lambda)}(t_1, t_2, z_1, z_2) \right],\end{aligned}$$

where the remainder term  $R_n^{(\lambda)}(t_1, t_2, z_1, z_2)$  satisfies

$$R_n^{(\lambda)}(t_1, t_2, z_1, z_2) = O \left[ (A_n - A)(B_n - B) + (B_n - B)^2 \right].$$

It follows from Lemmas A.1 and A.2 and Assumption 3.9 (a) that

$$\begin{aligned}A_n(t_1, t_2, z_1, z_2) &= A_n^{(1)}(t_1, t_2, z_1, z_2) + o_p \left[ (nh_{n1})^{-1/2} \right], \\ B_n(t_1, t_2, z_1, z_2) &= B_n^{(1)}(t_1, t_2, z_1, z_2) + o_p \left[ (nh_{n1})^{-1/2} \right], \\ |A_n(t_1, t_2, z_1, z_2) - A(t_1, t_2, z_1, z_2)| &= o_p \left[ (nh_{n1})^{-1/4} \right],\end{aligned}$$

and

$$|B_n(t_1, t_2, z_1, z_2) - B(t_1, t_2, z_1, z_2)| = o_p \left[ (nh_{n1})^{-1/4} \right]$$

uniformly over  $(t_1, t_2, z_1, z_2) \in [0, \tau] \times S_T \times S_Z \times S_Z$ . Thus, it follows that

$$\begin{aligned}\lambda_{n0}(t) - \lambda_0(t) &= \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 C(t_1, t_2, z_1, z_2) A_n^{(1)}(t_1, t_2, z_1, z_2) \\ &\quad - \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 D(t_1, t_2, z_1, z_2) B_n^{(1)}(t_1, t_2, z_1, z_2) + o_p \left[ (nh_{n1})^{-1/2} \right] \quad (31) \\ &\equiv I_{n1}(t) + I_{n2}(t) + o_p \left[ (nh_{n1})^{-1/2} \right]\end{aligned}$$

uniformly over  $t \in [0, \tau]$ .

It now remains to evaluate the integrals in (31). Observe that by a change of variables and a Taylor series expansion,

$$\begin{aligned}I_{n1}(t) &= \frac{1}{nh_{n1}h_{nz}^2} \sum_{i=1}^n \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 C(t, t_2, z_1, z_2) 1(T_{i2} > t_2) \\ &\quad \times K_T \left( \frac{t - T_{i1}}{h_{n1}} \right) K_Z \left( \frac{z_1 - Z_{i1}}{h_{nz}} \right) K_Z \left( \frac{z_2 - Z_{i2}}{h_{nz}} \right) \\ &= \frac{1}{nh_{n1}} \sum_{i=1}^n \left[ \int_{S_T} C(t, t_2, Z_{i1}, Z_{i2}) 1(T_{i2} > t_2) dt_2 \right] K_T \left( \frac{t - T_{i1}}{h_{n1}} \right) + O(h_{nz}^r) \\ &\equiv \zeta_n^{(A1)}(t) + O(h_{nz}^r)\end{aligned}$$

uniformly over  $t \in [0, \tau]$ . Similarly,

$$\begin{aligned}
I_{n2}(t) &= -\frac{1}{nh_{n2}h_{nz}^2} \sum_{i=1}^n \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 D(t, t_2, z_1, z_2) 1(T_{i1} > t) \\
&\quad \times K_T \left( \frac{t_2 - T_{i2}}{h_{n2}} \right) K_Z \left( \frac{z_1 - Z_{i1}}{h_{nz}} \right) K_Z \left( \frac{z_2 - Z_{i2}}{h_{nz}} \right) \\
&= -\frac{1}{n} \sum_{i=1}^n D(t, T_{i2}, Z_{i1}, Z_{i2}) 1(T_{i1} > t) + O_p(h_{n2}^q) + O(h_{nz}^r) \\
&\equiv \zeta_n^{(B1)}(t) + O(h_{n2}^q) + O(h_{nz}^r)
\end{aligned}$$

uniformly over  $t \in [0, \tau]$ .

Combining these results with the condition that  $(nh_{n1})^{1/2}h_{nz}^r \rightarrow 0$  and  $(nh_{n1})^{1/2}h_{n2}^q \rightarrow 0$  gives

$$\lambda_{n0}(t) - \lambda_0(t) = \zeta_n^{(A1)}(t) + \zeta_n^{(B1)}(t) + o_p \left[ (nh_{n1})^{-1/2} \right] \quad (32)$$

uniformly over  $t \in [0, \tau]$ . It is straightforward to show that  $E[\zeta_n^{(B1)}(t)] = -\lambda_0(t)$ . Furthermore, it is not difficult to show that by Theorem 2.37 of Pollard (1984),

$$\zeta_n^{(B1)}(t) - E[\zeta_n^{(B1)}(t)] = o \left( \log n / n^{1/2} \right)$$

uniformly over  $t \in [0, \tau]$ . Therefore, (32) can be rewritten as

$$\lambda_{n0}(t) - \lambda_0(t) = \zeta_n^{(A1)}(t) - \lambda_0(t) + o_p \left[ (nh_{n1})^{-1/2} \right] \quad (33)$$

uniformly over  $t \in [0, \tau]$ . Using integration by parts and a change of variables, it is not difficult to show that

$$E[\zeta_n^{(A1)}(t) - \lambda_0(t)] = B_\lambda(t) + o(h_{n1}^q). \quad (34)$$

Part (a) now follows by combining (33)-(34).

**Part (b):** As in the proof of part (a), using Lemmas A.1 and A.2, it can be shown that

$$\begin{aligned}
&\frac{A_n(t_1, t_2, z_1, z_2)}{B_n(t_1, t_2, z_1, z_2)} - \frac{A(t_1, t_2, z_1, z_2)}{B(t_1, t_2, z_1, z_2)} \\
&= C(t_1, t_2, z_1, z_2) A_n^{(1)}(t_1, t_2, z_1, z_2) - D(t_1, t_2, z_1, z_2) B_n^{(1)}(t_1, t_2, z_1, z_2) \\
&\quad + \left[ C(t_1, t_2, z_1, z_2) A^{(2)}(t_1, t_2, z_1, z_2) - D(t_1, t_2, z_1, z_2) B^{(2)}(t_1, t_2, z_1, z_2) \right]' S_\Omega + o_p \left( n^{-1/2} \right)
\end{aligned} \quad (35)$$

uniformly over  $(t_1, t_2, z_1, z_2) \in [0, \tau] \times S_T \times S_Z \times S_Z$ . Observe that Assumption 3.9 (b) is necessary to ensure that the remainder term is of order  $o_p(n^{-1/2})$ .

It now remains to integrate the leading terms in (35) over  $(t_1, t_2, z_1, z_2)$ . It follows from the proof of part (a) that

$$\begin{aligned} & \int_0^t dt_1 \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 C(t_1, t_2, z_1, z_2) A_n^{(1)}(t_1, t_2, z_1, z_2) \\ &= \frac{1}{nh_{n1}} \sum_{i=1}^n \int_0^t \left[ \int_{S_T} C(t_1, t_2, Z_{i1}, Z_{i2}) 1(T_{i2} > t_2) dt_2 \right] K_T \left( \frac{t_1 - T_{i1}}{h_{n1}} \right) dt_1 + O(h_{nz}^r) \\ &\equiv Q_n^{(A)}(t) + O(h_{nz}^r) \end{aligned}$$

uniformly over  $t \in [0, \tau]$ .

Define  $\tilde{R}_n^{(A)}(t) = Q_n^{(A)}(t) - \tilde{Q}_n^{(A)}(t)$ , where

$$\tilde{Q}_n^{(A)}(t) = \frac{1}{n} \sum_{i=1}^n \left[ \int_{S_T} C(T_{i1}, t_2, Z_{i1}, Z_{i2}) 1(T_{i2} > t_2) dt_2 \right] 1(0 \leq T_{i1} \leq t).$$

By integration by parts, it is easy to show that  $E\tilde{Q}_n^{(A)}(t) = \Lambda_0(t)$ . Combining (34) with Fubini's theorem yields  $EQ_n^{(A)}(t) = \Lambda_0(t) + O_p(h_{n1}^q)$  uniformly over  $t \in [0, \tau]$ . Thus,  $E\tilde{R}_n^{(A)}(t) = O_p(h_{n1}^q)$  uniformly over  $t \in [0, \tau]$ . Furthermore, we can show that the summand in  $\tilde{R}_n^{(A)}(t)$  is Euclidean. Therefore, by Theorem 2.37 of Pollard (1984),

$$\sup_t \left| \tilde{R}_n^{(A)}(t) - E\tilde{R}_n^{(A)}(t) \right| = o \left[ h_{n1}^{1/2} (\log n) / n^{1/2} \right]$$

almost surely. Therefore,  $Q_n^{(A)}(t) = \tilde{Q}_n^{(A)}(t) + o_p(n^{-1/2})$  uniformly over  $t \in [0, \tau]$ .

Now consider the second term in (35). Again, by the result of the proof of part (a),

$$\begin{aligned} & \int_0^t dt_1 \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 D(t_1, t_2, z_1, z_2) B_n^{(1)}(t_1, t_2, z_1, z_2) \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^t D(t_1, T_{i2}, Z_{i1}, Z_{i2}) 1(T_{i1} > t_1) dt_1 + O(h_{n2}^q) + O(h_{nz}^r) \end{aligned}$$

uniformly over  $t \in [0, \tau]$ .

