Journal of Multivariate Analysis **78**, 37–61 (2001)
doi:10.1006/jmva.2000.1944, available online at http://www.idealibrary.com on IDEAL®

Censored Partial Linear Models and Empirical Likelihood

Gengsheng Qin and Bing-Yi Jing[1]

Hong Kong University of Science and Technology, Kowloon, Hong Kong

Received March 15, 1999; published online April 6, 2001

Consider the partial linear model $Y_i = X_i^\tau \beta + g(T_i) + \varepsilon_i$, $i = 1, ..., n$, where β is a $p \times 1$ unknown parameter vector, g is an unknown function, X_i's are $p \times 1$ observable covariates, T_i's are other observable covariates in $[0, 1]$, and Y_i's are the response variables. In this paper, we shall consider the problem of estimating β and g and study their properties when the response variables Y_i are subject to random

particular, an empirical log-likelihood ratio for β is proposed and shown to have a limiting distribution of a weighted sum of independent chi-square distributions, which can be used to construct an approximate confidence region for β. Some simulation studies are conducted to compare the empirical likelihood and normal approximation-based method.

AMS 1991 subject classifications: 62E15; 62E20.

Key words and phrases: censored partial linear model; asymptotic normality; empirical likelihood.

1. INTRODUCTION

Consider the partial linear model

$$Y_i = X_i^\tau \beta + g(T_i) + \varepsilon_i, \qquad i = 1, ..., n, \tag{1.1}$$

where β is a $p \times 1$ unknown parameter vector, g is an unknown function, X_i's are $p \times 1$ observable covariates, T_i's are another observable covariates in $[0, 1]$, Y_i's are the response variables, and the residuals ε_i's are i.i.d. r.v.'s with zero mean and finite variance σ^2 and independent of (X_i^τ, T_i)'s. Following Speckman [31], we may further assume that (X_i^τ, T_i)'s are i.i.d. random vectors satisfying

$$X_i = \mu(T_i) + u_i, \qquad i = 1, ..., n, \tag{1.2}$$

[1] Research was supported by Hong Kong RGC CERG No. HKUST6162/97P.

0047-259X/01 $35.00

where $\mu(\cdot)$ is an unknown measurable function on $[0, 1]$, the residuals u_i's are i.i.d. random vectors with zero mean and positive definite variance-covariance matrix Σ_u and independent of T_i's. The partial linear model, as a semiparametric model, was introduced by Engle *et al.* [11] to allow the study of the effect of weather on electricity demand and further studied by Heckman [16], Speckman [31], Chen [3], Chen and Shiau [4, 5], Hong and Cheng [17, 18], Hamilton and Truong [15] and Mammen and Van de Geer [24] amongst others. Various estimators for β and $g(\cdot)$ were given by using different methods such as the kernel method, the penalized spline method, the piecewise constant smooth method, the smoothing splines, and the trigonometric series approach.

In this paper, we shall study the above partial linear model when the response variable Y_i may be censored from the right. So instead of observing (Y_i, X_i^τ, T_i), we now have observations $(Z_i, \delta_i, X_i^\tau, T_i)$, $i=1, ..., n$, where

$$Z_i = \min(Y_i, C_i), \qquad \delta_i = I(Y_i \leqslant C_i),$$

with C_i's being i.i.d. censoring r.v.'s that are independent of Y_i's and the censoring distribution G (i.e. the distribution of C_i's) being unknown. Under the censored partial linear model, our interest lies in the estimation of unknown parameter vector β and the unknown nonlinear function $g(\cdot)$. Recently, the censored partial linear model in which the censoring distribution G is assumed to be known is studied by Qin [28] and a central limit theorem for the estimator of β was proved by a method different from the method used in the present paper. In Section 2 of this paper, we shall consider estimating β by the least squares estimates and estimating g by the kernel regression method, respectively. Asymptotic normality of the LSE and consistency of the kernel regression estimate will be established. In Section 3, we shall investigate the application of the empirical likelihood method of Owen [25] to the censored partial linear model. An empirical log-likelihood ratio for β is defined and shown to have a limiting distribution of a weighted sum of independent chi-square r.v.'s with 1 degree of freedom. In order to use this to construct confidence region for β, one has to estimate the weights in the limiting distributions, which depend on the underlying distributions. Despite this, the estimated empirical likelihood in the censored partial linear regression problem still retains some of the good properties of the usual empirical likelihood of Owen [25], such as the range-preserving property. Theoretically, both the empirical likelihood and the normal approximation approach provide asymptotically correct coverage probability for β. So a small-scale simulation study is conducted in Section 4 to compare both methods.

In the standard survival analysis setting, the censored partial linear model is a special case of partial linear transformation models. The later is again a special case of generalized Cox proportional hazard model and arises frequently in regression diagnostics (see Dabrowska [10]). Special cases of the censored partial linear model have been considered by many researchers. For instance, if $g(t)=0$ in (1.1), the censored partial linear model reduces to usual censored linear model, which have been studied by Buckley and James [2], Koul *et al.* [21], Zheng [33, 34], Leurgans [23], and Lai *et al.* [22] among others. On the other hand, the case of non-parametric regression model with censored data is considered by Fan and Gijbels [12] by using local linear approximations. Finally, we note that the empirical likelihood method has been successfully applied in many areas, e.g., in linear regression models (Owen [27] and Chen [7, 8]) generalized linear models (Kolaczyk [20]), quantile (Chen and Hall [9]), general estimating equation (Qin and Lawless [29]), semiparametric model (Qin and Wong [30]), dependent processes (Kitamura [19]), and survival analysis (Adimari [1]).

2. LEAST SQUARES ESTIMATION AND ITS ASYMPTOTIC PROPERTIES

2.1. *The Methodology*

Let F and G be the distribution functions of Y_i and C_i, respectively. That is, $F(x)=P(Y_i\leqslant x)$ and $G(x)=P(C_i\leqslant x)$. Denote $\tau_F=\inf\{t\colon F(t)=1\}$ and $\tau_G=\inf\{t\colon G(t)=1\}$. Throughout this paper, we assume that $\tau_G\geqslant\tau_F$.

Because of the censoring, the usual methods for estimating β and $g(\cdot)$ can not be applied directly here. The problem is due to the fact that the censored observation Z_i and the true but unobservable r.v. Y_i have different expectations. To overcome this difficulty, assuming first that G is known, we define the transformed data

$$Y_{iG}=\frac{\delta_i Z_i}{1-G(Z_i)},\qquad i=1, ..., n.$$

Note that $E(Y_{iG}\mid X_i, T_i)=E(Y_i\mid X_i, T_i)=X_i^\tau\beta+g(T_i)$. Usually, however, the censoring d.f. G is unknown, in which case one can replace it by its Kaplan–Meier estimator G_n, resulting in

$$Y_{iG_n}=\frac{\delta_i Z_i}{1-G_n(Z_i)}.$$

This approach was first introduced by Koul *et al.* [21] and subsequently extended by Zheng [33] and Leurgans [23] in the context of censored linear regression model. For simplicity, we shall use the Y_{iG_n} defined above.

