Econometric Theory, **24**, 2008, 1254–1276. Printed in the United States of America.
doi:10.1017/S0266466608080493

ESTIMATING PANEL DATA DURATION MODELS WITH CENSORED DATA

Sokbae Lee
Centre for Microdata Methods and Practice
Institute for Fiscal Studies
and
University College London

This paper presents a method for estimating a class of panel data duration models, under which an unknown transformation of the duration variable is linearly related to the observed explanatory variables and the unobserved heterogeneity (or frailty) with completely known error distributions. This class of duration models includes a panel data proportional hazards model with fixed effects. The proposed estimator is shown to be $n^{1/2}$-consistent and asymptotically normal with dependent right censoring. The paper provides some discussions on extending the estimator to the cases of longer panels and multiple states. Some Monte Carlo studies are carried out to illustrate the finite-sample performance of the new estimator.

1. INTRODUCTION

Panel durations consist of multiple, sequentially observed durations of the same kind of events on each individual. In a large number of applications across different scientific fields, these panel durations are observed along with possible explanatory variables. Examples of panel durations include recurrences of a given illness (Wei, Lin, and Weissfeld, 1989), unemployment spells and job durations (Heckman and Borjas, 1980; Topel and Ward, 1992), birth intervals (Newman and McCullogh, 1984), car insurance claim durations (Abbring, Chiappori, and Pinquet, 2003), and household interpurchase times of a give

I thank Richard Blundell, Hide Ichimura, Roger Koenker, Bruce Meyer, Jeffrey Wooldridge (co-editor), two anonymous referees, and seminar participants at the 2003 Econometric Society North American Summer Meeting, the CAM Workshop on Nonlinear Dynamic Panel Data Models, the LSE Joint Econometrics and Statistics Workshop, the Iowa Alumni Workshop, the department seminar at the University of Mannheim, and the UCL Econometrics Lunch Seminar for many helpful comments and suggestions. I also thank the Leverhulme Trust for the funding of the Centre for Microdata Methods and Practice and of the research program Evidence, Inference and Inquiry. An earlier version of the paper was circulated under the title "Semiparametric Estimation of Panel Data Duration Models with Fixed Effects." Address correspondence to Sokbae Lee, Department of Economics, University College London, London, WC1E 6BT, United Kingdom; e-mail: l.simon@ucl.ac.uk.

  0266-4666/08 $15.00

product (Jain and Vilcassim, 1991). This paper is concerned with estimating a class of panel data duration models that can be viewed as panel data transformation models.

One econometric model that has been widely used in duration analysis is the mixed proportional hazards model. This model is often defined in terms of the hazard function of a positive random variable T (duration variable) conditional on a vector of observed explanatory variables X (covariates) and an unobserved random variable U (the unobserved heterogeneity or frailty). One form of this model is

$$\lambda(t|x,u) = \lambda_0(t)\exp(x'\beta + u), \tag{1}$$

where $\lambda(t|x,u)$ is the hazard that $T = t$ conditional on $X = x$ and $U = u$, the function λ_0 is the baseline hazard function, and β is the vector of unknown parameters. Here, x' denotes the transpose of x.

It is well known (see, e.g., Van der Berg, 2001, Sect. 4) that the mixed proportional hazards model (1) can be written as the linear transformation model

$$\log \Lambda_0(T) = -X'\beta - U + \varepsilon, \tag{2}$$

where $\Lambda_0(t) \equiv \int_0^t \lambda_0(u)\, du$ is the integrated baseline hazard function and ε is an unobserved random variable that is independent of X and U and has the type 1 extreme value distribution function. The model (2) belongs to a class of linear transformation models

$$H(T) = -X'\beta - U + \varepsilon, \tag{3}$$

where $H(\cdot)$ is an unknown strictly increasing function and ε has a completely specified distribution function $F(\cdot)$. If F is the type 1 extreme value distribution $F(u) = 1 - \exp(-e^u)$, model (3) is the mixed proportional hazards model in (2). If F is the logistic distribution $F(u) = e^u/(1 + e^u)$, model (3) can be called a mixed proportional odds model. For example, see Cheng, Wei, and Ying (1995) and Horowitz (1996, 1998, Ch. 5) for detailed discussions of applications of the transformation models.

This paper considers a panel data version of (3):

$$H_i(T_{ij}) = -X_{ij}'\beta - U_i + \varepsilon_{ij} \qquad (i = 1, \ldots, n, \qquad j = 1, \ldots, J), \tag{4}$$

where i denotes an individual and j denotes a duration. For example, T_{ij} denotes the ith individual's jth duration. It is assumed here that duration variables are successive and observed sequentially. That is, T_{i1} is followed by T_{i2}, which is followed by T_{i3}, and so on.

The observed covariates X_{ij} are assumed to be constant within each spell but to vary over spells, whereas the unobserved heterogeneity U_i is assumed to be identical over spells. Thus, U_i represents unobserved, permanent attributes of the ith individual. Covariates that are constant over spells are not included explicitly. They can be included in U_i, and their β coefficients are not identified. We allow U_i to be arbitrarily correlated with X_{ij} and do not impose any distributional assumptions on U_i, and therefore U_i is a *fixed effect*. Panel data structure allows unobserved heterogeneity to have a very general form, compared to unobserved heterogeneity in the single-spell duration models (e.g., Murphy, 1995).

It is also assumed that the unknown link function $H_i(\cdot)$ is strictly increasing but can be different across individuals. Therefore, the model (4) allows for unobserved heterogeneity in the shape of the link function also. Finally, it is assumed that ε_{ij} are independent of X_{ij} and independently and identically distributed (i.i.d.) across individuals and durations with a completely specified distribution. As in the cross-sectional transformation model (3), model (4) includes a panel data mixed proportional hazards model as a special case.

The focus of this paper is on estimating β in (4) when T_{ij} is censored.[1] It is well known (see, e.g., Kalbfleisch and Prentice, 1980, Sect. 8.1.2; Chamberlain, 1985; Ridder and Tunalı, 1999; Lancaster, 2000) that β can be estimated by a "stratified" partial likelihood approach when T_{ij} is uncensored or independently censored and $F(u) = 1 - \exp(-e^u)$. The usefulness of the stratified partial likelihood approach for panel duration data would be limited because dependent right censoring is almost inevitable in the analysis of panel duration data. The standard independent censoring assumption is likely to be violated if panel durations T_{ij} are correlated. For example, see Visser (1996), Wang and Wells (1998), and Lin, Sun, and Ying (1999) for discussions of the dependent censoring problem in terms of estimating survivor functions without covariates.

The contribution of this paper is on developing an estimator of β when T_{ij} is dependently censored and F is known, not necessarily the type 1 extreme value distribution. Therefore, this paper extends the transformation regression approach of Cheng et al. (1995) to panel duration data and provides alternatives to the marginal regression approach of Wei et al. (1989). In a related paper, Horowitz and Lee (2004) developed an estimator of β (among other things) when T_{ij} is dependently censored, $H_i(\cdot)$ is the same across individuals, and $F(u) = 1 - \exp(-e^u)$. The proposed estimator in this paper is based on a simple idea that the effect of censoring can be taken into account by using some proper weights. The use of weighting is widespread in many contexts, and there are many estimators based on weighting to deal with censoring. See, for example, Koul, Susarla, and Van Ryzin (1981) and Cheng et al. (1995) among many others.[2]

The paper is organized as follows. The next section describes the duration model and gives an informal description of the estimator of β. Asymptotic properties of the proposed estimator are given in Section 3. Extensions are discussed in Section 4. Section 5 presents results of some Monte Carlo studies. Concluding remarks are in Section 6. The proof of the main theorem is in the Appendix.

2. ESTIMATION OF THE PANEL DATA DURATION MODEL

It is useful to begin with a description of the censoring mechanism. It is assumed in this section that the number of durations $J = 2$. Let T_1 and T_2 be the duration variables of two consecutive and adjacent events. For $J = 2$, the model (4) has the form

$$H(T_1) = -X_1'\beta - U + \varepsilon_1 \quad \text{and} \quad H(T_2) = -X_2'\beta - U + \varepsilon_2. \tag{5}$$

Censoring is an inevitable part of modeling in duration analysis. To describe a censoring mechanism for successive durations T_1 and T_2, we assume that T_1 and T_2 are observed consecutively over a time period C, where C is random with an unknown probability distribution. As discussed in Visser (1996) and Wang and Wells (1998), there are three possible cases:

1. if $C \geq T_1 + T_2$, both T_1 and T_2 are uncensored;
2. if $T_1 \leq C < T_1 + T_2$, T_1 is uncensored but T_2 is censored;
3. if $C < T_1$, T_1 is censored and T_2 is unobserved.