Finally, consider the remaining terms in (35). Use integration by parts and (2) to obtain

$$\begin{aligned} & \int_0^t dt_1 \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 CA^{(2)} - DB^{(2)} \\ &= EX_1 \int_0^t dt_1 \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 \frac{\partial}{\partial z_1} \varphi(t_2, z_1, z_2) \frac{A(t_1, t_2 | z_1, z_2)}{B(t_1, t_2 | z_1, z_2)} \\ &\quad + EX_2 \int_0^t dt_1 \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 \frac{\partial}{\partial z_2} \varphi(t_2, z_1, z_2) \frac{A(t_1, t_2 | z_1, z_2)}{B(t_1, t_2 | z_1, z_2)} \\ &= -\Lambda_0(t) E[X_1 - X_2] \int_{S_Z} dz_1 \int_{S_Z} dz_2 \frac{w_z(z_1) w_z(z_2)}{p_z(z_1, z_2)}. \end{aligned}$$

Part (b) follows by combining these results.  $\square$

*Proof of Corollary 3.2.* Part (a) follows from Theorem 3.1 (a) and an application of a triangular-array central limit theorem. It is not difficult to show that the summand in  $\Gamma_n(t)$  is Euclidean. Then part (b) can be easily proved by combining Theorem 3.1 (b) with the empirical process method described in Pollard (1984) and Pakes and Pollard (1989).  $\square$

## A.2 Estimation of $V_\lambda(t)$ and $E[\chi_\lambda(t)\chi_\lambda(t')]$

The asymptotic variance  $V_\lambda(t)$  and the covariance function  $E[\chi_\lambda(t)\chi_\lambda(t')]$  can be estimated consistently by replacing unknown quantities with sample analogs. Define

$$\begin{aligned}\varphi_n(t_2, z_1, z_2) &= p_{nz}(z_1, z_2)^{-1}w(t_2, z_1, z_2)\exp(z_2 - z_1), \\ C_n(t_1, t_2, z_1, z_2) &= B_n(t_1, t_2|z_1, z_2)^{-1}\varphi_n(t_2, z_1, z_2),\end{aligned}$$

and

$$D_n(t_1, t_2, z_1, z_2) = B_n(t_1, t_2|z_1, z_2)^{-2}A_n(t_1, t_2|z_1, z_2)\varphi_n(t_2, z_1, z_2),$$

where  $p_{nz}$ ,  $A_n$ , and  $B_n$  are defined in Section 2.1. Let  $\bar{X}_j$  be the sample means of  $X_j$  and let  $\Omega_n$  be a consistent estimator of  $\Omega$ . It is easy to obtain the formula for calculating  $\Omega_n$  corresponding to the partial likelihood estimator of  $\beta$ . Define

$$\gamma_{nt}(T_{i1}, T_{i2}, X_{i1}, X_{i2}) = \left[ \int_{S_T} C_n(t, t_2, Z_{ni1}, Z_{ni2})1(T_{i2} > t_2) dt_2 \right]^2 \frac{1}{h_{n1}} K_T \left( \frac{t - T_{i1}}{h_{n1}} \right),$$

and

$$\begin{aligned}\Gamma_{nt}(T_{i1}, T_{i2}, X_{i1}, X_{i2}) &= \left[ \int_{S_T} C_n(T_{i1}, t_2, Z_{ni1}, Z_{ni2})1(T_{i2} > t_2) dt_2 \right] 1(0 \leq T_{i1} \leq t) \\ &\quad - \int_0^t D_n(t_1, T_{i2}, Z_{ni1}, Z_{ni2})1(T_{i1} > t_1) dt_1 \\ &\quad - \Lambda_{n0}(t) \left[ \int_{S_Z} dz_1 \int_{S_Z} dz_2 \frac{w_z(z_1)w_z(z_2)}{p_{nz}(z_1, z_2)} \right] [\bar{X}_1 - \bar{X}_2]' \Omega_n(T_{i1}, T_{i2}, X_{i1}, X_{i2}).\end{aligned}$$

Using the fact that  $p_{nz}$ ,  $A_n$ ,  $B_n$ , and  $\Lambda_{n0}$  converge in probability uniformly, It is straightforward to show that under the assumptions of Theorem 3.1,  $V_\lambda(t)$  is estimated consistently by

$$n^{-1} \sum_{i=1}^n \gamma_{nt}(T_{i1}, T_{i2}, X_{i1}, X_{i2}) \int_{-1}^1 K_T^2(u) du$$

and that  $E[\chi_\lambda(t)\chi_\lambda(t')]$  is estimated consistently by

$$n^{-1} \sum_{i=1}^n \Gamma_{nt}(T_{i1}, T_{i2}, X_{i1}, X_{i2}) \Gamma_{nt'}(T_{i1}, T_{i2}, X_{i1}, X_{i2}).$$



## B Appendix: Censored Case

### B.1 The Asymptotic Distribution of $n^{1/2}(\beta_n - \beta)$

This section of Appendix B presents conditions under which  $n^{1/2}(\beta_n - \beta)$  is asymptotically normally distributed. In this section, let  $q$  and  $r$  be integers such that  $q \geq 2$  and  $r > dq/(q-1)$ . We maintain Assumptions 3.2-3.4 and make the following additional assumptions:

**Assumption 3.1' (Random Sampling).**  $\{(Y_{i1}, Y_{i2}, X_{i1}, X_{i2}, \Delta_{i1}, \Delta_{i2}, C_i) : i = 1, \dots, n\}$  is a random sample of  $(Y_1, Y_2, X_1, X_2, \Delta_1, \Delta_2, C)$ .

**Assumption 3.5' (Smoothness).** The distribution of  $(Y_1, Y_2, X_1, X_2, \Delta_1, \Delta_2)$  is absolutely continuous with respect to the product of Lebesgue measure on  $\mathbf{R}^{2(1+d)}$  and counting measure on  $\{0, 1\}^2$ .

Furthermore, there are an interval of the real line,  $I_T$ , and an open rectangle of  $\mathbf{R}^d$ ,  $I_X$ , such that

- (a)  $I_T = [0, \tau_T)$ , where  $\tau_T \leq \infty$ ,
- (b)  $f(t_1, t_2 | x_1, x_2)$  and  $p_x(x_1, x_2)$  are bounded on  $I_T \times I_T \times I_X \times I_X$ ,
- (c)  $f(t_1, t_2 | x_1, x_2)$  and  $p_x(x_1, x_2)$  are positive for all  $(t_1, t_2, x_1, x_2) \in \text{int}(I_T \times I_T \times I_X \times I_X)$ , and
- (d)  $f(t_1, t_2 | x_1, x_2)$  and  $p_x(x_1, x_2)$  have bounded partial derivatives up to order  $q$  with respect to  $t_j$ , and up to order  $r$  with respect to  $x_j$  for  $j = 1, 2$ .

The conditions in Assumption 3.5' are parallel to those in Section 3.

**Assumption 3.6' (Weight Functions).** (a) The weight function  $w_\beta(\cdot)$  is a bounded, non-negative function with compact support  $S_\beta \subset I_T$  such that  $\int_{S_\beta} w_\beta(t) dt = 1$  and  $w_\beta$  is  $q$  times continuously differentiable on  $S_\beta$ .

(a) The weight function  $w_x(\cdot)$  is a bounded, non-negative function with compact support  $S_X \subset I_X$  such that  $w_x$  is continuously differentiable on  $S_X$ .

**Assumption 3.8' (Kernels).** (a)  $K_T$  has support  $[-1, 1]$ , is bounded and symmetrical about 0, has bounded variation, and satisfies

$$\int_{-1}^1 w^j K_T(u) du = \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{if } j = 1 \leq j \leq q-1, \\ C_T & \text{if } j = q, \end{cases}$$

where  $C_T$  is a positive constant.

(b)  $K_X$  has support  $[-1, 1]^d$ , is bounded and symmetrical about 0, has bounded variation, and satisfies

$$\int_{[-1, 1]^d} w^j K_X(u) du = \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{if } j = 1 \leq j \leq r-1, \\ C_X & \text{if } j = r, \end{cases}$$

where  $C_X$  is a positive constant.

**Assumption 3.9' (Bandwidths).**  $nh_{n1}^{2q} \rightarrow 0$ ,  $nh_{n2}^{2q} \rightarrow 0$ ,  $nh_{nx}^{4r} \rightarrow 0$ ,  $nh_{nx}^{4d} \rightarrow \infty$ ,  $\log n / (nh_{n1}^2 h_{nx}^{4d})^{1/4} \rightarrow 0$ , and  $\log n / (nh_{n2}^2 h_{nx}^{4d})^{1/4} \rightarrow 0$ .

Assumptions 3.8' and 3.9' are satisfied, for example, if  $K_T$  is a fourth-order kernel,  $K_X$  is a  $r$ -th-order kernel,  $h_{n1} = h_{n2} \propto n^{-1/7}$ , and  $h_{nx} \propto n^{-\kappa_x}$ , where  $1/(4r) < \kappa_x < 5/(28d)$ .