Denote $x^{\otimes 2} = xx^\tau$ for $x \in \mathbf{R}^p$. Also define

$$\tilde{X}_i = X_i - \sum_{j=1}^n W_{nj}(T_i)\, X_j, \qquad \tilde{Y}_{iG_n} = Y_{iG_n} - \sum_{j=1}^n W_{nj}(T_i)\, Y_{jG_n},$$

and

$$W_{nj}(t) = K\left(\frac{t-T_j}{h}\right) \Big/ \sum_{k=1}^n K\left(\frac{t-T_k}{h}\right),$$

where $K(\cdot)$ is a known nonnegative function, $h \equiv h(n) > 0$ is a sequence of bandwidths. Then the LSE of β, $\hat{\beta}_n$, and kernel regression estimator of g, $\hat{g}_n(t)$, are given by

$$\hat{\beta}_n = \arg\min_\beta \sum_{i=1}^n (\tilde{Y}_{iG_n} - \tilde{X}_i^\tau \beta)^2 = \left(\sum_{i=1}^n \tilde{X}_i^{\otimes 2}\right)^{-1} \sum_{i=1}^n \tilde{X}_i \tilde{Y}_{iG_n},$$
$$\hat{g}_n(t) = \sum_{j=1}^n W_{nj}(t)(Y_{jG_n} - X_j^\tau \hat{\beta}_n). \tag{2.1}$$

2.2. *Main Results*

Before stating the main results, we list the following conditions. For simplicity, we denote $\bar{K}(x) = 1 - K(x)$ for any distribution function $K(\cdot)$. Also, we shall use $\|\cdot\|$ to denote the Euclidean norm.

(A1) The functions $g(t)$, $\mu(t)$ and $g_1(t) \equiv E[Y_1 \mid T_1 = t]$ satisfy Lipschitz condition of order 1.

(A2) $\sup_t E[Y_1^2 \bar{G}^{-1}(Y_1) \mid T_1 = t] < \infty$ and $\sup_t E[|Y_1|\, \bar{G}^{-1}(Y_1) \mid T_1 = t] < \infty$.

(A3) T has density function $r(t)$ and $0 < \inf_{0 \leqslant t \leqslant 1} r(t) \leqslant \sup_{0 \leqslant t \leqslant 1} r(t) < \infty$.

(A4) $E\|X_1\|^4 < \infty$ and $E\varepsilon_1^4 < \infty$.

(A5) Let $\Gamma_1(t) = E(u_1 I(Y_1 > t))$. The d.f.'s F and G satisfy

$$\int_{-\infty}^{\tau_F} \frac{\left(\int_{t>y} t\, d\bar{F}(t)\right)^2}{\bar{G}^2(y)\, \bar{F}(y)}\, dG(y) < \infty \quad \text{and} \quad \int_{-\infty}^{\tau_F} \frac{\left\|\int_{t>y} t\, d\Gamma_1(t)\right\|^2}{\bar{G}^2(y)\, \bar{F}(y)}\, dG(y) < \infty.$$

(A6) For every $\varepsilon > 0$ there exists $y(\varepsilon) < \tau_F$ such that

$$\int_{u \in \mathbf{R}^p} \int_{y = y(\varepsilon)}^{\tau_F} \frac{\|u\| \, |y| \, dF(u, y)}{\bar{G}(y) \, \bar{F}^{1/2}(y)} < \varepsilon \qquad \text{and} \qquad \int_{y = y(\varepsilon)}^{\tau_F} \frac{|y| \, dF(y)}{\bar{G}(y) \, \bar{F}^{1/2}(y)} < \varepsilon$$

where $F(u, y) = P(u_1 < u, Y_1 < y)$.

(A7) There exists absolute constants M_1, M_2 and $\rho > 0$ such that

$$M_1 I[|t| \leqslant \rho] \leqslant K(t) \leqslant M_2 I[|t| \leqslant \rho],$$

and $K(\cdot)$ is uniformly continuous and of bounded variation on $[-\rho, \rho]$.

(A8) $\liminf nh^2/\log^4 n > 0$ and $nh^4 \to 0$.

(A9) $E(\|X_1\|^2 \, Y_1^2 \bar{G}^{-1}(Y_1)) < \infty$.

(A10) G is continuous, $\lim_{t \to \tau_F} \bar{F}^{-1}(t)(\int_t^{\tau_F} \bar{F}(dG)^{1-p} < \infty$ for some $0 < p < 1/2$.

(A11) G is continuous, $\bar{G}^{\alpha}(s) = O(\bar{F}(s-))$ as $s \uparrow \tau_G$ for some $0 \leqslant \alpha < 1$.

We shall now elaborate more on the above conditions. Conditions (A1)–(A4) and (A7)–(A8) are the usual conditions in the study of partial linear model, see Hong and Cheng [18], for example. Conditions (A5) and (A6) are similar to conditions (C2) and (C3) in Lai *et al.* [22] in the study of censored linear regression model. In the absence of censoring (i.e., $G(t) = 0$ for all $t < \tau_F$), conditions (A2) and (A9) reduce to $\sup_t E[Y_1^2 \mid T_1 = t] < \infty$ and $E(\|X_1\|^2 \, Y_1^2) < \infty$, respectively. Conditions (A10) and (A11) will be only used for obtaining the convergence rate of $\hat{g}_n(t)$, later in the proof of the theorem, and they will ensure the desired convergence rate of KM estimate G_n for G. Conditions (A10) is also a sufficient and necessary condition for the convergence of $G_n - G$ in the case $\tau_G > \tau_F$ (see Chen and Lo [6]). With Conditions (A11), Gu and Lai [14] proved the law of the iterated logarithm of G_n in the case of $\tau_G = \tau_F$, we do not know any other results on the rate of convergence of G_n for this case. Among these conditions, (A2), (A4), (A5) and (A9) are the moment conditions needed to derive the central limit theorem of $\hat{\beta}_n$, (A6), (A10) and (A11) are the conditions being used to control the behaviors near the endpoint τ_F of lifetime distribution F and censoring distribution G.

Define $V = \Sigma - V_1$, where

$$\Sigma = E(u_1^{\otimes 2}[Y_{1G} - E(Y_{1G} \mid X_1, T_1)]^2),$$

$$V_1 = \int_{-\infty}^{\tau_F} \left(\int_{t > y} t \, d\Gamma_1(t) \right)^{\otimes 2} \frac{dG(y)}{\bar{G}(y) \, \bar{F}(y-) \, \bar{G}(y-)}.$$

An estimate of V is given by $\hat{V}=\hat{\Sigma}-\hat{V}_1$, where

$$\hat{\Sigma}=n^{-1}\sum_{i=1}^{n}\tilde{X}_i^{\otimes 2}(\tilde{Y}_{iG_n}-\tilde{X}_i^{\tau}\hat{\beta}_n)^2$$

$$\hat{V}_1=n^{-1}\sum_{j=1}^{n}\left(\frac{\sum_{i=1}^{n}(X_i-\hat{\mu}_n(T_i))\,Y_{i\hat{g}_n}I(Z_i>Z_j)}{Y_n(Z_j)-\Delta N_n(Z_j)}\right)^{\otimes 2}\left(\sum_{i=1}^{n}I(Z_i>Z_j)\right)$$
$$\times\frac{\Delta N_n(Z_j)}{Y_n(Z_j)}$$

with $Y_n(s)=\sum_{i=1}^{n}I\{Z_i\geqslant s\}$, $N_n(s)=\sum_{i=1}^{n}I\{Z_i\leqslant s,\delta_i=0\}$ and $\hat{\mu}_n(t)=\sum_{j=1}^{n}W_{nj}(t)\,X_j$.

THEOREM 1. *Assume the conditions* (A1)–(A8) *hold.*

(i) *We have*

$$\sqrt{n}(\hat{\beta}_n-\beta)\xrightarrow{\mathscr{L}}N(0,\Sigma_u^{-1}V\Sigma_u^{-1}).$$

(ii) *If in addition* (A9) *holds, then the limiting covariance matrix* $\Sigma_u^{-1}V\Sigma_u^{-1}$ *in* (i) *can be consistently estimated by* $\hat{\Sigma}_u^{-1}\hat{V}\hat{\Sigma}_u^{-1}$, *where* $\hat{\Sigma}_u=n^{-1}\sum_{i=1}^{n}\tilde{X}_i^{\otimes 2}$.