Notice that T_1 is censored by $C_1 \equiv C$ and that T_2 is censored by $C_2 \equiv (C_1 - T_1)1(T_1 \leq C_1)$, where $1(\cdot)$ is the usual indicator function. Under this censoring mechanism, C_2 is correlated with T_2 because T_1 and T_2 are correlated by unobserved heterogeneity. This indicates that it would be quite difficult to estimate a (cross-sectional) duration model for T_2 in separation from T_1 with censored data. However, we will show subsequently that β in (4) can be estimated consistently.

To do so, we assume that one observes a pair of (Y_j, Δ_j) not T_j, where $Y_j = \min(T_j, C_j)$ and $\Delta_j = 1(T_j \leq C_j)$ for $j = 1, 2$. The observed data consist of i.i.d. realizations $\{(Y_{i1}, Y_{i2}, X_{i1}, X_{i2}, \Delta_{i1}, \Delta_{i2}) : i = 1, \ldots, n\}$ from $(Y_1, Y_2, X_1, X_2, \Delta_1, \Delta_2)$. Let $G(c)$ denote the survivor function of C, that is, $G(c) = \Pr(C \geq c)$, and let $\Delta X = X_1 - X_2$. Assume that C is independent of (T_1, T_2, X_1, X_2, U). Let $L(u) = \Pr[(\varepsilon_1 - \varepsilon_2) > u]$ for any real value u. Also, let $l(u) = -dL(u)/du$; that is, $l(\cdot)$ is the probability density function of $(\varepsilon_1 - \varepsilon_2)$. Then if we assume that ε_1 and ε_2 are i.i.d. with the common distribution function F,

$$L(u) = \int_{-\infty}^{\infty} [1 - F(u + v)]\, dF(v). \tag{6}$$

Assume further that ε_1 and ε_2 are independent of X_1 and X_2. Notice that under the assumptions made previously,

$$
\begin{aligned}
&\mathrm{E}\left[\frac{\Delta_1\Delta_2}{G(Y_1+Y_2)}\{1(Y_1>Y_2)-L(\Delta X'\beta)\}|X_1,X_2\right]\\
&\quad=\mathrm{E}\left[\mathrm{E}\left[\frac{1(T_1+T_2\le C)}{G(T_1+T_2)}\{1(T_1>T_2)-L(\Delta X'\beta)\}|T_1,T_2,X_1,X_2\right]\middle|X_1,X_2\right]\\
&\quad=\mathrm{E}[1(T_1>T_2)-L(\Delta X'\beta)|X_1,X_2]\\
&\quad=\mathrm{E}[1\{H(T_1)>H(T_2)\}-L(\Delta X'\beta)|X_1,X_2]\\
&\quad=\Pr[(\varepsilon_1-\varepsilon_2)>(X_1-X_2)'\beta|X_1,X_2]-L(\Delta X'\beta)\\
&\quad=0. \qquad\qquad \textbf{(7)}
\end{aligned}
$$

This implies that β satisfies the moment condition

$$
\mathrm{E}\left\{w_h(\Delta X'\beta)\Delta X\frac{\Delta_1\Delta_2}{G(Y_1+Y_2)}[1(Y_1>Y_2)-L(\Delta X'\beta)]\right\}=0, \qquad \textbf{(8)}
$$

where $w_h(\cdot)$ is a weight function.[3]

Our estimation strategy in this paper is to solve the sample analog of the population moment condition (8). In other words, our estimator b_n of β is the solution to the following estimating equation:

$$
n^{-1}\sum_{i=1}^{n}\left\{w_h(\Delta X_i'b)\Delta X_i\frac{\Delta_{i1}\Delta_{i2}}{G_n(Y_{i1}+Y_{i2})}[1(Y_{i1}>Y_{i2})-L(\Delta X_i'b)]\right\}=0, \qquad \textbf{(9)}
$$

where G_n is an estimator of G. Because C is censored independently by $T_1 + T_2$, we will use the Kaplan–Meier estimator of G for G_n. Specifically, G_n is estimated based on the data $\{(Y_{i1} + Y_{i2}, 1 - \Delta_{i1}\Delta_{i2}) : i = 1, \dots, n\}$.[4]

We end this section by mentioning some connection to well-known estimation methods. The estimating equation (9) can be viewed as a modification of the estimating equation of Cheng et al. (1995, see eqns. (2.1) and (2.3)), who used pairwise comparisons in a cross section. Without censoring, our estimator is the same as the uncensored estimator of Cheng et al. (1995, eqn. (2.1)), although differencing across time instead of individuals. If $w_h(\cdot) = l(\cdot)/\{L(\cdot)[1 - L(\cdot)]\}$, the estimator defined in (9) can be thought of as a weighted maximum-likelihood type estimator, meaning that b_n is the solution to

$$
\begin{aligned}
\max_b n^{-1}\sum_{i=1}^{n}\frac{\Delta_{i1}\Delta_{i2}}{G_n(Y_{i1}+Y_{i2})}&\{1(Y_{i1}>Y_{i2})\log[L(\Delta X_i'b)]\\
&+1(Y_{i1}\le Y_{i2})\log[1-L(\Delta X_i'b)]\}.
\end{aligned}
$$

When F is the type 1 extreme value distribution function, it can be seen that the proposed estimator is a weighted logit estimator with weight equal to the inverse of the probability that T_1 and T_2 are uncensored.

3. ASYMPTOTIC PROPERTIES OF THE ESTIMATOR

This section establishes the $n^{-1/2}$-consistency and asymptotic normality of b_n. To do so, we make the following assumptions.

Assumption 1. β is an interior point of the parameter space $\mathbf{B}$, which is a compact subset of $\mathbf{R}^d$.

Assumption 2. The data $\{(Y_{i1}, Y_{i2}, X_{i1}, X_{i2}, \Delta_{i1}, \Delta_{i2}) : i = 1, \ldots, n\}$ are i.i.d. realizations from $(Y_1, Y_2, X_1, X_2, \Delta_1, \Delta_2)$ in (5).

It is possible that X_{i1} and X_{i2} are missing when durations of interest are censored, especially when X_{i1} and X_{i2} are observed characteristics of durations. This does not cause any problem for the estimation procedure in Section 2 because the estimating equation (9) mainly uses observations corresponding to complete durations (i.e., $\Delta_{i1} = \Delta_{i2} = 1$). Observations with incomplete durations are only used to obtain an estimator of G to take into account the effect of dependent right censoring. Because C is independent of X_1 and X_2, it is unnecessary to observe X_1 and X_2 when durations are censored.

Assumption 3.

(a) ε_1 and ε_2 have the same distribution function $F(\cdot)$, which is completely specified.
(b) There exists a corresponding probability density function $f(\cdot)$, which is bounded, continuous, and positive everywhere along the real line.
(c) Furthermore, ε_1 and ε_2 are independent of each other and independent of (X_1, X_2).

As already discussed, this condition is satisfied by the panel data proportional hazards model with unobserved heterogeneity.

Assumption 4. The function $H(\cdot)$ is strictly increasing.

It can be seen from (7) that the link function $H(\cdot)$ can be different across individuals. This allows for arbitrary heterogeneity in the shape of the link function. As a matter of fact, U is not identified from $H(\cdot)$ because $H(\cdot)$ can vary over individuals. However, the model is expressed in the form of (4) to emphasize connections between our model (4) and duration models with unobserved heterogeneity.[5]

Assumption 5. The weight function $w_h(\cdot)$ is bounded and positive everywhere along the real line and has a bounded, continuous derivative.

A simple choice of w_h would be to set $w_h \equiv 1$. As suggested by Cheng et al. (1995), one might use $w_h(\cdot) = l(\cdot)/\{L(\cdot)[1 - L(\cdot)]\}$ to mimic the quasi-likelihood approach. Let $\| \ \|$ denote the euclidean norm.

Assumption 6. $\mathrm{E}\|\Delta X\|^2 < \infty$ and $\mathrm{E}[\Delta X \Delta X']$ is nonsingular.

This condition requires that covariates vary over spells, thereby excluding the constant term and spell-constant covariates.[6]

Assumption 7.

(a) The censoring variable C is random with an unknown continuous probability distribution. In addition, C is independent of (T_1, T_2, X_1, X_2, U).
(b) The survivor function of C, $G(c) \equiv \Pr(C \geq c)$ is positive for every $c \in \mathbf{R}$.

Assumption 7(a) is a convenient assumption under which we utilize results of counting process and martingale methods for the Kaplan–Meier estimator of $G(\cdot)$.[7] Assumption 7(b) is a rather strong condition, and especially it excludes the case of fixed censoring.[8] The same condition is assumed in Koul et al. (1981, Assump. A1).