**Assumption B.1 (Censoring).** *The censoring threshold  $C$  is independent of  $T_1$  and  $T_2$  given  $X_1$  and  $X_2$ . The conditional distribution of  $C$  given  $X_1 = x_1$  and  $X_2 = x_2$  is absolutely continuous with respect to Lebesgue measure for all  $x_1$  and  $x_2$ . Furthermore,  $G(c|x_1, x_2)$  is positive for every  $(c, x_1, x_2)$ , and  $G(c|x_1, x_2)$  is continuously differentiable with respect to  $x_1$  and  $x_2$  for each  $c$ .*

**Assumption B.2 (Full Rank Condition).** *The matrix  $\Phi_\beta \equiv E[w_x(X_1)w_x(X_2)\Delta X\Delta X']$  is non-singular.*

Define

$$\begin{aligned} & \tilde{\Omega}(Y_{i1}, Y_{i2}, X_{i1}, X_{i2}, \Delta_{i1}, \Delta_{i2}) \\ &= \Phi_\beta^{-1} \frac{\Delta_{i1}\Delta_{i2} w_{xi}\Delta X_i}{\exp(\Delta X_i'\beta) G(Y_{i1} + Y_{i2}|X_{i2}, X_{i2})} \left[ \frac{w_\beta(Y_{i1})1(Y_{i2} > Y_{i1})}{B_\beta(Y_{i1}|X_{i1}, X_{i2})} - \frac{w_\beta(Y_{i2})\exp(\Delta X_i'\beta)1(Y_{i1} > Y_{i2})}{B_\beta(Y_{i2}|X_{i1}, X_{i2})} \right]. \end{aligned}$$

The following proposition provides the main result of this section.

**Proposition B.1.** *Let Assumptions 3.1', 3.2-3.4, 3.5', 3.6', 3.8', 3.9', and B.1-B.2 hold. As  $n \rightarrow \infty$ ,*

$$\beta_n - \beta = \frac{1}{n} \sum_{i=1}^n \tilde{\Omega}(Y_{i1}, Y_{i2}, X_{i1}, X_{i2}, \Delta_{i1}, \Delta_{i2}) + o_p(n^{-1/2}).$$

*In particular,  $n^{1/2}(\beta_n - \beta)$  is asymptotically normal with mean zero and covariance matrix  $V_\beta \equiv E[\tilde{\Omega}(Y_1, Y_2, X_1, X_2, \Delta_1, \Delta_2)\tilde{\Omega}(Y_1, Y_2, X_1, X_2, \Delta_1, \Delta_2)']$ .*

The covariance matrix  $V_\beta$  can be estimated consistently by a sample analog estimator:

$$V_{n\beta} = n^{-1} \sum_{i=1}^n \tilde{\Omega}_n(Y_{i1}, Y_{i2}, X_{i1}, X_{i2}, \Delta_{i1}, \Delta_{i2})\tilde{\Omega}_n(Y_{i1}, Y_{i2}, X_{i1}, X_{i2}, \Delta_{i1}, \Delta_{i2})',$$

where  $\Phi_{n\beta} = n^{-1} \sum_{i=1}^n w_x(X_{i1})w_x(X_{i2})\Delta X_i\Delta X_i'$  and

$$\begin{aligned} & \tilde{\Omega}_n(Y_{i1}, Y_{i2}, X_{i1}, X_{i2}, \Delta_{i1}, \Delta_{i2}) \\ &= \Phi_{n\beta}^{-1} \frac{\Delta_{i1}\Delta_{i2} w_{xi}\Delta X_i}{\exp(\Delta X_i'\beta_n) G_n(Y_{i1} + Y_{i2}|X_{i2}, X_{i2})} \left[ \frac{w_\beta(Y_{i1})1(Y_{i2} > Y_{i1})}{B_{n\beta}(Y_{i1}|X_{i1}, X_{i2})} - \frac{w_\beta(Y_{i2})\exp(\Delta X_i'\beta_n)1(Y_{i1} > Y_{i2})}{B_{n\beta}(Y_{i2}|X_{i1}, X_{i2})} \right]. \end{aligned}$$

*Proof of Proposition B.1.* Define  $\tilde{A}_\beta(t, x_1, x_2) = \tilde{A}_\beta(t|x_1, x_2)p_x(x_1, x_2)$ ,  $\tilde{B}_\beta(t, x_1, x_2) = \tilde{B}_\beta(t|x_1, x_2) \times p_x(x_1, x_2)$ ,  $\tilde{A}_{n\beta}(t, x_1, x_2) = \tilde{A}_{n\beta}(t|x_1, x_2)p_{nx}(x_1, x_2)$ , and  $\tilde{B}_{n\beta}(t, x_1, x_2) = \tilde{B}_{n\beta}(t|x_1, x_2)p_{nx}(x_1, x_2)$ . Equation (14) can be rewritten as

$$\tilde{R}_{n\beta}(t|x_1, x_2) = \tilde{A}_{n\beta}(t, x_1, x_2)/\tilde{B}_{n\beta}(t, x_1, x_2). \quad (36)$$

As in the uncensored case, it is more convenient to use (36) than (14). We will split the proof into several steps.

**Step 1.** We first establish asymptotic linear approximations of  $\tilde{A}_{n\beta}(t, x_1, x_2)$  and  $\tilde{B}_{n\beta}(t, x_1, x_2)$ . Write

$$\tilde{A}_{n\beta}(t, x_1, x_2) = \tilde{A}_{n\beta}^{(1)}(t, x_1, x_2) + \tilde{A}_{n\beta}^{(2)}(t, x_1, x_2)$$

and

$$\tilde{B}_{n\beta}(t, x_1, x_2) = \tilde{B}_{n\beta}^{(1)}(t, x_1, x_2) + \tilde{B}_{n\beta}^{(2)}(t, x_1, x_2),$$

where

$$\begin{aligned} \tilde{A}_{n\beta}^{(1)}(t, x_1, x_2) &= \frac{1}{nh_{n1}h_{nx}^{2d}} \sum_{i=1}^n \frac{\Delta_{i1}\Delta_{i2}1(Y_{i2} > t)}{G(Y_{i1} + Y_{i2}|X_{i1}, X_{i2})} K_T \left( \frac{t - Y_{i1}}{h_{n1}} \right) K_X \left( \frac{x_1 - X_{i1}}{h_{nx}} \right) K_X \left( \frac{x_2 - X_{i2}}{h_{nx}} \right) \\ \tilde{A}_{n\beta}^{(2)}(t, x_1, x_2) &= -\frac{1}{nh_{n1}h_{nx}^{2d}} \sum_{i=1}^n \frac{\Delta_{i1}\Delta_{i2}1(Y_{i2} > t)}{G(Y_{i1} + Y_{i2}|X_{i1}, X_{i2})} K_T \left( \frac{t - Y_{i1}}{h_{n1}} \right) K_X \left( \frac{x_1 - X_{i1}}{h_{nx}} \right) K_X \left( \frac{x_2 - X_{i2}}{h_{nx}} \right) \\ &\quad \times G_n^{-1}(Y_{i1} + Y_{i2}|X_{i1}, X_{i2}) [G_n(Y_{i1} + Y_{i2}|X_{i1}, X_{i2}) - G(Y_{i2} + Y_{i2}|X_{i1}, X_{i2})], \\ \tilde{B}_{n\beta}^{(1)}(t, x_1, x_2) &= \frac{1}{nh_{n2}h_{nx}^{2d}} \sum_{i=1}^n \frac{\Delta_{i1}\Delta_{i2}1(Y_{i1} > t)}{G(Y_{i1} + Y_{i2}|X_{i1}, X_{i2})} K_T \left( \frac{t - Y_{i2}}{h_{n2}} \right) K_X \left( \frac{x_1 - X_{i1}}{h_{nx}} \right) K_X \left( \frac{x_2 - X_{i2}}{h_{nx}} \right), \end{aligned}$$

and

$$\begin{aligned} \tilde{B}_{n\beta}^{(2)}(t, x_1, x_2) &= -\frac{1}{nh_{n2}h_{nx}^{2d}} \sum_{i=1}^n \frac{\Delta_{i1}\Delta_{i2}1(Y_{i1} > t)}{G^2(Y_{i1} + Y_{i2}|X_{i1}, X_{i2})} K_T \left( \frac{t - Y_{i2}}{h_{n2}} \right) K_X \left( \frac{x_1 - X_{i1}}{h_{nx}} \right) K_X \left( \frac{x_2 - X_{i2}}{h_{nx}} \right) \\ &\quad \times G_n^{-1}(Y_{i1} + Y_{i2}|X_{i1}, X_{i2}) [G_n(Y_{i1} + Y_{i2}|X_{i1}, X_{i2}) - G(Y_{i2} + Y_{i2}|X_{i1}, X_{i2})]. \end{aligned}$$

Observe that  $G(Y_{i1} + Y_{i2}|X_{i1}, X_{i2})$  is bounded away from zero as long as  $\Delta_{i1} = \Delta_{i2} = 1$ . Thus,  $G_n^{-1}(Y_{i1} + Y_{i2}|X_{i1}, X_{i2}) = O_p(1)$  uniformly in  $\{i : \Delta_{i1} = \Delta_{i2} = 1\}$ . Combining this with uniform consistency of  $G_n$  to  $G$  on a compact set gives

$$\tilde{A}_{n\beta}^{(2)}(t, x_1, x_2) = \tilde{A}_{n\beta}^{(1)}(t, x_1, x_2)[1 + o_p(1)]$$

and

$$\tilde{B}_{n\beta}^{(2)}(t, x_1, x_2) = \tilde{B}_{n\beta}^{(1)}(t, x_1, x_2)[1 + o_p(1)]$$

uniformly over  $(t, x_1, x_2) \in S_T \times S_X \times S_X$ . In addition, arguments similar to those used in the proof of Lemma A.1, it can be shown that

$$\tilde{A}_{n\beta}^{(1)}(t, x_1, x_2) = \tilde{A}_\beta(t, x_1, x_2) + O(h_{n1}^q) + O(h_{nx}^r) + o\left[\log n / (nh_{n1}h_{nx}^{2d})^{1/2}\right] \quad a.s.$$

and

$$\tilde{B}_{n\beta}^{(1)}(t, x_1, x_2) = \tilde{B}_\beta(t, x_1, x_2) + O(h_{n2}^q) + O(h_{nx}^r) + o\left[\log n / (nh_{n2}h_{nx}^{2d})^{1/2}\right] \quad a.s.$$

uniformly over  $(t, x_1, x_2) \in S_T \times S_X \times S_X$ .