(iii) *In the case of* $\tau_F<\tau_G$, *if in addition* (A10) *holds, then*

$$\hat{g}_n(t)-g(t)=O(h)+O_p((nh)^{-1/2})+O_p(n^{-p}),$$

which achieves the optimal rate $O_p(n^{-1/3})$ *when* $h=O(n^{-1/3})$ *and* $p=1/3$.

In the case of $\tau_F=\tau_G$, *if in addition* (A11) *holds, then*

$$\hat{g}_n(t)-g(t)=O(h)+O_p((nh)^{-1/2}),$$

which achieves the optimal rate $O_p(n^{-1/3})$ *when* $h=O(n^{-1/3})$.

Using Theorem 1, we can construct a large-sample $(1-\alpha)$-level confidence region for β as

$$I_{1\alpha}(\beta)=\{\beta:n(\hat{\beta}_n-\beta)^{\tau}\,\hat{\Sigma}_u\hat{V}^{-1}\hat{\Sigma}_u(\hat{\beta}_n-\beta)\leqslant\chi^2_{p,\alpha}\},\qquad(2.2)$$

where $\chi^2_{p,\alpha}$ is the $(1-\alpha)$th quantile of the chi-square distribution with p degree of freedom.

3. EMPIRICAL LIKELIHOOD METHOD

Let us define

$$U_i = (X_i - E[X_i \mid T_i])(Y_{iG} - E[Y_{iG} \mid T_i] - (X_i - E[X_i \mid T_i])^\tau \beta),$$
$$i = 1, ..., n.$$

It is easy to see that $EU_i = 0$, $i = 1, ..., n$, when β is the true parameter. Hence, the problem of testing whether β is the true parameter is equivalent to testing whether $EU_i = 0$, for $i = 1, ..., n$. By Owen [27], this may be done using empirical likelihood. Let $p_1, ..., p_n$ be nonnegative numbers summing to unity. Then, the empirical log-likelihood ratio, evaluated at true parameter value β, is defined by

$$l(\beta) = -2 \min_{\sum p_i U_i = 0} \sum \log(np_i). \tag{3.1}$$

Note that U_i in (3.1) depends on unknown functions G, $\mu(t)$ and $g_1(t) = E[Y_1 \mid T_1 = t]$, so $l(\beta)$ can not be used directly to make inference on β. To solve the problem, a natural way is to replace G by its KM estimator G_n, and replace $\mu(t)$ and $g_1(t)$ in $l(\beta)$ by their estimators $\hat{\mu}_n(t)$ and $\hat{g}_{1n}(t)$ defined respectively by

$$\hat{\mu}_n(t) = \sum_{j=1}^{n} W_{nj}(t)\, X_j, \qquad \hat{g}_{1n}(t) = \sum_{j=1}^{n} W_{nj}(t)\, Y_{jG_n}.$$

Then, the estimated empirical log-likelihood can be defined as

$$\tilde{l}(\beta) = -2 \min_{\sum p_i \tilde{U}_i = 0} \sum_{i=1}^{n} \log(np_i). \tag{3.2}$$

where $\tilde{U}_i = \tilde{X}_i(\tilde{Y}_{iG_n} - \tilde{X}_i^\tau \beta)$. Using Lagrange multipliers, the optimal value for p_i satisfying (3.2) may be shown to be

$$p_i = \frac{1}{n}(1 + \tilde{\lambda}^\tau \tilde{U}_i)^{-1},$$

where $\tilde{\lambda} = (\tilde{\lambda}_1, \tilde{\lambda}_2, ..., \tilde{\lambda}_n)^\tau$ is the solution of the equation

$$\frac{1}{n} \sum_{i=1}^{n} \frac{\tilde{U}_i}{1 + \lambda^\tau \tilde{U}_i} = 0. \tag{3.3}$$

The corresponding empirical log-likelihood ratio is then

$$\tilde{l}(\beta) = 2 \sum_{i=1}^{n} \log\{1 + \tilde{\lambda}^{\tau} \tilde{U}_i\}. \tag{3.4}$$

The following theorem indicates that the limiting distribution of empirical log-likelihood ratio for β is a weighted sum of independent chi-square distributions.

THEOREM 2. *Under the above conditions* (A1)–(A9), *if* β_0 *is the true value of* β, *then*

$$\tilde{l}(\beta_0) \xrightarrow{\mathscr{L}} l_1\chi^2_{1,1} + l_2\chi^2_{2,1} + \cdots + l_p\chi^2_{p,1},$$

where $\chi^2_{i,1}$*'s are* p *independent chi-squared distributions with one degree of freedom and the weights* l_i*'s,* $i = 1, 2, ..., p$, *are the eigenvalues of* $\Sigma^{-1}V$ *and can be consistently estimated by the corresponding eigenvalues* $\hat{l}_i$*'s of* $\hat{\Sigma}^{-1}\hat{V}$.

COROLLARY 3. *When* $p = 1$, *let* $\hat{l}(\beta) = \hat{r}\tilde{l}(\beta)$, *where* $\hat{r} = \hat{V}^{-1/2}\hat{\Sigma}$. *Then* $\hat{l}(\beta_0)$ *has an asymptotic chi-square distribution with* 1 *degree of freedom, that is,*

$$\hat{l}(\beta_0) \xrightarrow{\mathscr{L}} \chi^2_1.$$

Remark. When $\Gamma_1(t) = 0$, V reduces to Σ. It is easily seen that a sufficient condition for this is that the residuals u_i's are independent of Y_i's. Another case of V reducing to Σ is that there is no censoring in observations. In these two cases, we have $\tilde{l}(\beta_0) \xrightarrow{\mathscr{L}} \chi^2_p$.

Clearly, Theorem 2 can be used not only to test the hypothesis $H_0: \beta = \beta_0$, but also to construct confidence region for β. Let

$$I_{2\alpha}(\beta) = \{\beta: \tilde{l}(\beta) \leqslant c_\alpha\} \tag{3.5}$$

where c_α is the $(1-\alpha)$th quantile of the weighted chi-square distribution $\hat{l}_1\chi^2_{1,1} + \cdots + \hat{l}_p\chi^2_{p,1}$. In practice, one can get c_α through Monte Carlo simulation. Then, by Theorem 2, $I_{2\alpha}(\beta)$ gives an approximate confidence region for β with asymptotically correct coverage probability $1-\alpha$, i.e.,

$$P(\beta_0 \in I_{2\alpha}(\beta)) = 1 - \alpha + o(1).$$

4. A SIMULATION STUDY

We shall now conduct some simulation studies to compare the performance in terms of coverage probabilities between the empirical likelihood method and the normal approximation method based on $\hat{\beta}_n$. To do that, we first generate (Y_i, X_i^τ, T_i), $i=1, ..., n$, from the models (1.1) and (1.2), where we have taken $\beta^\tau=(5, 10)$, $g(t)=t^2$, $\mu(t)^\tau=(2t, t)$, $X_i^\tau=(X_{1i}, X_{2i})$ and $u_i^\tau=(u_{1i}, u_{2i})$. Furthermore, T_i's are drawn from the uniform distribution $U[0, 1]$, u_{1i} and u_{2i} are generated from the standard normal distribution $N(0, 1)$, respectively. Two models are considered in this simulation study. In model A, the error ε_i's are generated from the standard normal distribution $N(0, 1)$. In model B, the error ε_i's are generated from the centered chi-square distribution with degree of freedom 20, i.e., ε_i's are iid. $\chi^2_{20}-20$. On the other hand, the censoring times C_i's are generated from the exponential distribution with rate 0.02 (the censoring proportion is about 20%). Finally, simulated observations from the censored partial linear model are $(Z_i, \delta_i, X_i^\tau, T_i)$ for $i=1, ..., n$, where

$$Z_i=\min(Y_i, C_i), \qquad \delta_i=I\{Y_i \leqslant C_i\}.$$

We can repeat the above process B times to generate B sets of data, where B is chosen to be 2000 here. Then the approximate coverage probabilities for the empirical likelihood and normal approximation methods based on these 2000 simulated data sets are simply the proportions of these data sets which satisfy the inequalities (3.5) and (2.2), respectively.