To present our main result, define $\pi(s) = \Pr(Y_1 + Y_2 \geq s)$ and

$$M_i(s) = 1(Y_{i1} + Y_{i2} \leq s, \Delta_{i1}\Delta_{i2} = 0) - \int_0^s 1(Y_{i1} + Y_{i2} \geq c)\, d\Lambda_C(c),$$

where Λ_C is the cumulative hazard function of C. In addition, define

$$\Omega = \mathrm{E}[w_h(\Delta X'\beta) l(\Delta X'\beta) \Delta X \Delta X']$$

and

$$\Gamma(s) = \mathrm{E}\left\{ w_h(\Delta X'\beta)\Delta X \frac{\Delta_1 \Delta_2}{G(Y_1 + Y_2)} [1(Y_1 > Y_2) - L(\Delta X'\beta)] \right.$$
$$\left. \times 1(Y_1 + Y_2 \geq s) \right\}.$$

The following theorem provides the main result of the paper.

THEOREM 1. *Let Assumptions 1–7 hold. Let*

$$\Phi = \mathrm{E}\left[[w_h(\Delta X'\beta)]^2 \Delta X \Delta X' \frac{\Delta_1 \Delta_2}{[G(Y_1 + Y_2)]^2} L(\Delta X'\beta)[1 - L(\Delta X'\beta)] \right]$$
$$- \int_0^\infty \frac{\Gamma(s)\Gamma(s)'}{\pi(s)}\, d\Lambda_C(s). \quad \textbf{(10)}$$

Assume that Φ *is finite. Then* $n^{1/2}(b_n - \beta)$ *is asymptotically normal with mean zero and covariance matrix* $V_\beta \equiv \Omega^{-1}\Phi\Omega^{-1}$.

The first term of Φ may not be finite if the right tail of C is much thinner than that of $T_1 + T_2$.[9] We can provide a couple of sufficient conditions that ensure that the first term of Φ is finite. If ΔX is bounded and $\mathrm{E}[1/G(T_1 + T_2)]$ is finite, then the first term of Φ is finite. Alternatively, if $\mathrm{E}[\|\Delta X\|^4]$ and $\mathrm{E}[1/G^2(T_1 + T_2)]$ are finite, then the first term of Φ is finite by Cauchy–Schwartz inequality.

Notice that the covariance matrix V_β is smaller (in the matrix sense) than one that would be obtained with a true $G(\cdot)$ instead of an estimated $G_n(\cdot)$.[10] It is straightforward to obtain a consistent estimator of the covariance matrix V_β. Define $\hat{\pi}(s) = n^{-1}\sum_{i=1}^n 1(Y_{i1} + Y_{i2} \geq s)$ and

$$\hat{\Gamma}(s) = n^{-1}\sum_{i=1}^n \left\{ w_h(\Delta X_i' b_n)\Delta X_i \frac{\Delta_{i1}\Delta_{i2}}{G_n(Y_{i1} + Y_{i2})} \times [1(Y_{i1} > Y_{i2}) - L(\Delta X_i' b_n)] 1(Y_{i1} + Y_{i2} \geq s) \right\}.$$

One can estimate V_β by its sample analog estimator $\hat{V}_\beta = \hat{\Omega}^{-1}\hat{\Phi}\hat{\Omega}^{-1}$, where

$$\hat{\Omega} = n^{-1}\sum_{i=1}^n w_h(\Delta X_i' b_n) \frac{\Delta_{i1}\Delta_{i2}}{[G_n(Y_{i1} + Y_{i2})]} l(\Delta X_i' b_n)\Delta X_i \Delta X_i'$$

and

$$\hat{\Phi} = n^{-1}\sum_{i=1}^n [w_h(\Delta X_i' b_n)]^2 \Delta X_i \Delta X_i' \frac{\Delta_{i1}\Delta_{i2}}{[G_n(Y_{i1} + Y_{i2})]^2} L(\Delta X_i' b_n)[1 - L(\Delta X_i' b_n)] - n^{-1}\sum_{i=1}^n (1 - \Delta_{i1}\Delta_{i2}) \frac{\hat{\Gamma}(Y_{i1} + Y_{i2})\hat{\Gamma}(Y_{i1} + Y_{i2})'}{[\hat{\pi}(Y_{i1} + Y_{i2})]^2}.$$

Notice that the second term of $\hat{\Phi}$ is a sample analog of the second term of Φ using the Nelson cumulative hazard estimator of Λ_C.[11]

4. EXTENSIONS

4.1. Estimation with Longer Panels

The estimation method in Section 2 easily extends to the case of longer panels. To consider estimation when $J > 2$, it is important to notice that panel durations T_{ij} in (4) are censored by C_{ij}, where $C_{i1} = C_i$ and $C_{ij} = (C_i - \sum_{k=1}^{j-1} T_{ik}) 1(T_{i,j-1} \leq C_{i,j-1})$ for $j = 2, \ldots, J$. As before, one observes $Y_{ij} = \min(T_{ij}, C_{ij})$ and $\Delta_{ij} = 1(T_{ij} \leq C_{ij})$ together with covariates X_{ij} for $j = 1, \ldots, J$

and $i = 1, \ldots, n$. Using the fact that the sum of T_{ij}'s is censored independently by C, the estimating equation (9) can be extended to longer panels. To do so, let S be a set of pairs of indices such that $S = \{(j,k) : j < k, j = 1, \ldots, J, k = 1, \ldots, J\}$, $\Delta X_{ijk} = X_{ij} - X_{ik}$, and $W_{ij} = \sum_{k=1}^{j} Y_{ik}$, that is, the sum of the first j observed spells. Then an estimator of β is the solution to the following estimating equation:

$$n^{-1} \sum_{i=1}^{n} \sum_{(j,k)\in S} \left\{ w_h(\Delta X_{ijk}' b) \Delta X_{ijk} \frac{\Delta_{ij} \Delta_{ik}}{G_n(W_{ik})} [1(Y_{ij} > Y_{ik}) - L(\Delta X_{ijk}' b)] \right\} = 0. \tag{11}$$

As in Section 2, the effect of censoring is adjusted by multiplying the inverse of the estimates $G_n(W_{ik})$ of the probability that Y_{ij} and Y_{ik} are uncensored for $j < k$. It is straightforward to obtain asymptotic properties of this estimator.

4.2. Estimation with Multiple States

This section shows how the estimation method in Section 2 can be extended to the case of multiple-state duration models. The censoring mechanism described in Section 2 considers a pure renewal process in the sense that T_1 and T_2 are the durations of the same kind and there is no time spent on other states. This pure renewal process assumption might be implausible in some applications, for example, employment and unemployment durations in labor economics. Fortunately, it is easy to extend the estimation method in Section 2 to multiple-state duration models.

Assume now that there is a different type of duration between two durations of interest, say, $\tilde{T}$. For example, T_1 may be the duration of the first job, $\tilde{T}$ the duration of being unemployed or out of the labor force, and T_2 the duration of the second job. Assume that C is independent of T_1, T_2, X_1, X_2, and $\tilde{T}$. One observes uncensored durations of T_1 and T_2 when $C \geq T_1 + \tilde{T} + T_2$. Hence, $\Delta_{i1} \Delta_{i2} = 1(C_i \geq T_{i1} + \tilde{T}_i + T_{i2})$. Then a consistent estimator of β can be obtained by solving the same estimating equation as (9), except that $G_n(Y_{i1} + Y_{i2})$ is now replaced with $G_n[\min((T_{i1} + \tilde{T}_i + T_{i2}), C_i)]$.

Basically, the estimation method in Section 2 can be extended to any censoring mechanism, provided that the probability of at least two durations of interest being uncensored is positive and can be estimated consistently. The main idea behind the estimation method is to use only observations corresponding to complete durations and to correct for the induced selection bias by using proper weights, namely, the inverse of the probability of two durations being uncensored.