**Step 2.** Using the fact that  $\log V_i = \Delta X_i' \beta$ , write

$$\beta_n - \beta = \left( n^{-1} \sum_{i=1}^n w_{xi} \Delta X_i \Delta X_i' \right)^{-1} \left( n^{-1} \sum_{i=1}^n w_{xi} \Delta X_i [\log V_{ni} - \log V_i] \right). \quad (37)$$

By a Taylor series expansion,

$$\log V_{ni} - \log V_i = V_i^{-1}(V_{ni} - V_i) + O_p[(V_{ni} - V_i)^2]. \quad (38)$$

Observe that by a Taylor series expansion, the result of Step 1, and Assumption 3.9',

$$\begin{aligned} V_{ni} - V_i &= \int_{S_\beta} w_\beta(t) [R_{n\beta}(t, X_{i1}, X_{i2}) - R_\beta(t, X_{i1}, X_{i2})] dt \\ &= \int_{S_\beta} \frac{w_\beta(t)}{B_\beta(t, X_{i1}, X_{i2})} \left[ A_{n\beta}^{(1)}(t, X_{i1}, X_{i2}) - \exp(\Delta X_i' \beta) B_{n\beta}^{(1)}(t, X_{i1}, X_{i2}) \right] dt [1 + o_p(1)] + o_p(n^{-1/2}) \\ &\equiv I_{n\beta i} [1 + o_p(1)] + o_p(n^{-1/2}) \end{aligned}$$

uniformly over  $(X_{i1}, X_{i2}) \in S_X \times S_X$ . By a change of variables and a Taylor series expansion,

$$\begin{aligned} I_{n\beta i} &= \frac{1}{nh_{nx}^{2d}} \sum_{j=1}^n \frac{\Delta_{j1} \Delta_{j2}}{G(Y_{j1} + Y_{j2} | X_{j1}, X_{j2})} K_X \left( \frac{X_{i1} - X_{j1}}{h_{nx}} \right) K_X \left( \frac{X_{i2} - X_{j2}}{h_{nx}} \right) \\ &\quad \times \left[ \frac{w_\beta(Y_{j1}) 1(Y_{j2} > Y_{j1})}{B_\beta(Y_{j1}, X_{i1}, X_{i2})} - \frac{w_\beta(Y_{j2}) \exp(\Delta X_i' \beta) 1(Y_{j1} > Y_{j2})}{B_\beta(Y_{j2}, X_{i1}, X_{i2})} \right] + O(h_{n1}^q) + O(h_{n2}^q) \\ &\equiv \tilde{I}_{n\beta i} + O(h_{n1}^q) + O(h_{n2}^q) \end{aligned}$$

uniformly over  $(X_{i1}, X_{i2}) \in S_X \times S_X$ . Therefore, we have

$$V_{ni} - V_i = \tilde{I}_{n\beta i} [1 + o_p(1)] + o_p(n^{-1/2})$$

uniformly over  $(X_{i1}, X_{i2}) \in S_X \times S_X$ .

**Step 3.** Combining the result of Step 2 with (38) gives

$$\begin{aligned}
& n^{-1} \sum_{i=1}^n w_{xi} \Delta X_i [\log V_{ni} - \log V_i] \\
&= \frac{1}{n} \sum_{i=1}^n \frac{w_{xi} \Delta X_i}{\exp(\Delta X'_i \beta)} \tilde{I}_{n\beta i} [1 + o_p(1)] + o_p(n^{-1/2}) \\
&= \frac{1}{n^2 h_{nx}^{2d}} \sum_{i=1}^n \sum_{j=1}^n \frac{w_{xi} \Delta X_i}{\exp(\Delta X'_i \beta)} \frac{\Delta_{j1} \Delta_{j2}}{G(Y_{j1} + Y_{j2} | X_{j1}, X_{j2})} K_X \left( \frac{X_{i1} - X_{j1}}{h_{nx}} \right) K_X \left( \frac{X_{i2} - X_{j2}}{h_{nx}} \right) \\
&\quad \times \left[ \frac{w_\beta(Y_{j1}) 1(Y_{j2} > Y_{j1})}{B_\beta(Y_{j1}, X_{i1}, X_{i2})} - \frac{w_\beta(Y_{j2}) \exp(\Delta X'_i \beta) 1(Y_{j1} > Y_{j2})}{B_\beta(Y_{j2}, X_{i1}, X_{i2})} \right] [1 + o_p(1)] + o_p(n^{-1/2}) \\
&\equiv \frac{1}{n^2 h_{nx}^{2d}} \sum_{i=1}^n \sum_{j=1}^n \xi_{ij} [1 + o_p(1)] + o_p(n^{-1/2}).
\end{aligned}$$

Write further the leading term as

$$\begin{aligned}
\frac{1}{n^2 h_{nx}^{2d}} \sum_{i=1}^n \sum_{j=1}^n \xi_{ij} &= \frac{1}{n^2 h_{nx}^{2d}} \sum_{j=1}^n \sum_{i=1, i \neq j}^n \xi_{ij} + \frac{1}{n^2 h_{nx}^{2d}} \sum_{i=1}^n \xi_{ii} \\
&\equiv I_{n\beta 1} + I_{n\beta 2}.
\end{aligned}$$

The order of  $I_{n\beta 2}$  is at most of order  $O_p[1/(nh_{nx}^{2d})]$ , so that using Assumption 3.9',  $I_{n\beta 2} = o_p(n^{-1/2})$ .

In particular, we require here that  $nh_{nx}^{4d} \rightarrow \infty$ . To deal with  $I_{n\beta 1}$ , observe

$$\begin{aligned}
I_{n\beta 1} &= \frac{1}{n} \sum_{j=1}^n \frac{\Delta_{j1} \Delta_{j2}}{G(Y_{j1} + Y_{j2} | X_{j1}, X_{j2})} \\
&\quad \times \frac{1}{n h_{nx}^{2d}} \sum_{i=1, i \neq j}^n \frac{w_{xi} \Delta X_i}{\exp(\Delta X'_i \beta)} K_X \left( \frac{X_{i1} - X_{j1}}{h_{nx}} \right) K_X \left( \frac{X_{i2} - X_{j2}}{h_{nx}} \right) \\
&\quad \times \left[ \frac{w_\beta(Y_{j1}) 1(Y_{j2} > Y_{j1})}{B_\beta(Y_{j1}, X_{i1}, X_{i2})} - \frac{w_\beta(Y_{j2}) \exp(\Delta X'_i \beta) 1(Y_{j1} > Y_{j2})}{B_\beta(Y_{j2}, X_{i1}, X_{i2})} \right] \\
&= \frac{1}{n} \sum_{j=1}^n \frac{\Delta_{j1} \Delta_{j2}}{G(Y_{j1} + Y_{j2} | X_{j1}, X_{j2})} \frac{w_{xj} \Delta X_j}{\exp(\Delta X'_j \beta)} p_x(X_{j1}, X_{j2}) \\
&\quad \times \left[ \frac{w_\beta(Y_{j1}) 1(Y_{j2} > Y_{j1})}{B_\beta(Y_{j1}, X_{j1}, X_{j2})} - \frac{w_\beta(Y_{j2}) \exp(\Delta X'_j \beta) 1(Y_{j1} > Y_{j2})}{B_\beta(Y_{j2}, X_{j1}, X_{j2})} \right] [1 + o_p(1)]
\end{aligned}$$

by using arguments similar to those used to prove the uniform consistency of the kernel density estimator. The proposition follows easily by combining the result of this step with (37).  $\square$

## B.2 Asymptotic Properties of $\lambda_{n0}$ and $\Lambda_{n0}$

This section of Appendix B presents conditions under which the estimators of  $\lambda_0$  and  $\Lambda_0$  in Section 2.2 are uniformly consistent and asymptotically normally distributed. We maintain Assumptions 3.1', 3.2-3.4, 3.6, and 3.8-3.9 and make the following additional assumptions:

**Assumption 3.5'' (Smoothness).** *The distribution of  $(Y_1, Y_2, Z_1, Z_2, \Delta_1, \Delta_2)$  is absolutely continuous with respect to the product of Lebesgue measure on  $\mathbf{R}^4$  and counting measure on  $\{0, 1\}^2$ .*

*Furthermore, there are intervals of the real line,  $I_T$  and  $I_Z$ , such that*

- (a)  $I_T = [0, \tau_T)$ , where  $\tau_T \leq \infty$ , and  $I_Z$  is open,
- (b)  $f(t_1, t_2|z_1, z_2)$  and  $p_z(z_1, z_2)$  are bounded on  $I_T \times I_T \times I_Z \times I_Z$ ,
- (c)  $f(t_1, t_2|z_1, z_2)$  and  $p_z(z_1, z_2)$  are positive for all  $(t_1, t_2, z_1, z_2) \in \text{int}(I_T \times I_T \times I_Z \times I_Z)$ , and
- (d)  $f(t_1, t_2|z_1, z_2)$  and  $p_z(z_1, z_2)$  have bounded partial derivatives up to order  $q$  with respect to  $t_j$  and up to order  $r$  with respect to  $z_j$  for  $j = 1, 2$ .