From the expression of estimator $\hat{V}$ for variance-covariance matrix V, we can see that the denominator $Y_n(Z_{(i)})-\Delta N_n(Z_{(i)})$ in the definition of $\hat{V}$ can be zero when $i=n$, so we add 0.5 to $Y_n(Z_{(n)})-\Delta N_n(Z_{(n)})$ whenever the term becomes zero in our simulation study, these modifications of the variance-covariance estimation will not change our large sample results. We should mention that in our simulation study we have chosen the kernel function $K(t)$ to be the biweight kernel

$$K(x)=\tfrac{15}{16}(1-x^2)^2, \qquad |x| \leqslant 1.$$

Two different bandwidths of h are selected to be $5n^{-1/3}$ and $15n^{-1/3}$. The sample size n has been chosen to be 60 and 100, respectively. The nominal confidence level α has been taken to be 0.90, 0.95 and 0.99, respectively. The results of the simulation are presented in Tables I and II.

We make the following observations from these tables.

(1) For the empirical likelihood method, the coverage accuracies for β in both models increase as the sample size n increases. For the normal approximation method, the coverage accuracies for β in model B increase

TABLE I

Coverage Probability Comparisons for β by Empirical Likelihood and Normal Approximation under Model A, Where $\varepsilon \sim N(0, 1)$

		$h = 5n^{-1/3}$		$h = 15n^{-1/3}$	
Nominal levels	n	Normal approximation	Empirical likelihood	Normal approximation	Empirical likelihood
0.90	60	0.865	0.853	0.859	0.854
	100	0.843	0.890	0.830	0.895
0.95	60	0.900	0.911	0.894	0.911
	100	0.883	0.938	0.864	0.942
0.99	60	0.941	0.973	0.938	0.971
	100	0.929	0.980	0.911	0.979

as the sample size n increases, but the coverage accuracies for β in model A start to decrease as n gets large, part of reason could be that the variance estimator is not very stable. However, a more detailed study seems worthwhile.

(2) Generally, the empirical likelihood method performs better than the normal approximation method for moderate sample size ($n = 60$, 100), particularly when the error distribution is not symmetric (see Table II).

(3) The empirical likelihood method and the normal approximation based method have very similar performances for the two chosen different bandwidths.

TABLE II

Coverage Probability Comparisons for β by Empirical Likelihood and Normal Approximation under Model B, Where $\varepsilon \sim \chi^2_{20} - 20$

		$h = 5n^{-1/3}$		$h = 15n^{-1/3}$	
Nominal levels	n	Normal approximation	Empirical likelihood	Normal approximation	Empirical likelihood
0.90	60	0.809	0.845	0.813	0.841
	100	0.823	0.873	0.810	0.875
0.95	60	0.867	0.917	0.862	0.903
	100	0.877	0.930	0.861	0.932
0.99	60	0.926	0.973	0.918	0.969
	100	0.929	0.981	0.925	0.982

We also investigate the performances of empirical likelihood method and the normal approximation based method in fixed sample size ($n = 100$) but with different censoring proportions being controlled by choosing different censoring times (Here the censoring times are generated from the exponential distributions with different rates). The result is reported in Tables III–IV.

Tables III and IV indicate that:

(1) The empirical likelihood outperforms the normal approximation methods in both models. Particularly, the empirical likelihood performs much better than the normal approximation method under higher censoring proportions.

(2) At each nominal level and for both the empirical likelihood and normal approximation methods, the coverage accuracies for β in both models decrease as the censoring proportions increase. However, the coverage accuracies for the empirical likelihood are acceptable even under higher censoring proportions.

TABLE III

Coverage Probability Comparisons under Different Censoring Proportions for Models A and B, Where $n = 100$ and $h = 5n^{-1/3}$

Nominal levels	Models	Censoring proportion	Normal approximation	Empirical likelihood
0.90	A	10%	0.922	0.893
		20%	0.843	0.890
		30%	0.766	0.848
	B	10%	0.841	0.876
		20%	0.823	0.873
		30%	0.765	0.839
0.95	A	10%	0.935	0.940
		20%	0.883	0.938
		30%	0.817	0.903
	B	10%	0.879	0.938
		20%	0.877	0.930
		30%	0.835	0.893
0.99	A	10%	0.957	0.985
		20%	0.929	0.980
		30%	0.888	0.962
	B	10%	0.930	0.982
		20%	0.929	0.981
		30%	0.917	0.962

TABLE IV

Coverage Probability Comparisons under Different Censoring Proportions for Models A and B, Where $n=100$ and $h=15n^{-1/3}$

Nominal levels	Models	Censoring proportion	Normal approximation	Empirical likelihood
0.90	A	10%	0.899	0.899
		20%	0.830	0.895
		30%	0.756	0.855
	B	10%	0.828	0.880
		20%	0.810	0.875
		30%	0.753	0.833
0.95	A	10%	0.927	0.943
		20%	0.864	0.942
		30%	0.806	0.912
	B	10%	0.865	0.937
		20%	0.861	0.932
		30%	0.820	0.894
0.99	A	10%	0.952	0.980
		20%	0.911	0.979
		30%	0.873	0.960
	B	10%	0.928	0.984
		20%	0.925	0.982
		30%	0.911	0.963

5. PROOF OF THEOREMS

Before proving the main theorems, we give a series of lemmas.

LEMMA 1. *Assume that* (A4), (A7), *and* (A8) *hold. Then*

(i) $$\max_{1\leqslant i\leqslant n}\left\|\sum_{j=1}^{n} W_{nj}(T_i)\,u_j\right\| = o((\log n)^{-1}),\ a.s.$$

(ii) $$\max_{1\leqslant j\leqslant n}\left\|\sum_{i=1}^{n} W_{nj}(T_i)\,u_i\right\| = o((\log n)^{-1}),\ a.s.$$

(iii) $$\max_{1\leqslant j\leqslant n}\sum_{i=1}^{n} W_{nj}(T_i) = O(1),\ a.s.$$

(iv) $$EW_{ni}^2(t) = O((n^2h)^{-1}).$$

Proof. Proofs of (iii) and (iv) can be found in Hong and Cheng [18] and Qin [28] respectively. The proofs of (i) and (ii) are similar to that of Lemma 2 in Hong and Cheng [18] and hence are omitted here.

LEMMA 2. *If $f(t)$ satisfies Lipschitz condition of order* 1 *and* (A7) *holds, then*

$$\sup_{0\leqslant t\leqslant 1} \|\tilde{f}(t)\| \equiv \sup_{0\leqslant t\leqslant 1} \left\| f(t)-\sum_{j=1}^{n} W_{nj}(t)\, f(T_j)\right\| = O(h).$$

Proof. For each $t\in[0,1]$, let $D_t=\{j\colon |T_j-t|\leqslant \rho h,\ j=1, ..., n\}$. By (A7),

$$\|\tilde{f}(t)\| = \left\|\sum_{j=1}^{n} W_{nj}(t)(f(t)-f(T_j))\right\| = \left\|\sum_{j\in D_t} W_{nj}(t)(f(t)-f(T_j))\right\|$$

$$\leqslant \sum_{j\in D_t} W_{nj}(t)\, \|f(t)-f(T_j)\| \leqslant Ch \sum_{j\in D_t} W_{nj}(t) \leqslant Ch.$$

Since Ch does not depend on t, we get $\sup_{0\leqslant t\leqslant 1} \|\tilde{f}(t)\| = O(h)$.