5. MONTE CARLO STUDIES

This section presents the results of some simulation studies that illustrate the finite-sample performance of the estimator. For each Monte Carlo experiment, 1,000 samples were generated from the following model with $J = 2$:

$$H(T_1) = -X_{11}\beta_1 - X_{12}\beta_2 - X_{13}\beta_3 - U + \varepsilon_1,$$

$$H(T_2) = -X_{21}\beta_1 - X_{22}\beta_2 - X_{23}\beta_3 - U + \varepsilon_2,$$

where H is the natural log function, X_{11} and X_{21} were independently drawn from a uniform distribution on $[0,1]$, X_{12} and X_{22} are independent dummy variables equal to one with probability 0.5, X_{13} and X_{23} are also dummy variables such that $X_{31} = 0$ and $X_{32} = 1$, and ε_1 and ε_2 were independently drawn from the type 1 extreme value distribution. The unobserved heterogeneity U was generated by $U = (X_{11} + X_{21})/2$ and is the only source of correlation between T_1 and T_2. The true parameters are $(\beta_1, \beta_2, \beta_3) = (-1, -1, -1)$. Finally, we experiment with two types of distributions for the censoring mechanism. First, the censoring threshold C was generated from the exponential distribution with mean μ, and second, C was from the uniform distribution with support $[0, \nu]$, where different μ's and ν's were chosen to investigate the effects of censoring. Assumption 7(b) is satisfied by the exponential distribution but not by the uniform distribution. The latter distribution is considered to see how the estimator performs when Assumption 7(b) is violated. The simulations used sample sizes of $n = 100$, 200, 400, and 800, and all the simulations were carried out in GAUSS using GAUSS pseudo-random number generators. Throughout the simulations, the weight function was $w_h \equiv 1$.

Table 1 reports the mean bias and standard deviation (S.D.) for the estimate of each coefficient for the case of censoring with the exponential distribution. It can be seen that for each coefficient and for each level of censoring, the bias is negligible. Furthermore, the standard deviation decreases quite quickly as the sample size increases at about a rate of $n^{-1/2}$, although the estimator does not perform well when the proportion of censoring exceeds 50%. Table 2 reports the mean bias and standard deviation for the estimate of each diagonal component of the variance matrix V_β/n. To compute the biases and standard deviations, the finite-sample variances of estimates of coefficients (obtained by 1,000 simulations) are treated as the true values of the variances. Again the variance estimator performs well except for heavy censoring. Note that the standard deviation shrinks quite fast with the sample size because the true variance also shrinks.

We now consider the case of censoring with the uniform distribution. The results are summarized in Tables 3 and 4. Not surprisingly, the performance of the estimator is worse compared to the case with the exponential distribution. Note that the asymptotic biases are quite small for light censoring (up to 30%) and they get larger for heavier censoring. Similar conclusions can be drawn for variance estimates.

In summary, our simulation results suggest that (1) the new estimator and its variance estimator perform very well in finite samples for light and moderate censoring (up to 50%) when the censoring variable has infinite support, (2) they

Table 1. Simulation results for estimates of coefficients (Censoring variable: exponential distribution)

Proportion of censoring	Sample size	β_1 Bias	β_1 S.D.	β_2 Bias	β_2 S.D.	β_3 Bias	β_3 S.D.
10%	100	−0.060	0.672	−0.035	0.393	−0.032	0.271
	200	0.009	0.466	−0.010	0.261	−0.010	0.188
	400	−0.013	0.323	−0.017	0.185	−0.004	0.129
	800	−0.005	0.221	0.003	0.130	−0.005	0.090
20%	100	−0.060	0.708	−0.034	0.417	−0.022	0.285
	200	0.011	0.493	−0.005	0.275	−0.008	0.200
	400	−0.013	0.335	−0.014	0.193	−0.005	0.135
	800	−0.005	0.232	0.003	0.135	−0.002	0.095
30%	100	−0.065	0.795	−0.026	0.447	−0.023	0.312
	200	0.019	0.536	0.000	0.302	−0.005	0.214
	400	−0.006	0.358	−0.013	0.208	−0.005	0.147
	800	−0.002	0.252	0.006	0.145	0.003	0.101
40%	100	−0.062	0.890	−0.024	0.502	−0.010	0.346
	200	0.026	0.608	0.000	0.334	0.006	0.237
	400	−0.003	0.401	−0.009	0.235	−0.003	0.162
	800	0.010	0.288	0.010	0.157	0.008	0.112
50%	100	−0.056	1.054	−0.012	0.578	0.000	0.404
	200	0.024	0.706	0.005	0.393	0.017	0.275
	400	0.022	0.473	0.001	0.270	0.014	0.185
	800	0.011	0.332	0.017	0.179	0.014	0.130
60%	100	−0.016	1.290	0.010	0.713	0.020	0.492
	200	0.053	0.828	0.036	0.471	0.042	0.334
	400	0.038	0.571	0.024	0.326	0.033	0.235
	800	0.031	0.418	0.036	0.220	0.036	0.167
70%	100	−0.004	1.736	0.001	0.997	0.048	0.681
	200	0.077	1.077	0.074	0.601	0.074	0.426
	400	0.072	0.750	0.045	0.424	0.053	0.311
	800	0.049	0.539	0.063	0.302	0.068	0.218

Note: Bias denotes the mean bias, and S.D. stands for standard deviation.

perform quite well for light censoring (up to 30%) when the censoring variable has finite support, and (3) the performance deteriorates rapidly as the proportion of censoring exceeds 50% for both cases of censoring. In view of these results, we recommend the new estimator when the censoring involves less than 50% of observations, especially with small sample sizes.

Table 2. Simulation results for estimates of the variances (Censoring variable: exponential distribution)

Proportion of censoring	Sample size	β_1 Bias	β_1 S.D.	β_2 Bias	β_2 S.D.	β_3 Bias	β_3 S.D.
10%	100	−0.015	0.102	0.002	0.038	0.001	0.018
	200	−0.015	0.030	0.005	0.011	0.000	0.005
	400	−0.005	0.010	0.001	0.004	0.001	0.002
	800	0.000	0.003	0.001	0.001	0.000	0.001
20%	100	0.006	0.130	0.008	0.049	0.006	0.024
	200	−0.006	0.041	0.010	0.015	0.001	0.007
	400	0.003	0.013	0.004	0.005	0.002	0.002
	800	0.003	0.005	0.002	0.002	0.001	0.001
30%	100	−0.001	0.201	0.023	0.069	0.011	0.033
	200	0.004	0.069	0.013	0.022	0.005	0.011
	400	0.014	0.022	0.008	0.008	0.004	0.004
	800	0.007	0.008	0.004	0.003	0.002	0.001
40%	100	0.021	0.332	0.034	0.116	0.021	0.056
	200	0.006	0.105	0.024	0.037	0.011	0.019
	400	0.028	0.045	0.012	0.016	0.008	0.008
	800	0.013	0.017	0.009	0.006	0.005	0.003
50%	100	0.009	0.604	0.058	0.192	0.033	0.103
	200	0.023	0.191	0.038	0.072	0.020	0.037
	400	0.044	0.093	0.023	0.032	0.015	0.015
	800	0.031	0.044	0.018	0.015	0.009	0.007
60%	100	0.009	1.150	0.098	0.502	0.051	0.212
	200	0.073	0.353	0.061	0.148	0.032	0.081
	400	0.084	0.176	0.042	0.064	0.022	0.034
	800	0.046	0.089	0.032	0.035	0.014	0.016
70%	100	0.055	3.859	0.099	1.173	0.055	0.700
	200	0.079	0.842	0.108	0.321	0.048	0.159
	400	0.149	0.467	0.084	0.168	0.043	0.097
	800	0.107	0.250	0.058	0.091	0.030	0.046

Note: Bias denotes the mean bias, and S.D. stands for standard deviation. The finite-sample variances of estimates of coefficients (obtained by 1,000 simulations) are treated as the true value of the variances.

6. CONCLUSIONS

This paper has considered the estimation of panel data duration models with unobserved heterogeneity. In particular, this paper has provided a method for estimating the regression coefficients under dependent right censoring. The new estimator has its strengths and weaknesses. The strengths are that the estimator

Table 3. Simulation results for estimates of coefficients (Censoring variable: uniform distribution)

Proportion of censoring	Size sample	β_1 Bias	β_1 S.D.	β_2 Bias	β_2 S.D.	β_3 Bias	β_3 S.D.
10%	100	−0.057	0.671	−0.032	0.392	−0.029	0.271
	200	0.013	0.466	−0.009	0.260	−0.009	0.188
	400	−0.012	0.322	−0.016	0.186	−0.003	0.129
	800	−0.004	0.222	0.004	0.130	−0.003	0.090
20%	100	−0.046	0.713	−0.024	0.419	−0.012	0.285
	200	0.028	0.491	0.006	0.273	0.004	0.199
	400	−0.001	0.338	−0.005	0.193	0.005	0.136
	800	0.007	0.235	0.011	0.135	0.007	0.095
30%	100	−0.027	0.786	0.012	0.449	0.019	0.313
	200	0.053	0.543	0.037	0.297	0.037	0.217
	400	0.034	0.372	0.024	0.210	0.034	0.151
	800	0.041	0.261	0.040	0.146	0.043	0.103
40%	100	0.034	0.869	0.053	0.498	0.068	0.350
	200	0.090	0.619	0.084	0.338	0.085	0.242
	400	0.085	0.421	0.071	0.238	0.085	0.170
	800	0.089	0.293	0.092	0.166	0.093	0.120
50%	100	0.069	0.994	0.108	0.601	0.137	0.397
	200	0.160	0.692	0.143	0.394	0.152	0.265
	400	0.147	0.459	0.139	0.278	0.152	0.187
	800	0.146	0.342	0.165	0.189	0.155	0.136
60%	100	0.170	1.193	0.192	0.734	0.202	0.471
	200	0.217	0.803	0.211	0.474	0.224	0.323
	400	0.232	0.535	0.228	0.321	0.236	0.228
	800	0.238	0.394	0.245	0.228	0.231	0.161
70%	100	0.259	1.578	0.261	0.899	0.275	0.557
	200	0.366	0.940	0.344	0.560	0.307	0.379
	400	0.356	0.647	0.335	0.392	0.324	0.249
	800	0.326	0.490	0.354	0.281	0.334	0.192