**Assumption 3.7' (Estimator of  $\beta$ ).** *There is a  $d \times 1$ -vector-valued function  $\tilde{\Omega}(y_1, y_2, x_1, x_2, \delta_1, \delta_2)$  such that*

- (a)  $E\tilde{\Omega}(Y_1, Y_2, X_1, X_2, \Delta_1, \Delta_2) = 0$ ,
- (b) the components of  $E[\tilde{\Omega}(Y_1, Y_2, X_1, X_2, \Delta_1, \Delta_2)\tilde{\Omega}(Y_1, Y_2, X_1, X_2, \Delta_1, \Delta_2)']$  are finite, and
- (c) as  $n \rightarrow \infty$ ,

$$\beta_n - \beta = n^{-1} \sum_{i=1}^n \tilde{\Omega}(Y_{i1}, Y_{i2}, X_{i1}, X_{i2}, \Delta_{i1}, \Delta_{i2}) + o_p(n^{-1/2}).$$

This assumption is satisfied by  $\beta_n$ , as was shown in Proposition B.1.

**Assumption B.1' (Censoring).** *The censoring threshold  $C$  is independent of  $(T_1, T_2, X_1, X_2)$ . The distribution of  $C$  is absolutely continuous with respect to Lebesgue measure. Furthermore,  $G(c)$  is positive for every  $c$ .*

Define

$$\tilde{C}(t_1, t_2, z_1, z_2) = \tilde{B}(t_1, t_2|z_1, z_2)^{-1} \varphi(t_2, z_1, z_2)$$

and

$$\tilde{D}(t_1, t_2, z_1, z_2) = \tilde{B}(t_1, t_2|z_1, z_2)^{-2} \tilde{A}(t_1, t_2|z_1, z_2) \varphi(t_2, z_1, z_2).$$

Define

$$\begin{aligned} & \tilde{\Gamma}_t(Y_{i1}, Y_{i2}, X_{i1}, X_{i2}, \Delta_{i1}, \Delta_{i2}) \\ &= \frac{\Delta_{i1} \Delta_{i2}}{G(Y_{i1} + Y_{i2})} \left\{ \left[ \int_{S_T} \tilde{C}(Y_{i1}, t_2, Z_{i1}, Z_{i2}) 1(Y_{i2} > t_2) dt_2 \right] 1(0 \leq Y_{i1} \leq t) \right. \\ & \quad \left. - \int_0^t \tilde{D}(t_1, Y_{i2}, Z_{i1}, Z_{i2}) 1(Y_{i1} > t_1) dt_1 \right\} \\ & \quad - \Lambda_0(t) \left[ \int_{S_Z} dz_1 \int_{S_Z} dz_2 \frac{w_z(z_1) w_z(z_2)}{p_z(z_1, z_2)} \right] E[X_1 - X_2]' \tilde{\Omega}(Y_{i1}, Y_{i2}, X_{i1}, X_{i2}, \Delta_{i1}, \Delta_{i2}). \end{aligned}$$

In addition, define

$$\begin{aligned}\tilde{B}_\lambda(t) &= \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 \left[ \int_{S_T} \tilde{C}(t, s_2, z_1, z_2) 1(t_2 > s_2) ds_2 \right] \frac{\partial^q}{\partial t_1^q} \frac{f(t_1, t_2, z_1, z_2)}{G(t_1 + t_2)} \Big|_{t_1=t} \\ &\quad \times \frac{1}{q!} \int_{-1}^1 u^q K_T(u) du\end{aligned}$$

and

$$\begin{aligned}\tilde{V}_\lambda(t) &= \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 \left[ \int_{S_T} \tilde{C}(t, s_2, z_1, z_2) 1(t_2 > s_2) ds_2 / G(t + t_2) \right]^2 f(t, t_2, z_1, z_2) \\ &\quad \times \int_{-1}^1 K_T^2(u) du.\end{aligned}$$

The following proposition gives the main result of this section.

**Proposition B.2.** *Let Assumptions 3.1', 3.2-3.4, 3.5'', 3.6, 3.7', 3.8-3.9, and B.1' hold.*

(a) *Assume  $h_{n1} \propto n^{-1/(2q+1)}$ . For  $t \in [0, \tau]$ ,*

$$n^{q/(2q+1)}[\lambda_{n0}(t) - \lambda_0(t)] \rightarrow_d \mathbf{N} \left( \tilde{B}_\lambda(t), \tilde{V}_\lambda(t) \right).$$

(b) *On  $[0, \tau]$ ,*

$$n^{1/2}[\Lambda_{n0}(t) - \Lambda_0(t)] \Rightarrow \tilde{\chi}_\Lambda(t),$$

where  $\tilde{\chi}_\Lambda(t)$  is a tight Gaussian process with mean 0 and covariance function  $E[\tilde{\chi}_\Lambda(t)\tilde{\chi}_\Lambda(t')] = E[\tilde{\Gamma}_t(T_1, T_2, X_1, X_2)\tilde{\Gamma}_{t'}(T_1, T_2, X_1, X_2)]$ .

As in the uncensored case, the asymptotic variance  $\tilde{V}_\lambda$  of  $\lambda_{n0}$  and the covariance function of  $\tilde{\chi}_\Lambda$  can be estimated consistently by replacing unknown quantities with sample analogs.

*Proof of Proposition B.2.* The proof of Proposition B.2 is similar to those of Theorem 3.1 and Proposition B.1. We will only indicate the differences. Define  $\tilde{A}(t_1, t_2, z_1, z_2) = \tilde{A}(t_1, t_2 | z_1, z_2) p_z(z_1, z_2)$ ,  $\tilde{B}(t_1, t_2, z_1, z_2) = \tilde{B}(t_1, t_2 | z_1, z_2) p_z(z_1, z_2)$ ,  $\tilde{A}_n(t_1, t_2, z_1, z_2) = \tilde{A}_n(t_1, t_2 | z_1, z_2) p_{nz}(z_1, z_2)$ , and  $\tilde{B}_n(t_1, t_2, z_1, z_2) = \tilde{B}_n(t_1, t_2 | z_1, z_2) p_{nz}(z_1, z_2)$ . Equation (17) can be rewritten as

$$\tilde{R}_n(t_1, t_2 | z_1, z_2) = \tilde{A}_n(t_1, t_2, z_1, z_2) / \tilde{B}_n(t_1, t_2, z_1, z_2). \quad (39)$$

As before, it is more convenient to use (39) than (17).

**Part (a):** By the definition and a Taylor series expansion,

$$\begin{aligned}\lambda_{n0}(t) - \lambda_0(t) &= \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 w(t_2, z_1, z_2) \exp(z_2 - z_1) \left[ \frac{\tilde{A}_n(t_1, t_2, z_1, z_2)}{\tilde{B}_n(t_1, t_2, z_1, z_2)} - \frac{\tilde{A}(t_1, t_2, z_1, z_2)}{\tilde{B}(t_1, t_2, z_1, z_2)} \right] \\ &= \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 \\ &\quad \times \left[ \tilde{C}(t_1, t_2, z_1, z_2) \tilde{A}_n(t_1, t_2, z_1, z_2) - \tilde{D}(t_1, t_2, z_1, z_2) \tilde{B}_n(t_1, t_2, z_1, z_2) + \tilde{R}_n^{(\lambda)}(t_1, t_2, z_1, z_2) \right],\end{aligned}$$

where  $\tilde{R}_n^{(\lambda)}(t_1, t_2, z_1, z_2)$  is a remainder term.

Define

$$\tilde{A}_n^{(1)}(t_1, t_2, z_1, z_2) = \frac{1}{nh_{n1}h_{nz}^2} \sum_{i=1}^n \frac{\Delta_{i1}\Delta_{i2}1(Y_{i2} > t_2)}{G(Y_{i1} + Y_{i2})} K_T \left( \frac{t_1 - Y_{i1}}{h_{n1}} \right) K_Z \left( \frac{z_1 - Z_{i1}}{h_{nz}} \right) K_Z \left( \frac{z_2 - Z_{i2}}{h_{nz}} \right)$$

and

$$\tilde{B}_n^{(1)}(t_1, t_2, z_1, z_2) = \frac{1}{nh_{n2}h_{nz}^2} \sum_{i=1}^n \frac{\Delta_{i1}\Delta_{i2}1(Y_{i1} > t_1)}{G(Y_{i1} + Y_{i2})} K_T \left( \frac{t_2 - Y_{i2}}{h_{n2}} \right) K_Z \left( \frac{z_1 - Z_{i1}}{h_{nz}} \right) K_Z \left( \frac{z_2 - Z_{i2}}{h_{nz}} \right).$$

By arguments similar to those used in the proofs of Lemmas A.1-A.2 and Proposition B.1, it can be shown that

$$\begin{aligned} \lambda_{n0}(t) - \lambda_0(t) &= \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 \tilde{C}(t_1, t_2, z_1, z_2) \tilde{A}_n^{(1)}(t_1, t_2, z_1, z_2) \\ &\quad - \int_{S_T} dt_2 \int_{S_Z} dz_1 \int_{S_Z} dz_2 \tilde{D}(t_1, t_2, z_1, z_2) \tilde{B}_n^{(1)}(t_1, t_2, z_1, z_2) + o_p \left[ (nh_{n1})^{-1/2} \right] \end{aligned}$$

uniformly over  $t \in [0, \tau]$ .