LEMMA 3. *Let $\tilde{U}_{(n)} = \max_i \|\tilde{U}_i\|$. Assume that* (A1)–(A4) *and* (A7)–(A9) *hold. Then $\tilde{U}_{(n)} = o_p(n^{1/2})$.*

Proof. Let

$$\tilde{Y}_{iG} = Y_{iG} - \sum_{j=1}^{n} W_{nj}(T_i)\, Y_{jG},$$

$$\tilde{Y}_i = Y_i - \sum_{j=1}^{n} W_{nj}(T_i)\, Y_j, \qquad U_{i0} = \tilde{X}_i(\tilde{Y}_i - \tilde{X}_i^{\tau} \beta).$$

Since

$$\begin{aligned}\tilde{U}_i &= \tilde{X}_i(\tilde{Y}_{iG_n} - \tilde{X}_i^{\tau} \beta)\\ &= \tilde{X}_i(\tilde{Y}_{iG_n} - \tilde{Y}_{iG}) + \tilde{X}_i(\tilde{Y}_{iG} - \tilde{Y}_i) + U_{i0},\end{aligned}$$

we have

$$\tilde{U}_{(n)} \leqslant \max_i \|U_{i0}\| + \max_i \|\tilde{X}_i\|\, |\tilde{Y}_{iG_n} - \tilde{Y}_{iG}| + \max_i \|\tilde{X}_i\|\, |\tilde{Y}_i - \tilde{Y}_i|. \qquad (5.1)$$

For the first term in the right side of (5.1), from $Y_i = X_i\beta + g(T_i) + \varepsilon_i$ and the definitions of $\widetilde{X}_i$ and $\widetilde{Y}_i$, it follows that

$$\begin{aligned}\max_i \|U_{i0}\| \leqslant{}& \max_i \Big\|X_i - \sum_j W_{nj}(T_i)\, X_j\Big\| \max_i \Big|g(T_i) - \sum_j W_{nj}(T_i)\, g(T_j)\Big| \\ &+ \max_i \|X_i\varepsilon_i\| + \max_i \|X_i\| \max_i \Big|\sum_j W_{nj}(T_i)\,\varepsilon_j\Big| \\ &+ \max_i |\varepsilon_i| \max_i \Big|\sum_j W_{nj}(T_i)\,\varepsilon_j\Big| \\ &+ \max_i \left(\Big\|\sum_{j_1} W_{nj_1}(T_i)\, X_{j_1}\Big\|\right) \max_i \Big|\sum_{j_2} W_{nj_2}(T_i)\,\varepsilon_{j_2}\Big|. \end{aligned} \tag{5.2}$$

By Lemma 3 of Owen [26] and (A4), we have

$$\max_i \|X_i\| = o(n^{1/2}), \qquad \max_i \|X_i\varepsilon_i\| = o(n^{1/2}), \qquad \max_i |\varepsilon_i| = o(n^{1/2}). \tag{5.3}$$

Similar to the proof of Lemma 1(i), we can show that

$$\max_i \Big|\sum_{j=1}^{n} W_{nj}(T_i)\,\varepsilon_j\Big| = o((\log n)^{-1}), \ a.s.$$

By (A1), we have $\max_i \|\sum_j W_{nj}(T_i)\,\mu(T_j)\| = O(1)$ a.s. Combining this and Lemma 1(i), we get

$$\max_i \Big\|\sum_j W_{nj}(T_i)\, X_j\Big\| = O(1), \ a.s. \tag{5.4}$$

Therefore,

$$\max_i \|U_{i0}\| = o_p(n^{1/2}). \tag{5.5}$$

Noting that (A2) and (A9) imply $E\,|Y_{1G}|^2 < \infty$ and $E\,\|X_1 Y_{1G}\|^2 < \infty$, thus by Lemma 3 of Owen [26], we have

$$\|X_1 Y_{1G}\| = o(n^{1/2}), \qquad |Y_{1G}| = o(n^{1/2}).$$

Applying (5.4) and the result due to Srinivasan and Zhou [32],

$$\sup_{t \leqslant \max_i Z_i} \left|\frac{G_n(t) - G(t)}{1 - G_n(t)}\right| = O_p(1), \tag{5.6}$$

we get

$$\max_i \|\tilde{X}_i\| \, |\tilde{Y}_{iG_n} - \tilde{Y}_{iG}| \tag{5.7}$$

$$\leqslant \max_i \|\tilde{X}_i\| \left(|Y_{iG_n} - Y_{iG}| + \left| \sum_{j=1}^{n} W_{nj}(T_i)(Y_{jG_n} - Y_{jG}) \right| \right)$$

$$\leqslant 2 \max_i \|\tilde{X}_i\| \, |Y_{iG_n} - Y_{iG}|$$

$$\leqslant 2 \max_i \|X_i\| \, |Y_{iG_n} - Y_{iG}| + 2 \max_i \left(\left\| \sum_{j=1}^{n} W_{nj}(T_i) X_j \right\| |Y_{iG_n} - Y_{iG}| \right)$$

$$\leqslant 2 \max_i \|X_i\| \, |Y_{iG_n} - Y_{iG}| + 2 \max_i |Y_{iG_n} - Y_{iG}| \left(\max_i \left\| \sum_{j=1}^{n} W_{nj}(T_i) X_j \right\| \right)$$

$$\leqslant 2 \sup_{t \leqslant \max_i Z_i} \left| \frac{G_n(t) - G(t)}{1 - G_n(t)} \right| (\max_i \|X_i Y_{iG}\| + \max_i |Y_{iG}| \cdot O(1))$$

$$= o(n^{1/2}). \tag{5.8}$$

Similarly, for the third term in the right side of (5.1), we have

$$\max_i \|\tilde{X}_i\| \, |\tilde{Y}_{iG} - \tilde{Y}_i|$$

$$\leqslant 2 \max_i \|\tilde{X}_i\| \, |Y_{iG} - Y_i|$$

$$\leqslant 2 \max_i \|X_i\| \, |Y_{iG} - Y_i| + 2 \max_i \left(\left\| \sum_{j=1}^{n} W_{nj}(T_i) X_j \right\| |Y_{iG} - Y_i| \right)$$

$$\leqslant 2(\max_i \|X_i Y_{iG}\| + \max_i \|X_i Y_i\| + O(1) \cdot (\max_i |Y_{iG}| + \max_i |Y_i|))$$

$$= o(n^{1/2}). \tag{5.9}$$

Then, Lemma 3 follows from (5.1), (5.5)–(5.9).

LEMMA 4. *Assume* (A1)–(A8) *hold.*

(i) *We have* $n^{-1} \sum_{i=1}^{n} \tilde{X}_i^{\otimes 2} \to \Sigma_u$ *a.s.*

(ii) *Let* $\tilde{U}_{i0} = \tilde{X}_i(\tilde{Y}_{iG} - \tilde{X}_i^{\tau} \beta)$ *and* $\hat{\varepsilon}_i = Y_{iG} - E(Y_{iG} \mid X_i, T_i)$, *then we have*

$$n^{-1/2} \sum_{i=1}^{n} \tilde{U}_{i0} = n^{-1/2} \sum_{i=1}^{n} u_i \hat{\varepsilon}_i + O_p((nh^2)^{-1/2}) + O_p(n^{1/2} h^2).$$

(iii) $n^{-1/2} \sum_{i=1}^{n} \tilde{U}_i \xrightarrow{\mathscr{L}} N(0, V)$.

(iv) *If in addition* (A9) *holds, then* $n^{-1} \sum_{i=1}^{n} \tilde{U}_i^{\otimes 2} = n^{-1} \sum_{i=1}^{n} (u_i \hat{\varepsilon}_i)^{\otimes 2} + o_p(1)$.