Note: Bias denotes the mean bias, and S.D. stands for standard deviation.

is fairly easy to implement and can be extended easily to the cases of longer panels and multiple states. However, there are weaknesses regarding the regularity conditions on the censoring variable. The new estimator may not be consistent without infinite support for the censoring variable; however, when this assumption is not satisfied, the estimator performs quite well in the Monte Carlo experiments for the cases with light censoring (up to 30% of observations).

TABLE 4. Simulation results for estimates of the variances (Censoring variable: uniform distribution)

Proportion of censoring	Sample size	β_1 Bias	β_1 S.D.	β_2 Bias	β_2 S.D.	β_3 Bias	β_3 S.D.
10%	100	−0.013	0.101	0.003	0.037	0.001	0.018
	200	−0.013	0.031	0.006	0.012	0.000	0.005
	400	−0.005	0.010	0.001	0.004	0.001	0.002
	800	0.000	0.004	0.001	0.001	0.000	0.001
20%	100	0.002	0.135	0.006	0.050	0.007	0.024
	200	−0.003	0.043	0.010	0.015	0.002	0.007
	400	0.003	0.015	0.005	0.006	0.002	0.003
	800	0.003	0.006	0.003	0.002	0.001	0.001
30%	100	0.006	0.207	0.019	0.074	0.010	0.036
	200	0.002	0.096	0.017	0.027	0.005	0.015
	400	0.011	0.040	0.009	0.013	0.004	0.006
	800	0.007	0.017	0.005	0.006	0.003	0.003
40%	100	0.016	0.318	0.028	0.116	0.012	0.059
	200	−0.013	0.138	0.021	0.051	0.008	0.025
	400	0.013	0.082	0.012	0.030	0.006	0.016
	800	0.012	0.044	0.008	0.019	0.003	0.008
50%	100	0.005	0.519	−0.002	0.206	0.011	0.082
	200	−0.005	0.218	0.020	0.097	0.013	0.041
	400	0.032	0.121	0.011	0.045	0.009	0.025
	800	0.017	0.100	0.012	0.030	0.006	0.019
60%	100	−0.045	1.178	−0.035	0.499	0.010	0.173
	200	−0.010	0.358	0.011	0.135	0.008	0.076
	400	0.033	0.212	0.018	0.097	0.007	0.043
	800	0.026	0.187	0.013	0.069	0.008	0.044
70%	100	−0.356	2.610	−0.046	0.772	0.028	0.329
	200	0.009	0.691	0.014	0.244	0.016	0.188
	400	0.014	0.336	0.011	0.111	0.015	0.058
	800	0.000	0.258	0.010	0.068	0.006	0.060

Note: Bias denotes the mean bias, and S.D. stands for standard deviation. The finite-sample variances of estimates of coefficients (obtained by 1,000 simulations) are treated as the true value of the variances.

Another possible extension that is not included in Section 4 is to let $F(\cdot)$ be unknown. Recently, Khan and Tamer (2007) have proposed estimators for the regression coefficients in censored duration models with unknown $F(\cdot)$ and with general forms of censoring but excluding the dependent right censoring considered in this paper. When $F(\cdot)$ is unknown, (7) can be thought of as a single index mean regression model, in which $(\Delta_1 \Delta_2)/(G(Y_1 + Y_2))1(Y_1 > Y_2)$ is the dependent variable. Thus, it is expected that β can be estimated (up to scale) at an $n^{-1/2}$ rate by combining methods similar to those used in the analysis of single index models (see, e.g., Ichimura, 1993; Klein and Spady, 1993; Powell, Stock, and Stoker, 1989; Horowitz and Härdle, 1996; Hristache, Juditski, and Spokoiny, 2001) with some tail behavior restrictions on the Kaplan–Meier estimator of $G(\cdot)$. This is a topic for future research.

NOTES

1. In this paper, we regard β as parameters of interest, and we treat H_i as nuisance parameters. To give a specific example where β is of interest, consider a recent empirical work by Abbring et al. (2003). They test for moral hazard by checking whether car insurance claim intensities show negative occurrence dependence. This can be modeled semiparametrically in our setup by using dummy variables for panel durations of claims as part of X. A very general form of individual heterogeneity can be allowed by not specifying H_i.

2. See equations (3.51) and (3.52) of Powell (1994, p. 2505) for a concise explanation of the idea behind the estimator of Koul et al. (1981).

3. Obviously there are other moment conditions that can be derived from (7). It may be useful to develop a more efficient generalized method of moments (GMM) estimator using a set of possible moment conditions; however, it is beyond the scope of this paper to investigate the issue of efficiency.

4. Because C is also censored independently by T_1, the Kaplan–Meier estimator G_n could be estimated based on the data $\{(Y_{i1}, 1 - \Delta_{i1}) : i = 1, \ldots, n\}$ also.

5. The stratified partial likelihood approach also allows the baseline hazard function to vary over individuals. See, for example, Chamberlain (1985) and Ridder and Tunalı (1999) for details.

6. As is common among fixed-effects estimators, if regression coefficients of spell-constant covariates vary over spells, then the difference between two coefficients can be identified and estimated using the method developed in this paper.

7. See, for example, Assumption 6.2.2 of Fleming and Harrington (1991, p. 232). In principle, one could allow C to depend on X_1 and X_2. This would make the estimator and asymptotic theory more complicated because the conditional Kaplan–Meier estimator is then needed. See, for example, Dabrowska (1989) for details of the conditional Kaplan–Meier estimator.

8. Roughly speaking, this assumption requires that there is a chance of observing a complete spell no matter how large the spell is. This might not be palatable in some applications, and so we carry out Monte Carlo experiments that investigate how the proposed estimator performs when Assumption 7(b) is violated.

9. I am grateful to an anonymous referee who raised concern about this problem.

10. This result is not surprising; see, for example, Koul et al. (1981), Srinivasan and Zhou (1994), and Cheng et al. (1995) for cases of smaller asymptotic variances with estimated G_n. See also Wooldridge (2002) for similar results in the context of inverse probability weighted M-estimation for general selection problems.

11. One could estimate Ω using $\tilde{\Omega}$, where $\tilde{\Omega}$ is the same as $\hat{\Omega}$ without the weighting term $\Delta_{i1}\Delta_{i2}/G_n(Y_{i1} + Y_{i2})$. Instead we decide to use $\hat{\Omega}$ because it is expected that as a result of the use of weighting, $\hat{\Omega}$ might have a smaller variance than $\tilde{\Omega}$. This conjecture was confirmed by a small Monte Carlo experiment, although we did not calculate the asymptotic variances of $\hat{\Omega}$ and $\tilde{\Omega}$.

REFERENCES

Abbring, J.H, P.A. Chiappori, & J. Pinquet (2003) Moral hazard and dynamic insurance data. *Journal of the European Economic Association* 1, 767–820.

Billingsley, P. (1968) Weak convergence of probability measures. Wiley.

Chamberlain, G. (1985) Heterogeneity, omitted variable bias, and duration dependence. In J.J. Heckman & B. Singer (eds.), *Longitudinal Analysis of Labor Market Data*, pp. 3–38. Cambridge University Press.

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

Dabrowska, D. (1989) Uniform consistency of the kernel conditional Kaplan-Meier estimate. *Annals of Statistics* 17, 1157–1167.

Fleming, T.R. & D.P. Harrington (1991) *Counting Processes and Survival Analysis*. Wiley.

Gill, R. (1983) Large sample behaviour of the product-limit estimator on the whole line. *Annals of Statistics* 11, 49–58.

Heckman, J.J. & G.J. Borjas (1980) Does unemployment cause future unemployment? Definitions, questions and answers for a continuous time model of heterogeneity and state dependence. *Economica* 47, 247–283.