Repeating the same arguments given in the proof of Theorem 3.1 (a) gives

$$\begin{aligned} \lambda_{n0}(t) - \lambda_0(t) &= \frac{1}{nh_{n1}} \sum_{i=1}^n \frac{\Delta_{i1}\Delta_{i2}}{G(Y_{i1} + Y_{i2})} \left[ \int_{S_T} \tilde{C}(t, t_2, Z_{i1}, Z_{i2}) 1(Y_{i2} > t_2) dt_2 \right] K_T \left( \frac{t - Y_{i1}}{h_{n1}} \right) - \lambda_0(t) \\ &\quad + o_p \left[ (nh_{n1})^{-1/2} \right]. \end{aligned}$$

Then part (a) follows easily.

**Part (b):** This can be proved by repeating arguments similar to those used to prove part (b) of Theorem 3.1.  $\square$

### B.3 Alternative Estimator of $R_\beta(t|x_1, x_2)$

This part of Appendix B provides an alternative estimator of  $R_\beta(t|x_1, x_2)$ . There may be several methods for estimating  $R_\beta(t|x_1, x_2)$  under dependent right censoring, but we present here an alternative estimator of  $R_\beta(t|x_1, x_2)$  based on Visser (1996) and Wang and Wells (1998). See Lin, Sun, and Ying (1999) and Wang and Wells (1998) for more possible methods. The same idea as those described here can be applied to estimate  $R(t_1, t_2|z_1, z_2)$  in Section 2.2.2.

To describe an alternative estimator of  $R_\beta(t|x_1, x_2)$ , it is useful to introduce some notation. Define the conditional distribution functions  $F_1(t_1|x_1) = \Pr(T_1 \leq t_1|X_1 = x_1)$  and  $F_2(t_2|t_1, x_2) =$



$\Pr(T_2 \leq t_2 | T_1 = t_1, X_2 = x_2)$ . Also, define  $f_1(t_1|x_1) = \partial F_1(t_1|x_1)/\partial t_1$  and  $f_2(t_2|t_1, x_2) = \partial F_2(t_2|t_1, x_2)/\partial t_2$ . Using the fact that

$$S(t_1, t_2|z_1, z_2) = \int_{t_1}^{\infty} [1 - F_2(t_2|s_1, x_2)] dF_1(s_1|x_1),$$

write

$$R_\beta(t|x_1, z_2) = \frac{[1 - F_2(t|t, x_2)]f_1(t|x_1)}{\int_t^{\infty} f_2(t|s_1, x_2)dF_1(s_1|x_1)}. \quad (40)$$

An alternative estimator of  $R_\beta(t|x_1, x_2)$  can be obtained by replacing  $f_1, f_2, F_1$ , and  $F_2$  in (40) with their sample analogs.  $F_1$  can be estimated by using the conditional Kaplan-Meier estimator. Although  $C_2$  is dependent on  $T_2$ , the conditional Kaplan-Meier estimator can also be used to estimate  $F_2$ . This is because  $C_2$  is conditionally independent of  $T_2$  given  $T_1$  and  $X_2$ . It is worthwhile to observe that Kaplan-Meier-type estimators are step functions, thereby implying that  $f_1$  and  $f_2$  cannot be estimated by  $dF_{n1}(t_1|x_1)/dt_1$  and  $dF_{n2}(t_2|t_1, x_2)/dt_2$ . However, it is not difficult to develop consistent estimators of  $f_1$  and  $f_2$  based on the kernel method. See Dabrowska (1987, 1989) for the details of the conditional Kaplan-Meier estimator.

There are advantages and disadvantages to using this alternative estimator as opposed to the estimator of  $R_\beta(t|x_1, x_2)$  in Section 2.2.1. The advantages are that (1) the alternative estimator uses more data than the proposed estimator in Section 2.2.1, and (2) the censoring variable  $C$  does not have to be random; however, the disadvantages are that (1) the alternative estimator is computationally burdensome, (2) it is more complicated to derive asymptotic properties, and (3) it is difficult to extend to the case of longer panels. We chose to use the estimator in Section 2.2.1 mainly because of its attractive simple form.

## References

- Abbring, J.H, P.A, Chiappori, and J. Pinquet, 2003, Moral hazard and dynamic insurance data, forthcoming, Journal of the European Economic Association.
- Burke, M.D., 1988, Estimation of a bivariate distribution function under random censorship, *Biometrika*, 75, 379-382.
- Chamberlain, G., 1985, Heterogeneity, omitted variable bias, and duration dependence, in: J.J. Heckman and B. Singer, eds. *Longitudinal analysis of labor market data* (Cambridge University Press, Cambridge) 3-38.

- Cox, D.R., 1972, Regression models and life tables, *Journal of the Royal Statistical Society, Series B*, 34, 187-220.
- Dabrowska, D., 1987, Non-parametric regression with censored survival time data, *Scandinavian Journal of Statistics*, 14, 181-197.
- Dabrowska, D., 1989, Uniform consistency of the kernel conditional Kaplan-Meier estimate, *Annals of Statistics*, 17, 1157-1167.
- Efromovich, S., 2001, Density estimation under random censorship and order restrictions: from asymptotic to small samples, *Journal of the American Statistical Association*, 96, 667-684.
- Gørgens, T., and J.L. Horowitz, 1999, Semiparametric estimation of a censored regression model with an unknown transformation of the dependent variable, *Journal of Econometrics*, 90, 155-191.
- Hahn, J., 1994, The efficiency bound of the mixed proportional hazard model, *Review of Economic Studies*, 61, 607-629.
- Honoré, B.E., 1993, Identification results for duration models with multiple spells, *Review of Economic Studies*, 60, 241-246.
- 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* (Springer-Verlag, New York).
- 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 W. Härdle, 1996, Direct semiparametric estimation of single-index models with discrete covariates, *Journal of the American Statistical Association*, 91, 1632-1640.
- Kalbfleisch, J.D., and R.L. Prentice, 1980, *The statistical analysis of failure time data* (Wiley, New York).
- Lancaster, T., 2000, The incidental parameter problem since 1948, *Journal of Econometrics*, 95, 391-413.

- Lin, D.Y., W. Sun, and Z. Ying, 1999, Nonparametric estimation of the gap time distributions for serial events with censored data, *Biometrika*, 86, 59-70.
- Linton, O.B., and W. Härdle, 1996, Estimating additive regression models with known links, *Biometrika*, 83, 529-540.
- Müller, H.-G., 1984, Smooth optimal kernel estimators of densities, regression curves and modes, *Annals of Statistics*, 12, 766-774.
- Pakes, A., and D. Pollard, 1989, Simulation and the asymptotics of optimization estimators, *Econometrica*, 57, 1027-1057.
- Pollard, D., 1984, *Convergence of stochastic processes* (Springer-Verlag, New York).
- Powell, J.L., J.H. Stock, and T.M. Stoker, 1989, Semiparametric estimation of index coefficients, *Econometrica*, 57, 474-523.
- Ridder, G., and İ. Tunalı, 1999, Stratified partial likelihood estimation, *Journal of Econometrics*, 92, 193-232.
- Robinson, P.M., 1988, Root-n-consistent semiparametric regression, *Econometrica*, 56, 931-954.
- Van der Berg, G.J., 2001, Duration models: specification, identification, and multiple durations, in: J.J. Heckman and E. Leamer, eds. *Handbook of Econometrics*, Vol V. (North-Holland, Amsterdam) Chapter 55.
- Visser, M., 1996, Nonparametric estimation of the bivariate survival function with application to vertically transmitted AIDS, *Biometrika*, 83, 507-518.
- Wang, W., and M.T. Wells, 1998, Nonparametric estimation of successive duration times under dependent censoring, *Biometrika*, 85, 561-572.
- Woutersen, T.M., 2000, Consistent estimators for panel duration data with endogenous censoring and endogenous regressors, unpublished manuscript, Department of Economics, University of Western Ontario.

Table 1. *Weight functions and means of data-driven bandwidths used in estimation of  $\lambda_0$  and  $\Lambda_0$ . [Uncensored Case]*

	$n = 100$		$n = 500$	
	$\lambda_{n0}$	$\Lambda_{n0}$	$\lambda_{n0}$	$\Lambda_{n0}$
Weibull Model				
$h_{n1}$	4.20	2.83	2.65	1.56
$h_{n2}$	3.79	2.83	2.31	1.56
$h_{nz}$	6.69	7.34	4.57	5.19
$w_t(u)$	$1(0.5 \leq u \leq 3.5)/3$			
$w_z(u)$	Equation (22)			
U-shaped Hazard Model				
$h_{n1}$	3.16	2.13	2.29	1.34
$h_{n2}$	2.85	2.13	1.99	1.34
$h_{nz}$	4.98	5.47	3.93	4.45
$w_t(u)$	$1(0.2 \leq u \leq 5)/4.8$			
$w_z(u)$	Equation (22)			

Table 2. *Weight functions and bandwidths used in estimation of  $\beta$ ,  $\lambda_0$ , and  $\Lambda_0$ .*  
 [Censored Case]

Estimation of $\lambda_0$ and $\Lambda_0$					Estimation of $\beta$		
		$n = 100$		$n = 500$			
	$\lambda_{n0}$	$\Lambda_{n0}$	$\lambda_{n0}$	$\Lambda_{n0}$	$\beta_n$	$\beta_n$	
Weibull Model							
$h_{n1}$	3.5	3.0	2.5	2.0	$h_{n1}$	4.5	3.5
$h_{n2}$	3.5	3.0	2.5	2.0	$h_{n2}$	4.5	3.5
$h_{nz}$	5.0	7.0	4.0	5.0	$h_{nx}$	1.0	0.7
$w_t(u)$	$1(0.5 \leq u \leq 3.5)/3$				$w_\beta(u)$	$1(0.5 \leq u \leq 3.5)/3$	
$w_z(u)$	Equation (22)				$w_x(u)$	$1( u  \leq 1)$	
U-shaped Hazard Model							
$h_{n1}$	3.0	2.5	2.0	1.5	$h_{n1}$	5.0	4.0
$h_{n2}$	3.0	2.5	2.0	1.5	$h_{n2}$	5.0	4.0
$h_{nz}$	6.0	7.0	4.0	5.0	$h_{nx}$	1.2	0.9
$w_t(u)$	$1(0.2 \leq u \leq 5)/4.8$				$w_\beta(u)$	$1(0.2 \leq u \leq 5)/4.8$	
$w_z(u)$	Equation (22)				$w_x(u)$	$1( u  \leq 1)$	

Table 3. Monte Carlo results for the estimator of  $\beta$  (Censored Case).