Proof. Part (i) is the conclusion of Lemma 4 in Hong and Cheng [18]. Similar to the proof of Theorem 2.1 in Qin [28], we can get (ii). For (iii), we have the decomposition

$$n^{-1/2} \sum_{i=1}^{n} \tilde{U}_i = n^{-1/2} \sum_{i=1}^{n} \tilde{U}_{i0} + I$$
$$= n^{-1/2} \sum_{i=1}^{n} u_i \hat{\varepsilon}_i + n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{\tau_F} A(y)\, dM_i(y) + R_n, \tag{5.10}$$

where

$$I = n^{-1/2} \sum_{i=1}^{n} \tilde{X}_i (\tilde{Y}_{iG_n} - \tilde{Y}_{iG}),$$

$$A(y) = -\int_{t>y} t\, d\Gamma_1(t)/(\bar{G}(y)\, \bar{F}(y-)),$$

$$\Lambda(s) = \int_{-\infty}^{s} (1 - G(t-))^{-1}\, dG(t),$$

$$M_i(s) = I(Z_i \leqslant s, \delta_i = 0) - \int_{-\infty}^{s} I(C_i \geqslant s,\ Y_i > s)\, d\Lambda(s),$$

$$R_n = I - n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{\tau_F} A(y)\, dM_i(y) + O_p((nh^2)^{-1/2}) + O_p(n^{1/2} h^2).$$

Note that the first term in the right side of (5.10), which is a sum of i.i.d. r.v.'s, has a limiting normal distribution with zero mean and variance–covariance matrix $E(u_1 \hat{\varepsilon}_1)^{\otimes 2}$, and the second term, which is a martingale, has a limiting normal distribution with zero mean and variance V_1 by Rebolledo's martingale central limit theorem. If $R_n = o_p(1)$, then (iii) follows by a similar argument to the proof of Theorem 2 in Lai *et al.* [22]. So we only need to prove $R_n = o_p(1)$. Write

$$\tilde{U}_i = u_i - \sum_{j=1}^{n} W_{nj}(T_i)\, u_j, \qquad \tilde{\mu}_i = \mu(T_i) - \sum_{j=1}^{n} W_{nj}(T_i)\, \mu(T_j).$$

Then we have

$$
\begin{aligned}
I &= n^{-1/2} \sum_{j=1}^{n} \left(\tilde{U}_j - \sum_{i=1}^{n} W_{nj}(T_i)\, \tilde{u}_i \right) (Y_{jG_n} - Y_{jG}) \\
&\quad + n^{-1/2} \sum_{j=1}^{n} \left(\tilde{\mu}_j - \sum_{i=1}^{n} W_{nj}(T_i)\, \tilde{\mu}_i \right) (Y_{jG_n} - Y_{jG}) \\
&= n^{-1/2} \sum_{j=1}^{n} u_j (Y_{jG_n} - Y_{jG}) + n^{-1/2} \sum_{j=1}^{n} \left(\tilde{\mu}_j - \sum_{i=1}^{n} W_{nj}(T_i)\, \tilde{\mu}_i \right) (Y_{jG_n} - Y_{jG}) \\
&\quad + n^{-1/2} \sum_{j=1}^{n} \left[\sum_{i=1}^{n} W_{ni}(T_j)\, u_i + \sum_{i=1}^{n} W_{nj}(T_i)\, u_i \right. \\
&\quad \left. + \sum_{i=1}^{n} \left(W_{nj}(T_i) \sum_{k=1}^{n} W_{nk}(T_i)\, u_k \right) \right] (Y_{jG_n} - Y_{jG}) \\
&\equiv I_1 + I_2 + I_3 .
\end{aligned}
\tag{5.11}
$$

First we investigate I_1. Applying the following martingale representation (Gill [13])

$$
\frac{G_n(t) - G(t)}{1 - G(t)} = \int_{-\infty}^{t} \frac{1 - G_n(t-)}{1 - G(t)} \frac{dM(t)}{Y_n(t)} = \int_{-\infty}^{t} D_n(t)\; Y_n^{-1}(t)\, dM(t),
\tag{5.12}
$$

where $D_n(s) = (1 - G_n(s-))/(1 - G(s))$ and $M(s) = \sum_{i=1}^{n} M_i(s)$, we have

$$
I_1 = n^{-1/2} \sum_{j=1}^{n} \frac{u_j \delta_j Z_j}{1 - G_n(t-)} \int_{-\infty}^{Z_j} D_n(t)\; Y_n^{-1}(t)\, dM(t).
$$

Similar to the proofs of (2.28) and (2.29) in Lai *et al.* [22], we can show that

$$
I_1 = n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{\tau_F} A(y)\, dM_i(y) + o_p(1).
\tag{5.13}
$$

Now let us look at I_2 and I_3. Note that

$$
\begin{aligned}
|Y_{jG_n} - Y_{jG}| &= \left| Y_{jG} \left(\left(1 - \frac{G_n(Z_j) - G(Z_j)}{1 - G(Z_j)} \right)^{-1} - 1 \right) \right| \\
&= \left| \frac{Y_{jG}}{1 - G(Z_j)} \left((G_n(Z_j) - G(Z_j)) + \frac{(G_n(Z_j) - G(Z_j))^2}{1 - G_n(Z_j)} \right) \right| \\
&\leqslant \left(1 + \sup_{t \leqslant \max_i Z_i} \frac{|G_n(t) - G(t)|}{|1 - G_n(t)|} \right) \left| \frac{G_n(Z_j) - G(Z_j)}{1 - G(Z_j)} \right| |Y_{jG}|.
\end{aligned}
\tag{5.14}
$$

By Lemmas 1 and 2, (5.6), (5.2), and (5.14), we get

$$|I_2| \leqslant n^{-1/2} \sup_{t \in [0,1]} \left\| \mu(t) - \sum_{j=1}^{n} W_{nj}(t)\, \mu(T_j) \right\|$$

$$\times \sum_{j=1}^{n} \left(1 + \sum_{i=1}^{n} W_{nj}(T_i)\right) |Y_{jG_n} - Y_{jG}|$$

$$\leqslant Cn^{-1/2} h \sum_{j=1}^{n} |Y_{jG_n} - Y_{jG}|$$

$$\leqslant Chn^{-1/2} \sum_{j=1}^{n} |Y_{jG}| \left| \int_{-\infty}^{Z_j} D_n(t)\, Y_n^{-1}(t)\, dM(t) \right|.$$

$$|I_3| \leqslant Cn^{-1/2} (\log^{-1} n) \sum_{j=1}^{n} |Y_{jG_n} - Y_{jG}|$$

$$\leqslant C(\log^{-1} n)\, n^{-1/2} \sum_{j=1}^{n} |Y_{jG}| \left| \int_{-\infty}^{Z_j} D_n(t)\, Y_n^{-1}(t)\, dM(t) \right|.$$

Using (A5) and an argument similar to the proofs of (2.28) and (2.29) in Lai *et al.* [22], we can show that

$$|I_2| \leqslant Ch \left| n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{\tau_F} A_1(y)\, dM_i(y) \right| + o_p(1) = o_p(1), \tag{5.15}$$

$$|I_3| \leqslant C(\log^{-1} n) \left| n^{-1/2} \sum_{i=1}^{n} \int_{-\infty}^{\tau_F} A_1(y)\, dM_i(y) \right| + o_p(1) = o_p(1), \tag{5.16}$$

where $A_1(y) = -\int_{t>y} |t|\, d\bar{F}(t)/(\bar{G}(y)\, \bar{F}(y-))$. From (5.11), (5.13), (5.15) and (5.16), it follows that $R_n = o_p(1)$. Next we turn to prove (iv). For any $a \in \mathbf{R}^p$, we have the decomposition

$$n^{-1} \sum_{i=1}^{n} (a^\tau \tilde{U}_i)^2 = n^{-1} \sum_{i=1}^{n} (a^\tau (\tilde{U}_i - \tilde{U}_{i0}))^2 + 2n^{-1} \sum_{i=1}^{n} (a^\tau (\tilde{U}_i - \tilde{U}_{i0}))(a^\tau \tilde{U}_{i0})$$