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.

Horowitz, J.L. & W. Härdle (1996) Direct semiparametric estimation of single-index models with discrete covariates. *Journal of the American Statistical Association* 91, 1632–1640.

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

Hristache, M., A. Juditski, & V. Spokoiny (2001) Direct estimation of the index coeffcients in a single index model. *Annals of Statistics* 29, 595–623.

Ichimura, H. (1993) Semiparametric least squares (SLS) and weighted SLS estimation of single index models. *Journal of Econometrics* 58, 71–120.

Jain, D. & N.J. Vilcassim (1991) Investigating household purchase timing decisions: A conditional hazard function approach. *Marketing Science* 10, 1–23.

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

Khan, S. & E. Tamer (2007) Partial rank estimation of duration models with general forms of censoring. *Journal of Econometrics* 136, 251–280.

Klein, R.W. & R.H. Spady (1993) An efficient semiparametric estimation for binary response models. *Econometrica* 61, 387–421.

Koul, H., V. Susarla, & J. Van Ryzin (1981) Regression analysis with randomly right-censored data. *Annals of Statistics* 9, 1276–1288.

Lancaster, T. (2000) The incidental parameter problem since 1948. *Journal of Econometrics* 95, 391–413.

Lin, D.Y., W. Sun, & Z. Ying (1999) Nonparametric estimation of the gap time distributions for serial events with censored data. *Biometrika* 86, 59–70.

Murphy, S.A. (1995) Asymptotic theory for the frailty model. *Annals of Statistics* 23, 182–198.

Newey, W.K. & D.L. McFadden (1994) Large sample estimation and hypothesis testing. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. IV, pp. 2111–2245. North-Holland.

Newman, J.L. & C.E. McCullogh (1984) A hazard rate approach to the timing of births. *Econometrika* 52, 939–961.

Powell, J.L. (1994) Estimation of semiparametric models. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. IV, pp. 2443–2521. North-Holland.

Powell, J.L., J.H. Stock, & T.M. Stoker (1989) Semiparametric estimation of index coefficients. *Econometrika* 51, 1403–1430.

Ridder, G. & İ. Tunalı (1999) Stratified partial likelihood estimation. *Journal of Econometrics* 92, 193–232.

Srinivasan, C. & M. Zhou (1994) Linear regression with censoring. *Journal of Multivariate Analysis* 49, 179–201.

Topel, R.H. & M.P. Ward (1992) Job mobility and the careers of young men. *Quarterly Journal of Economics* 107, 439–479.

Van der Berg, G.J. (2001) Duration models: Specification, identification, and multiple durations. In J.J. Heckman & E. Leamer (eds.), *Handbook of Econometrics*, vol. V, pp. 3381–3460. North-Holland.

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

Wang, W. & M.T. Wells (1998) Nonparametric estimation of successive duration times under dependent censoring. *Biometrika* 85, 561–572.

Wei, L.J., D.Y. Lin, & L. Weissfeld (1989) Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *Journal of the American Statistical Association* 84, 1065–1073.

Wooldridge, J.M. (2002) Inverse probability weighted M-estimators for sample selection, attrition, and stratification. *Portuguese Economic Journal* 1, 117–139.

Zhou, M. (1991) Some properties of the Kaplan-Meier estimator for independent non-identically distributed random variables. *Annals of Statistics* 19, 2266–2274.

APPENDIX: Proof of Theorem 1

It is assumed in the Appendix that Assumptions 1–7 hold. The following lemma is useful to prove Theorem 1.

LEMMA 1. *Let $\hat{S}_n(b)$ denote the left-hand side of (9). Then $\hat{S}_n(b)$ converges uniformly over b in probability to $S_0(b)$, where*

$$S_0(b) = \mathrm{E}[w_h(\Delta X'b)\Delta X\{L(\Delta X'\beta) - L(\Delta X'b)\}].$$

Proof of Lemma 1. Define $\bar{C}_n = \max_i\{Y_{i1} + Y_{i2}\}$. For any value of $\tau > 0$, write $\hat{S}_n(b) = \hat{S}_{n1}(b;\tau) + \hat{S}_{n2}(b;\tau)$, where

$$\hat{S}_{n1}(b;\tau) = n^{-1}\sum_{i=1}^{n}\left\{w_h(\Delta X_i'b)\Delta X_i 1(Y_{i1} + Y_{i2} \le \tau)\frac{\Delta_{i1}\Delta_{i2}}{G_n(Y_{i1} + Y_{i2})}\right.$$
$$\left.\times[1(Y_{i1} > Y_{i2}) - L(\Delta X_i'b)]\right\}$$

and

$$\hat{S}_{n2}(b;\tau) = n^{-1}\sum_{i=1}^{n}\left\{w_h(\Delta X_i'b)\Delta X_i 1(\tau < Y_{i1} + Y_{i2} \le \bar{C}_n)\frac{\Delta_{i1}\Delta_{i2}}{G_n(Y_{i1} + Y_{i2})}\right.$$
$$\left.\times[1(Y_{i1} > Y_{i2}) - L(\Delta X_i'b)]\right\}.$$

Let $S_n(b)$ denote the same expression as $\hat{S}_n(b)$ except that $G_n(Y_{i1} + Y_{i2})$ is replaced with $G(Y_{i1} + Y_{i2})$, so that $S_n(b) = S_{n1}(b;\tau) + S_{n2}(b;\tau)$, where

$$S_{n1}(b;\tau) = n^{-1}\sum_{i=1}^{n}\left\{w_h(\Delta X_i'b)\Delta X_i 1(Y_{i1} + Y_{i2} \le \tau)\frac{\Delta_{i1}\Delta_{i2}}{G(Y_{i1} + Y_{i2})}\right.$$
$$\left.\times[1(Y_{i1} > Y_{i2}) - L(\Delta X_i'b)]\right\}$$

and

$$S_{n2}(b;\tau) = n^{-1}\sum_{i=1}^{n}\left\{w_h(\Delta X_i' b)\Delta X_i 1(\tau < Y_{i1} + Y_{i2} \le \bar{C}_n)\,\frac{\Delta_{i1}\Delta_{i2}}{G(Y_{i1}+Y_{i2})}\right.$$
$$\left.\times\,[1(Y_{i1} > Y_{i2}) - L(\Delta X_i' b)]\right\}.$$

Finally, define

$$S_{01}(b;\tau) = \mathrm{E}[w_h(\Delta X' b)\Delta X 1(T_1 + T_2 \le \tau)\{L(\Delta X'\beta) - L(\Delta X' b)\}].$$

We show subsequently that (1) for each $\tau > 0$, $\hat{S}_{n1}(b;\tau)$ converges uniformly over b in probability to $S_{01}(b;\tau)$, (2) $\lim_{\tau\to\infty}\sup_{b\in\mathbf{B}}\|S_{01}(b;\tau) - S_0(b)\| = 0$, and (3) for each $\varepsilon > 0$, $\lim_{\tau\to\infty}\operatorname{limsup}_{n\to\infty}\Pr(\sup_{b\in\mathbf{B}}\|\hat{S}_n(b) - \hat{S}_{n1}(b;\tau)\| \ge \varepsilon) = 0$. Then the lemma follows from Theorem 4.2 of Billingsley (1968) and the fact that weak convergence to a constant term implies convergence in probability.

First, consider the limiting behavior of $\hat{S}_{n1}(b;\tau)$. Notice that $\sup\{c : G(c) > 0\} = \infty$. Thus, by the property of the Kaplan–Meier estimator (see, e.g., Fleming and Harrington, 1991), $G_n(c)$ converges to $G(c)$ uniformly on $[0,\tau]$ and $\{G_n(c) : c \in [0,\tau]\}$ and $\{G(c) : c \in [0,\tau]\}$ are bounded away from zero for sufficiently large n for any fixed but arbitrary $\tau > 0$. This implies that

$$\begin{aligned}|\hat{S}_{n1}(b;\tau) - S_{n1}(b;\tau)| &\le \sup_i\left|\frac{1(Y_{i1}+Y_{i2}\le\tau)}{G(Y_{i1}+Y_{i2})}\,\frac{G(Y_{i1}+Y_{i2}) - G_n(Y_{i1}+Y_{i2})}{G_n(Y_{i1}+Y_{i2})}\right|\\ &\qquad\times n^{-1}\sum_{i=1}^{n}\|\Delta X_i\|2|w_h(\Delta X_i' b)|\\ &= o_p(1)O_p(1) = o_p(1)\end{aligned}$$

uniformly over b for any fixed τ. In addition, because $\{G(c) : c \in [0,\tau]\}$ is bounded away from zero, by the uniform law of large numbers (e.g., Newey and McFadden, 1994, Lem. 2.4, p. 2129), $S_{n1}(b;\tau)$ converges uniformly over b in probability to $S_{01}(b;\tau)$ for each τ. Thus, we have proved part (1). It is obvious that part (2) holds.