	Mean Bias	Median Bias	Std. Dev.	RMSE	MAE
<i>Weibull Model</i>					
<i>Second-Order Kernel</i>					
$n = 100$	-0.209	-0.225	0.294	0.360	0.274
$n = 500$	-0.166	-0.155	0.150	0.216	0.174
<i>Fourth-Order Kernel</i>					
$n = 100$	-0.024	-0.073	0.381	0.381	0.230
$n = 500$	-0.015	-0.045	0.205	0.206	0.142
<i>U-shaped Hazard Model</i>					
<i>Second-Order Kernel</i>					
$n = 100$	-0.200	-0.198	0.308	0.367	0.230
$n = 500$	-0.120	-0.128	0.162	0.207	0.137
<i>Fourth-Order Kernel</i>					
$n = 100$	-0.089	-0.095	0.414	0.424	0.271
$n = 500$	-0.036	-0.058	0.291	0.293	0.188

Note: Table 3 presents the mean bias, median bias, standard deviation, root mean squared error (RMSE), and median absolute error (MAE) of the estimator.

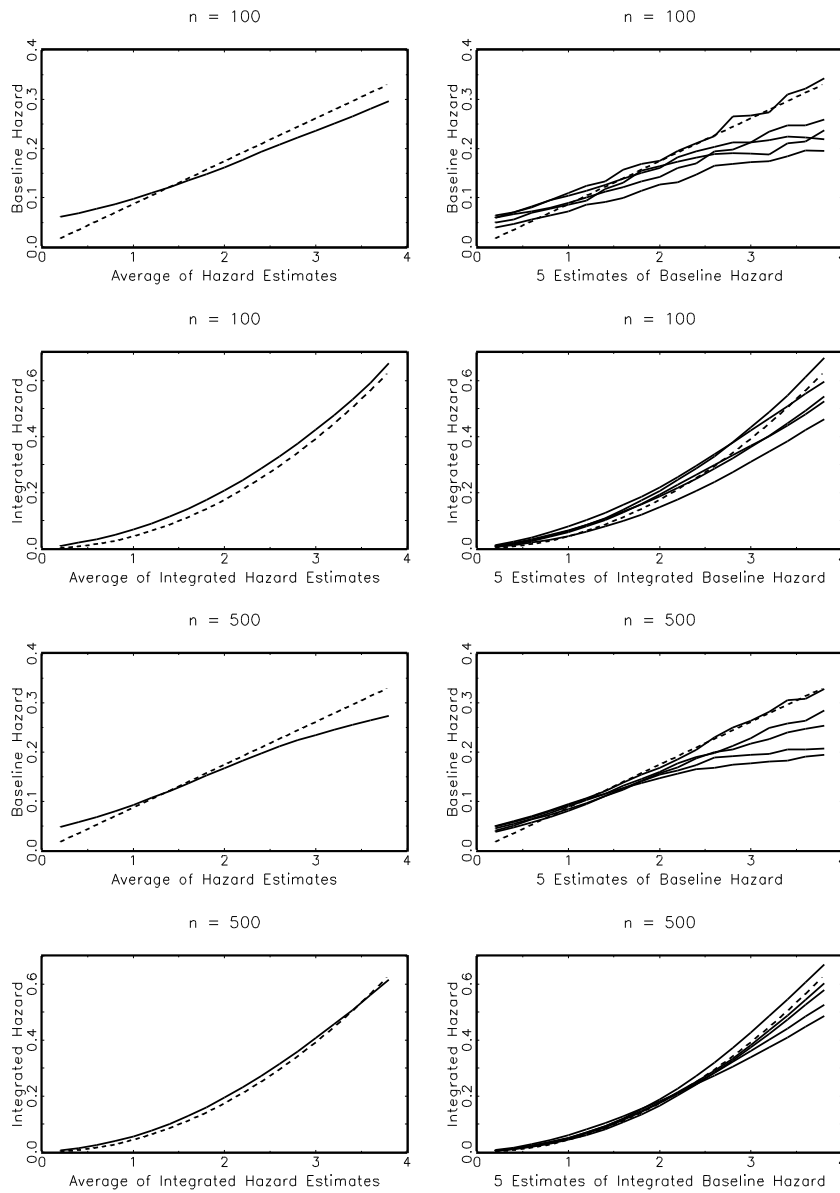


Figure 1. Monte Carlo results for the Weibull model (Uncensored Case).

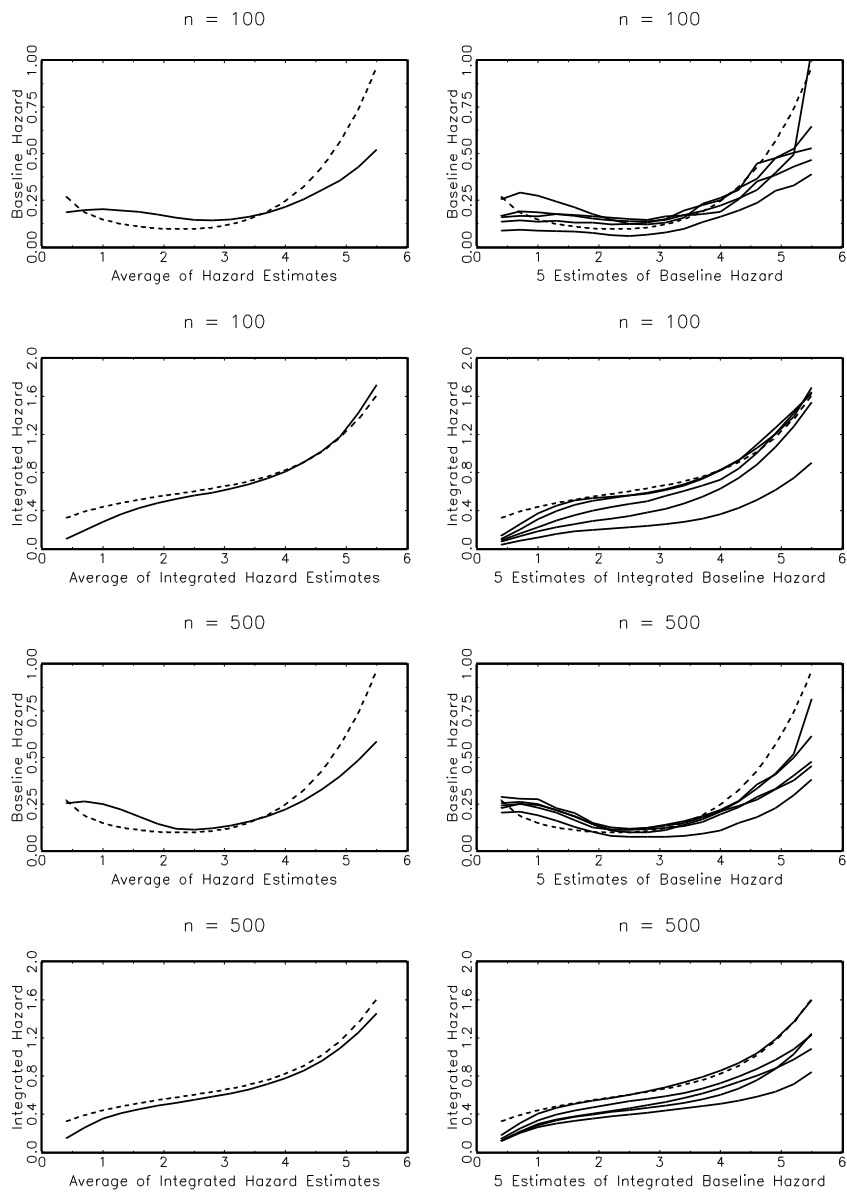


Figure 2. Monte Carlo results for the U-shaped model (Uncensored Case).