$$+ n^{-1} \sum_{i=1}^{n} (a^\tau \tilde{U}_{i0})^2$$

$$\equiv J_1 + J_2 + J_3, \tag{5.17}$$

where

$$J_1 = n^{-1} \sum_{i=1}^{n} (a^\tau \widetilde{X}_i)^2 (\widetilde{Y}_{iG_n} - \widetilde{Y}_{iG})^2$$

$$\leqslant \|a\|^2 (n^{-1/2} \max_i \|\widetilde{X}_i(\widetilde{Y}_{iG_n} - \widetilde{Y}_{iG})\|) \left(n^{-1/2} \sum_{i=1}^{n} \|\widetilde{X}_i(\widetilde{Y}_{iG_n} - \widetilde{Y}_{iG})\|\right),$$

$$J_2 \leqslant 2 \|a\|^2 (n^{-1/2} \max_i \|U_{i0}\|) \left(n^{-1/2} \sum_{i=1}^{n} \|\widetilde{X}_i(\widetilde{Y}_{iG_n} - \widetilde{Y}_{iG})\|\right).$$

Similarly to (5.11), (5.13), (5.15), and (5.16), we can show that

$$n^{-1/2} \sum_{i=1}^{n} \|\widetilde{X}_i(\widetilde{Y}_{iG_n} - \widetilde{Y}_{iG})\| = O_p(1).$$

Then by (5.5) and (5.8), we get

$$J_1 = o_p(1), \qquad J_2 = o_p(1). \tag{5.18}$$

For the term J_3, writing $\tilde{\varepsilon}_i = \sum_{j=1}^{n} W_{nj}(T_i)\, \hat{\varepsilon}_j$, we have

$$J_3 = n^{-1} \sum_{i=1}^{n} (a^\tau \widetilde{X}_i)^2 (\hat{\varepsilon}_i - \tilde{\varepsilon}_i + \tilde{g}(T_i))^2$$

$$= n^{-1} \sum_{i=1}^{n} (a^\tau \widetilde{X}_i)^2 \hat{\varepsilon}_i^2 + n^{-1} \sum_{i=1}^{n} (a^\tau \widetilde{X}_i)^2 \tilde{\varepsilon}_i^2 - 2n^{-1} \sum_{i=1}^{n} (a^\tau \widetilde{X}_i)^2 \hat{\varepsilon}_i \tilde{\varepsilon}_i$$

$$+ n^{-1} \sum_{i=1}^{n} (a^\tau \widetilde{X}_i)^2 \tilde{g}^2(T_i) + 2n^{-1} \sum_{i=1}^{n} (a^\tau \widetilde{X}_i)^2 \tilde{g}(T_i)(\hat{\varepsilon}_i - \tilde{\varepsilon}_i)$$

$$\equiv J_{31} + J_{32} + J_{33} + J_{34} + J_{35},$$

where $\tilde{g}(t) = g(t) - \sum_{j=1}^{n} W_{nj}(t)\, g(T_j)$. Similar to Lemma 1(i), we can prove

$$\max_i |\tilde{\varepsilon}_i| = \max_i \left| \sum_{j=1}^{n} W_{nj}(T_i)\, \hat{\varepsilon}_j \right| = o(\log^{-1} n), \ a.s. \tag{5.19}$$

Using (5.19), Lemma 2, and Lemma 4(i), we can get

$$J_{32} = o_p(\log^{-2} n), \qquad J_{33} = o_p(\log^{-1} n), \qquad J_{34} = O_p(h^2), \qquad J_{35} = O_p(h).$$

For the term J_{31}, writing $\bar{u}_i = \sum_{j=1}^{n} W_{nj}(T_i)\, u_j$, we can further decompose it into

$$J_{31} = n^{-1} \sum_{i=1}^{n} ((a^\tau u_i)\,\hat{\varepsilon}_i)^2 + n^{-1} \sum_{i=1}^{n} (a^\tau \bar{u}_i)^2\, \hat{\varepsilon}_i^2 - 2n^{-1} \sum_{i=1}^{n} (a^\tau u_i)(a^\tau \bar{u}_i)\, \hat{\varepsilon}_i^2$$

$$+ n^{-1} \sum_{i=1}^{n} (a^\tau \tilde{\mu}_i)^2\, \hat{\varepsilon}_i^2 + 2n^{-1} \sum_{i=1}^{n} (a^\tau \tilde{\mu}_i)(a^\tau \tilde{U}_i)\, \hat{\varepsilon}_i^2$$

$$\equiv K_1 + K_2 + K_3 + K_4 + K_5.$$

Applying (A2), Lemma 1(i) and Lemma 2, we can get

$$K_2 = o_p(\log^{-2} n), \qquad K_3 = o_p(\log^{-1} n), \qquad K_4 = O_p(h^2), \qquad K_5 = O_p(h).$$

Therefore, $J_{31} = n^{-1} \sum_{i=1}^{n} ((a^\tau u_i)\,\hat{\varepsilon}_i)^2 + o_p(1)$. Hence,

$$J_3 = n^{-1} \sum_{i=1}^{n} ((a^\tau u_i)\,\hat{\varepsilon}_i)^2 + o_p(1). \tag{5.20}$$

From (5.17), (5.18), and (5.20), we get

$$a^\tau \left(n^{-1} \sum_{i=1}^{n} \tilde{U}_i^{\otimes 2} - n^{-1} \sum_{i=1}^{n} (u_i \hat{\varepsilon}_i)^{\otimes 2} \right) a = o_p(1).$$

Thus (iv) is proved and the proof of the lemma is completed.

LEMMA 5. *Let* $Z \xrightarrow{\mathscr{L}} N(0, I_p)$, *where* I_p *is the* $p \times p$ *indentity matrix. Let* U *be a* $p \times p$ *nonnegative definite matrix with eigenvalues* $l_1, ..., l_p$. *Then,*

$$Z' U Z \xrightarrow{\mathscr{L}} l_1 \chi_{1,1}^2 + \cdots + l_p \chi_{p,1}^2.$$

where $\chi_{i,1}^2$ *are as defined in Theorem* 2.1.

Proof. By the assumption, there exists an orthonormal matrix P such that $U = P'DP$, where $D = \mathrm{diag}(l_1, ..., l_p)$ is a diagonal matrix with diagonal elements $l_1, ..., l_p$. Let $\tilde{Z} = P'Z = (\tilde{Z}_1, ..., \tilde{Z}_p)'$. Clearly, $\tilde{Z} \xrightarrow{\mathscr{L}} N(0, I_p)$. Therefore,

$$Z' U Z = (PZ)'\, D(PZ) = l_1 \tilde{Z}_1^2 + \cdots + l_p \tilde{Z}_p^2,$$

where $\tilde{Z}_1^2, ..., \tilde{Z}_p^2$ are i.i.d. random variables with the common limiting chi-square distribution with one degree of freedom. This completes the proof.

Now we give the proofs of the main theorems.