Next, consider part (3). Because $\hat{S}_{n2}(b;\tau) = \hat{S}_n(b) - \hat{S}_{n1}(b;\tau)$, it suffices to show that for each $\varepsilon > 0$, $\lim_{\tau\to\infty}\operatorname{limsup}_{n\to\infty}\Pr(\sup_{b\in\mathbf{B}}\|\hat{S}_{n2}(b;\tau)\| \ge \varepsilon) = 0$. Notice that

$$\begin{aligned}|\hat{S}_{n2}(b;\tau) - S_{n2}(b;\tau)| &\le \sup_{Y_{i1}+Y_{i2}\le\bar{C}_n}\left|\frac{G(Y_{i1}+Y_{i2}) - G_n(Y_{i1}+Y_{i2})}{G_n(Y_{i1}+Y_{i2})}\right|\\ &\qquad\times n^{-1}\sum_{i=1}^{n}\left|1(Y_{i1}+Y_{i2}>\tau)\,\frac{\Delta_{i1}\Delta_{i2}}{G(Y_{i1}+Y_{i2})}\right|\|\Delta X_i\|2|w_h(\Delta X_i' b)|.\end{aligned} \tag{A.1}$$

By Zhou (1991, Thm. 2.2),

$$\sup_{c<\bar{C}_n} \left| \frac{G(c) - G_n(c)}{G_n(c)} \right| = O_p(1). \quad \textbf{(A.2)}$$

Taking $G_n(\cdot)$ to be a left-continuous version of the Kaplan–Meier estimator (i.e., $G_n(\cdot -) = G_n(\cdot)$; see also Srinivasan and Zhou, 1994, eqn. (5.7))

$$\sup_{Y_{i1}+Y_{i2}\leq\bar{C}_n} \left| \frac{G(Y_{i1}+Y_{i2}) - G_n(Y_{i1}+Y_{i2})}{G_n(Y_{i1}+Y_{i2})} \right| = O_p(1). \quad \textbf{(A.3)}$$

By Markov inequality, for any $M > 0$ and for any $\tau > 0$,

$$\begin{aligned}
&\Pr\left(n^{-1}\sum_{i=1}^{n} 1(Y_{i1}+Y_{i2}>\tau)\,\frac{\Delta_{i1}\Delta_{i2}}{[G(Y_{i1}+Y_{i2})]}\,\|\Delta X_i\| > M\right)\\
&\quad = \Pr\left(n^{-1}\sum_{i=1}^{n} \frac{1(T_{i1}+T_{i2}>\tau)\Delta_{i1}\Delta_{i2}}{[G(T_{i1}+T_{i2})]}\,\|\Delta X_i\| > M\right)\\
&\quad \leq M^{-1}\mathrm{E}\left[1(T_1+T_2>\tau)\,\frac{\Delta_1\Delta_2}{G(T_1+T_2)}\,\|\Delta X\|\right]\\
&\quad = M^{-1}\mathrm{E}[1(T_1+T_2>\tau)\|\Delta X\|]. \qquad \textbf{(A.4)}
\end{aligned}$$

Combining (A.3) and (A.4) with (A.1) gives that for each $\varepsilon > 0$,

$$\lim_{\tau\to\infty} \operatorname{limsup}_{n\to\infty} \Pr\left(\sup_{b\in\mathbf{B}}\|\hat{S}_{n2}(b;\tau) - S_{n2}(b;\tau)\| \geq \varepsilon\right) = 0. \quad \textbf{(A.5)}$$

In view of (A.4), it can also be shown that

$$\lim_{\tau\to\infty} \operatorname{limsup}_{n\to\infty} \Pr\left(\sup_{b\in\mathbf{B}}\|S_{n2}(b;\tau)\| \geq \varepsilon\right) = 0. \quad \textbf{(A.6)}$$

Therefore, we have proved part (3) and consequently the lemma also. ■

Proof of Theorem 1. It is obvious that $S_0(b)$ is continuous and is zero only when $b = \beta$. Therefore, in view of Lemma 1, b_n is consistent, that is, $b_n \to_p \beta$.

Now a first-order Taylor series approximation of $\hat{S}_n(b_n)$ at β gives

$$0 = n^{1/2}\hat{S}_n(b_n) = n^{1/2}\hat{S}_n(\beta) + \frac{\partial\hat{S}_n(b_n^*)}{\partial b}\,n^{1/2}(b_n - \beta), \quad \textbf{(A.7)}$$

where b_n^* is between b_n and β and $\partial\hat{S}_n/\partial b$ is the matrix whose (l,k) element is the partial derivative of the lth component of $\hat{S}_n$ with respect to the kth component of b. Let $\dot{w}_h(u) = dw_h(u)/du$. Notice that for any b,

$$\frac{\partial\hat{S}_n(b)}{\partial b} = T_{n1}(b) + T_{n2}(b),$$

where

$$T_{n1}(b) = n^{-1} \sum_{i=1}^{n} w_h(\Delta X_i' b) \frac{\Delta_{i1}\Delta_{i2}}{G_n(Y_{i1}+Y_{i2})} l(\Delta X_i' b)\Delta X_i \Delta X_i'$$

and

$$T_{n2}(b) = n^{-1} \sum_{i=1}^{n} \left\{ \dot{w}_h(\Delta X_i' b)\Delta X_i \Delta X_i' \frac{\Delta_{i1}\Delta_{i2}}{G_n(Y_{i1}+Y_{i2})} [1(Y_{i1} > Y_{i2}) - L(\Delta X_i' b)] \right\}.$$

By arguments similar to those used to prove Lemma 1 with the assumption that $\mathrm{E}\|\Delta X\|^2 < \infty$, we have

$$\sup_{b\in\mathbf{B}} \| T_{n1} - \mathrm{E}[w_h(\Delta X' b) l(\Delta X' b)\Delta X \Delta X'] \| = o_p(1)$$

and

$$\sup_{b\in\mathbf{B}} \| T_{n2} - \mathrm{E}[\dot{w}_h(\Delta X' b)\Delta X \Delta X'(L(\Delta X'\beta) - L(\Delta X' b))] \| = o_p(1).$$

Therefore, an application of the continuous mapping theorem yields

$$\left\| \frac{\partial \hat{S}_n(b_n^*)}{\partial b} - \Omega \right\| = o_p(1). \qquad \textbf{(A.8)}$$

Now consider $n^{1/2}\hat{S}_n(\beta)$. Using the same notation as in the proof of Lemma 1, write $\hat{S}_n(\beta) = S_n(\beta) + [\hat{S}_n(\beta) - S_n(\beta)]$. That is,

$$S_n(\beta) = n^{-1} \sum_{i=1}^{n} w_h(\Delta X_i'\beta)\Delta X_i \frac{\Delta_{i1}\Delta_{i2}}{G(Y_{i1}+Y_{i2})} [1(Y_{i1} > Y_{i2}) - L(\Delta X_i'\beta)].$$

Define $\hat{R}_n = \hat{S}_n(\beta) - S_n(\beta)$. For each $\tau > 0$, write $\hat{R}_n = \hat{R}_{n1}(\tau) + \hat{R}_{n2}(\tau)$, where

$$\hat{R}_{n1}(\beta;\tau) = n^{-1} \sum_{i=1}^{n} \left\{ w_h(\Delta X_i'\beta)\Delta X_i 1(Y_{i1}+Y_{i2} \le \tau) \left[\frac{\Delta_{i1}\Delta_{i2}}{G_n(Y_{i1}+Y_{i2})} - \frac{\Delta_{i1}\Delta_{i2}}{G(Y_{i1}+Y_{i2})} \right] \right.$$
$$\left. \times [1(Y_{i1} > Y_{i2}) - L(\Delta X_i'\beta)] \right\}$$

and

$$\hat{R}_{n2}(\beta;\tau) = n^{-1} \sum_{i=1}^{n} \left\{ w_h(\Delta X_i'\beta)\Delta X_i 1(Y_{i1}+Y_{i2} > \tau) \left[\frac{\Delta_{i1}\Delta_{i2}}{G_n(Y_{i1}+Y_{i2})} - \frac{\Delta_{i1}\Delta_{i2}}{G(Y_{i1}+Y_{i2})} \right] \right.$$
$$\left. \times [1(Y_{i1} > Y_{i2}) - L(\Delta X_i'\beta)] \right\}.$$

For each $\tau > 0$, let $\mathcal{S}_{01}(\tau)$ denote a random vector that is normally distributed with mean zero and covariance matrix $\mathrm{E}[\varphi_i(\tau)\varphi_i(\tau)']$, where

$$\varphi_i(\tau) = w_h(\Delta X_i'\beta)\Delta X_i \frac{\Delta_{i1}\Delta_{i2}}{G(Y_{i1}+Y_{i2})}[1(Y_{i1} > Y_{i2}) - L(\Delta X_i'\beta)] + \int_0^\infty \frac{\Gamma(s;\tau)}{\pi(s)}\, dM_i(s)$$

and

$$\Gamma(s;\tau) = \mathrm{E}\Big\{ w_h(\Delta X'\beta)\Delta X 1(Y_1 + Y_2 \leq \tau)\frac{\Delta_1\Delta_2}{G(Y_1+Y_2)} \times [1(Y_1 > Y_2) - L(\Delta X'\beta)]1(Y_1+Y_2 \geq s)\Big\}. \tag{A.9}$$

Finally, let $\mathcal{S}_0$ denote a random vector that is normally distributed with mean zero and covariance matrix Φ, which is defined in (10).