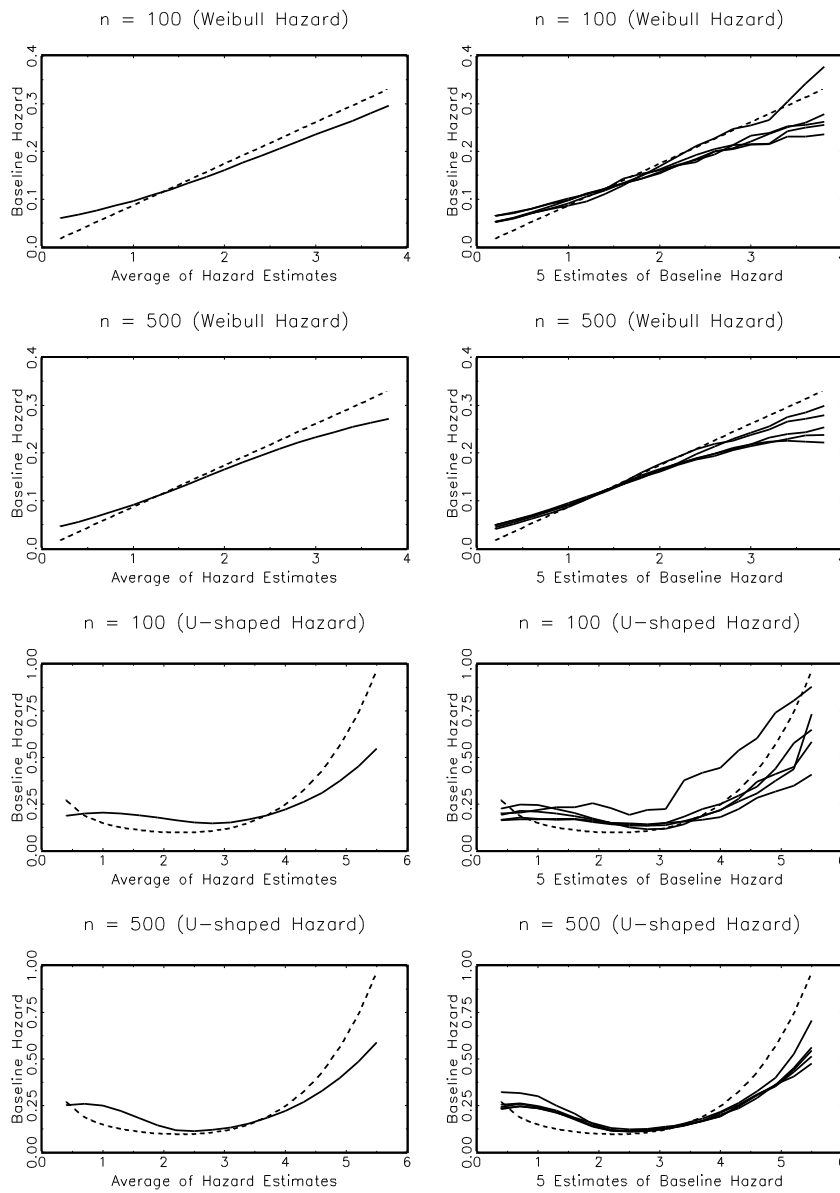


Figure 3. Monte Carlo results for the linearly combined estimator (Uncensored Case).

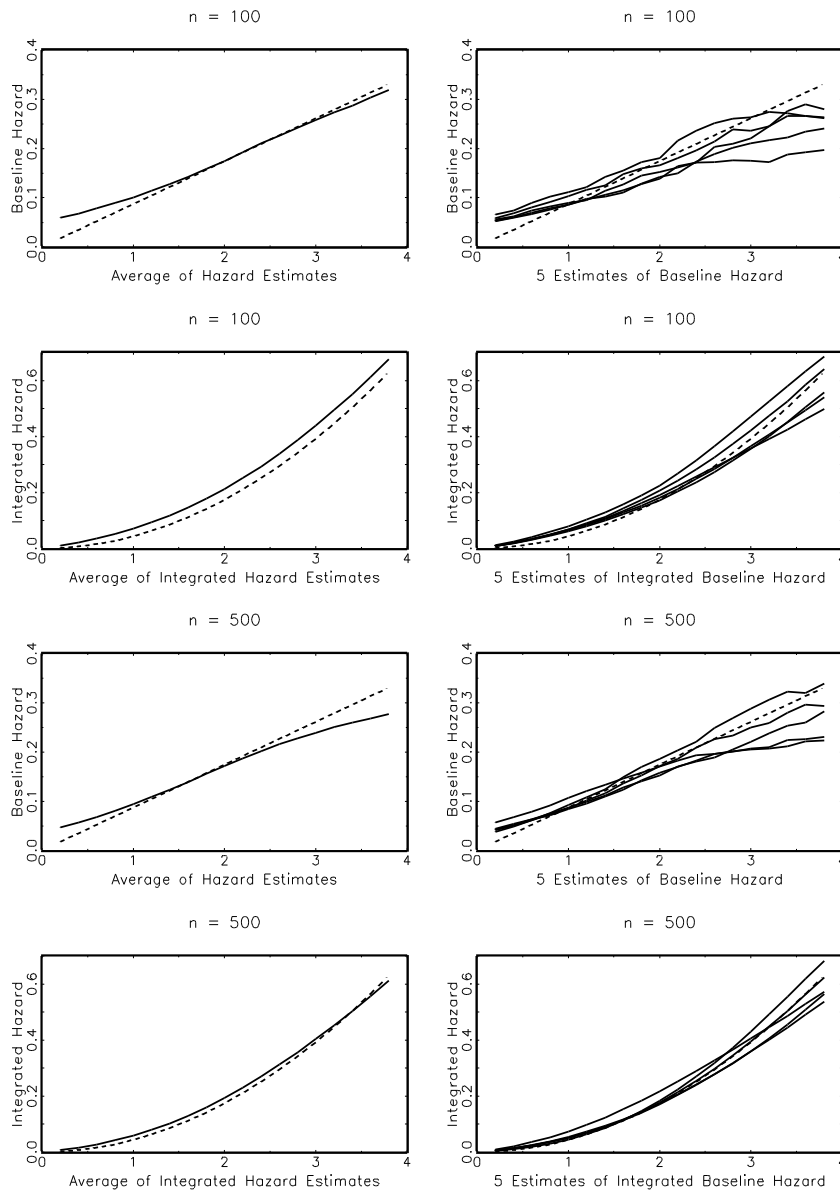


Figure 4. Monte Carlo results for the Weibull model (Censored Case).

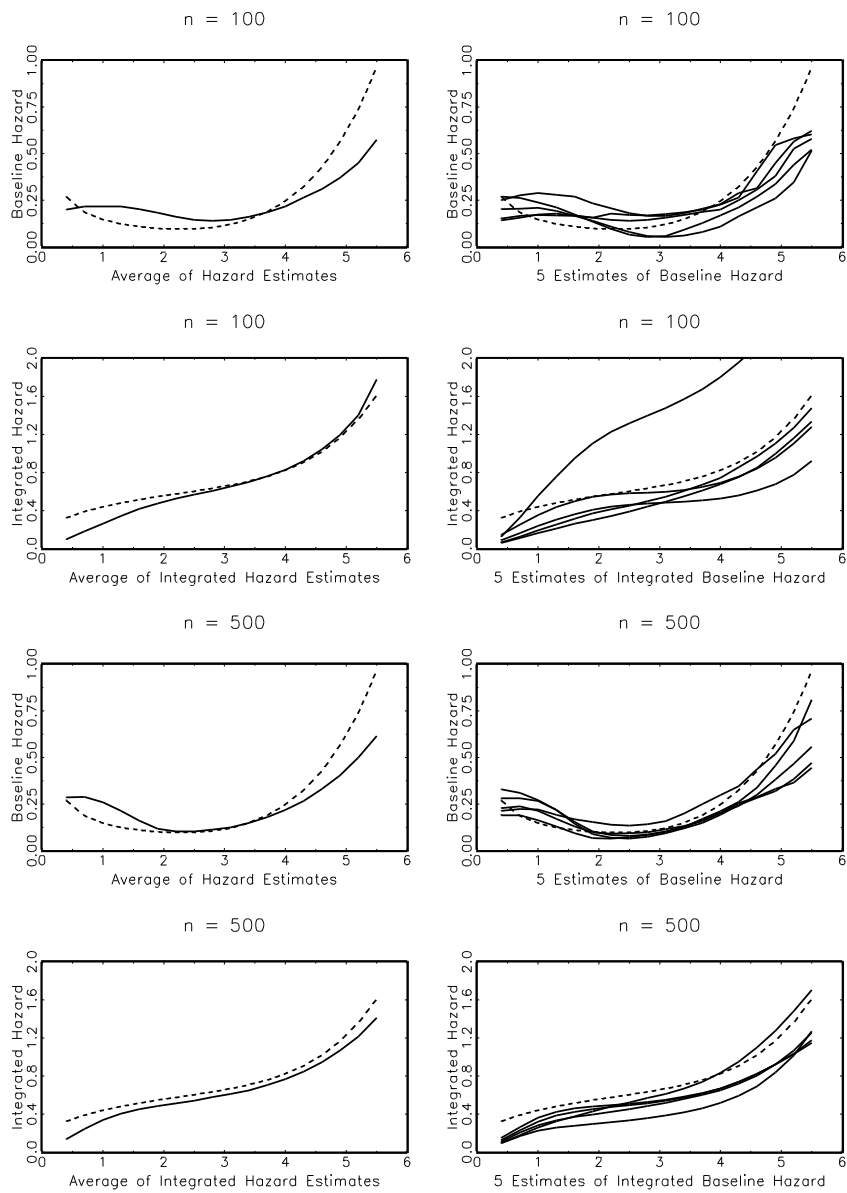


Figure 5. Monte Carlo results for the U-shaped model (Censored Case).