Again by Theorem 4.2 of Billingsley (1968), the theorem follows if we show that (1) for each $\tau > 0$, $n^{1/2}[S_n(\beta) + \hat{R}_{n1}(\beta;\tau)] \to_d \mathcal{S}_{01}(\tau)$, (2) $\mathcal{S}_{01}(\tau) \to_d \mathcal{S}_0$ as $\tau \to \infty$, and (3) for each $\varepsilon > 0$, $\lim_{\tau\to\infty}\limsup_{n\to\infty}\Pr(\|n^{1/2}\hat{R}_{n2}(\beta;\tau)\| \geq \varepsilon) = 0$.

We first show part (1). For any $c \leq \tau$, by a martingale integral representation for the Kaplan–Meier estimator (see, e.g., Fleming and Harrington, 1991),

$$\frac{G(c) - G_n(c)}{G_n(c)} = n^{-1}\sum_{k=1}^n \int_0^\infty \frac{1(c \geq s)}{\pi(s)}\, dM_k(s) + o_p(n^{-1/2}),$$

where $\pi(s)$ and $M_k(s)$ are defined in the main text. Using this, we have

$$\hat{R}_{n1}(\beta;\tau) = R_{n1}(\beta;\tau) + o_p(n^{-1/2})$$

for any arbitrary but fixed τ, where

$$R_{n1}(\beta;\tau) = n^{-1}\sum_{k=1}^n \int_0^\infty n^{-1}\sum_{i=1}^n S_{1i}(\beta;\tau)\frac{1}{\pi(s)}\, dM_k(s)$$

and

$$S_{1i}(\beta;\tau) = w_h(\Delta X_i'\beta)\Delta X_i 1(Y_{i1}+Y_{i2} \leq \tau)\frac{\Delta_{i1}\Delta_{i2}}{G(Y_{i1}+Y_{i2})}[1(Y_{i1} > Y_{i2}) - L(\Delta X_i'\beta)].$$

Then standard arguments for obtaining the projection of a U-statistic (see, e.g., Newey and McFadden, 1994, Lem. 8.4, p. 2201) give

$$R_{n1}(\beta;\tau) = n^{-1}\sum_{k=1}^n \int_0^\infty \frac{\Gamma(s;\tau)}{\pi(s)}\, dM_k(s) + o_p(n^{-1/2}),$$

where $\Gamma(s;\tau)$ is defined in (A.9). Thus, part (1) is proved. It is trivial to show that part (2) holds.

Now consider part (3). Write

$$\hat{R}_{n2}(\beta;\tau) = R_{n21}(\beta;\tau) + R_{n22}(\beta;\tau), \qquad \textbf{(A.10)}$$

where

$$R_{n21}(\beta;\tau) = n^{-1}\sum_{i=1}^{n} S_{2i}(\beta;\tau)\,\frac{G(Y_{i1}+Y_{i2}) - G_n(Y_{i1}+Y_{i2})}{G(Y_{i1}+Y_{i2})},$$

$$R_{n22}(\beta;\tau) = n^{-1}\sum_{i=1}^{n} S_{2i}(\beta;\tau)\,\frac{[G(Y_{i1}+Y_{i2}) - G_n(Y_{i1}+Y_{i2})]^2}{G(Y_{i1}+Y_{i2})G_n(Y_{i1}+Y_{i2})},$$

and

$$S_{2i}(\beta;\tau) = w_h(\Delta X_i'\beta)\Delta X_i 1(Y_{i1}+Y_{i2}>\tau)\,\frac{\Delta_{i1}\Delta_{i2}}{G(Y_{i1}+Y_{i2})}\,[1(Y_{i1}>Y_{i2}) - L(\Delta X_i'\beta)].$$

Note that

$$\|R_{n21}(\beta;\tau)\| \le \left[n^{-1}\sum_{i=1}^{n}\|S_{2i}(\beta;\tau)\|\right]\sup_{Y_{i1}+Y_{i2}\le\bar{C}_n}\left|\frac{G(Y_{i1}+Y_{i2}) - G_n(Y_{i1}+Y_{i2})}{G(Y_{i1}+Y_{i2})}\right|. \qquad \textbf{(A.11)}$$

In view of Theorem 1.1 of Gill (1983),

$$\sup_{Y_{i1}+Y_{i2}\le\bar{C}_n}\left|\frac{G(Y_{i1}+Y_{i2}) - G_n(Y_{i1}+Y_{i2})}{G(Y_{i1}+Y_{i2})}\right| = O_p(n^{-1/2}). \qquad \textbf{(A.12)}$$

Combining (A.4), (A.11), and (A.12) gives

$$\lim_{\tau\to\infty}\text{limsup}_{n\to\infty}\Pr(\|n^{1/2}\hat{R}_{n21}(\beta;\tau)\| \ge \varepsilon) = 0$$

for every $\varepsilon > 0$.

Now consider $R_{n22}(\beta;\tau)$. Note that

$$R_{n22}(\beta;\tau) \le \left[n^{-1}\sum_{i=1}^{n}\|S_{2i}(\beta;\tau)\|\right]\sup_{Y_{i1}+Y_{i2}\le\bar{C}_n}\left|\frac{G(Y_{i1}+Y_{i2}) - G_n(Y_{i1}+Y_{i2})}{G(Y_{i1}+Y_{i2})}\right|$$
$$\times \sup_{Y_{i1}+Y_{i2}\le\bar{C}_n}\left|\frac{G(Y_{i1}+Y_{i2}) - G_n(Y_{i1}+Y_{i2})}{G_n(Y_{i1}+Y_{i2})}\right|. \qquad \textbf{(A.13)}$$

Combining (A.3), (A.4), (A.12), and (A.13) gives

$$\lim_{\tau\to\infty}\text{limsup}_{n\to\infty}\Pr(\|n^{1/2}\hat{R}_{n22}(\beta;\tau)\| \ge \varepsilon) = 0$$

for every $\varepsilon > 0$. Thus, we have proved part (3).

It now remains to calculate the asymptotic variance, in particular Φ. First, note that by the variance calculation for a martingale (see, e.g., Fleming and Harrington, 1991, Thms. 2.4.5, 2.5.4)

$$\operatorname{var}\left\{\int_0^\infty \frac{\Gamma(s)}{\pi(s)}\, dM_i(s)\right\} = \int_0^\infty \frac{\Gamma(s)\Gamma(s)'}{\pi(s)}\, d\Lambda_C(s).$$

Furthermore,

$$\begin{aligned}
&2\operatorname{cov}\left\{w_h(\Delta X_i'\beta)\Delta X_i \frac{\Delta_{i1}\Delta_{i2}}{G(Y_{i1}+Y_{i2})}[1(Y_{i1}>Y_{i2})-L(\Delta X_i'\beta)], \int_0^\infty \frac{\Gamma(s)'}{\pi(s)}\, dM_i(s)\right\}\\
&\quad = -2\operatorname{cov}\left\{w_h(\Delta X_i'\beta)\Delta X_i \frac{\Delta_{i1}\Delta_{i2}}{G(Y_{i1}+Y_{i2})}[1(Y_{i1}>Y_{i2})-L(\Delta X_i'\beta)],\right.\\
&\qquad\qquad \left.\int_0^\infty 1(Y_{i1}+Y_{i2}\geq s)\frac{\Gamma(s)'}{\pi(s)}\, d\Lambda_c(s)\right\}\\
&\quad = -2\int_0^\infty \frac{\Gamma(s)\Gamma(s)'}{\pi(s)}\, d\Lambda_C(s).
\end{aligned}$$

Then the conclusion of the theorem follows immediately. ■