Proof of Theorem 1. By (2.1), we have

$$\sqrt{n}(\hat{\beta}_n-\beta)=\left(n^{-1}\sum_{i=1}^{n}\tilde{X}_i^{\otimes 2}\right)^{-1}\left(n^{-1/2}\sum_{i=1}^{n}\tilde{X}_i(\tilde{Y}_{iG_n}-\tilde{X}_i^{\tau}\beta)\right),$$

then the first part of the theorem follows from (i) and (iii) of Lemma 4. From the proof of Lemma 4(iv), we can see that for any $a\in\mathbf{R}^p$

$$n^{-1}\sum_{i=1}^{n}(a^{\tau}\tilde{U}_i)^2-n^{-1}\sum_{i=1}^{n}(a^{\tau}\tilde{U}_{i0})^2=o_p(1). \tag{5.21}$$

Now we have that

$$\begin{aligned}
n^{-1}&\sum_{i=1}^{n}(a^{\tau}\tilde{X}_i)^2(\tilde{Y}_{iG_n}-\tilde{X}_i^{\tau}\hat{\beta}_n)^2-n^{-1}\sum_{i=1}^{n}(a^{\tau}\tilde{U}_i)^2\\
&=-\frac{2}{n}\sum_{i=1}^{n}(a^{\tau}\tilde{X}_i)^2(\tilde{Y}_{iG_n}-\tilde{X}_i^{\tau}\beta)(\tilde{X}_i^{\tau}(\hat{\beta}_n-\beta))\\
&\quad+\frac{1}{n}\sum_{i=1}^{n}(a^{\tau}\tilde{X}_i)^2(\tilde{X}_i^{\tau}(\hat{\beta}_n-\beta))^2\\
&\equiv M_1+M_2.
\end{aligned}$$

Note that (5.3) and (5.4) imply $\max_i\|\tilde{X}_i\|=o_p(n^{1/2})$, by Lemma 4(i), Lemma 3, and $\hat{\beta}_n-\beta=O_p(n^{-1/2})$,

$$|M_1|\leqslant 2\max_i\|\tilde{U}_i\|\,\|\hat{\beta}_n-\beta\|\cdot n^{-1}\sum_{i=1}^{n}(a^{\tau}\tilde{X}_i)^2=o_p(1),$$

$$|M_2|\leqslant \max_i\|\tilde{X}_i\|^2\,|\hat{\beta}_n-\beta\|^2\cdot n^{-1}\sum_{i=1}^{n}(a^{\tau}\tilde{X}_i)^2=o_p(1).$$

Hence,

$$n^{-1}\sum_{i=1}^{n}(a^{\tau}\tilde{X}_i)^2(\tilde{Y}_{iG_n}-\tilde{X}_i^{\tau}\hat{\beta}_n)^2-n^{-1}\sum_{i=1}^{n}(a^{\tau}\tilde{U}_i)^2\xrightarrow{p}0. \tag{5.22}$$

Similarly to the proof of Theorem 2.2 in Qin [28], we can get $n^{-1}\sum_{i=1}^{n}\tilde{U}_{i0}^{\otimes 2}-\Sigma=o_p(1)$, so by (5.21) and (5.22), $\hat{\Sigma}-\Sigma=o_p(1)$. From Lai *et al.* [22], $\hat{V}_1$ consistently estimates V_1, therefore $\hat{V}=V+o_p(1)$.

Hence $\hat{\Sigma}_u^{-1}\hat{V}\hat{\Sigma}_u^{-1}=\Sigma_u^{-1}V\Sigma_u^{-1}+o_p(1)$ and the second part of the theorem is proved. For the third part of the theorem, we have the decomposition

$$\begin{aligned}\hat{g}_n(t)-g(t)&=\sum_j W_{nj}(t)(Y_{jG_n}-Y_{jG})+\sum_j W_{nj}(t)(Y_{jG}-X_j^\tau\beta-g(T_j))\\&\quad-\sum_j W_{nj}(t)\,X_j^\tau(\hat{\beta}_n-\beta)+\Big(\sum_j W_{nj}(t)\,g(T_j)-g(t)\Big)\\&\equiv J_1+J_2+J_3+J_4.\end{aligned}\tag{5.23}$$

For the term J_1, by (5.14)

$$\begin{aligned}|J_1|&\leqslant\sup_{t\leqslant\tau_F}|G_n(t)-G(t)|\left(1+\sup_{t\leqslant\max_i Z_i}\frac{|G_n(t)-G(t)|}{|1-G_n(t)|}\right)\\&\quad\times\sum_{j=1}^n W_{nj}(t)\,|Y_{jG}|\,\bar{G}^{-1}(Z_j).\end{aligned}$$

Using (A2), we get

$$E\left(\sum_{j=1}^n W_{nj}(t)\,|Y_{jG}|\,\bar{G}^{-1}(Z_j)\right)=\sum_{j=1}^n E(W_{nj}(t)\,E(|Y_{jG}|\,\bar{G}^{-1}(Z_j)\,|\,T_j))<\infty.$$

So $\sum_{j=1}^n W_{nj}(t)\,|Y_{jG}|\,\bar{G}^{-1}(Z_j)=O_p(1)$. In the case of $\tau_F<\tau_G$, from Chen and Lo [6] and (A10), $\sup_{t\leqslant\tau_F}|G_n(t)-G(t)|=O_p(n^{-p})$. In the case of $\tau_F=\tau_G$, from Gu and Lai [14] and (A11), $\sup_{t\leqslant\tau_F}|G_n(t)-G(t)|=O_p((n\log\log n)^{-1/2})$. Hence from (5.6), we have, in these two cases, $J_1=O_p(n^{-p})$ and $J_1=O_p((n\log\log n)^{-1/2})$, respectively. Using (A1), (A2), and Lemma 1(iv), we get

$$\begin{aligned}EJ_2^2&=\sum_j E(W_{nj}^2(t)\,E((Y_{jG}-X_j^\tau\beta-g(T_j))^2\,|\,X_j,T_j))\\&\leqslant O(nE(W_{nj}^2(t)))=O((nh)^{-1}).\end{aligned}$$

Hence, $J_2=O_p((nh)^{-1/2})$. By (5.4) and $\hat{\beta}_n-\beta=O_p(n^{-1/2})$, we get $J_3=O_p(n^{-1/2})$. From Lemma 2, $J_4=O(h)$. Thus we complete the proof of the third part of the theorem.

Proof of Theorem 2. Applying Taylor's expansion to (3.4), we have

$$\tilde{l}(\beta)=2\sum_{i=1}^n(\tilde{\lambda}^\tau\tilde{U}_i-\tfrac{1}{2}(\tilde{\lambda}^\tau\tilde{U}_i)^2)+r_n\tag{5.24}$$

with $|r_n| \leqslant C \sum_{i=1}^{n} (\tilde{\lambda}^\tau \tilde{U}_i)^3$ in probability. By Lemma 4(iv) and the same arguments as those used in the proof of (2.14) of Owen [26], we can prove

$$\|\tilde{\lambda}\| = O_p(n^{-1/2}). \tag{5.25}$$

By Lemma 3 and (5.25), and noting that

$$n^{-1} \sum_{i=1}^{n} \|\tilde{U}_i\|^2 = O_p(1), \tag{5.26}$$

we have

$$|r_n| \leqslant C\, \|\tilde{\lambda}\|^3 \max_i \|\tilde{U}_i\| \sum_{i=1}^{n} \|\tilde{U}_i\|^2 = o_p(1). \tag{5.27}$$

Note that

$$\begin{aligned}
\sum_{i=1}^{n} \frac{\tilde{U}_i}{1+\tilde{\lambda}^\tau \tilde{U}_i} &= \sum_{i=1}^{n} \tilde{U}_i \left[1 - \tilde{\lambda}^\tau \tilde{U}_i + \frac{(\tilde{\lambda}^\tau \tilde{U}_i)^2}{1+\tilde{\lambda}^\tau \tilde{U}_i} \right] \\
&= \sum_{i=1}^{n} \tilde{U}_i - \left(\sum_{i=1}^{n} \tilde{U}_i^{\otimes 2} \right) \tilde{\lambda} + \sum_{i=1}^{n} \frac{\tilde{U}_i (\tilde{\lambda}^\tau \tilde{U}_i)^2}{1+\tilde{\lambda}^\tau \tilde{U}_i}.
\end{aligned} \tag{5.28}$$

From (3.3), (5.25), (5.26), (5.28), and Lemma 3, it follows that

$$\tilde{\lambda} = \left(\sum_{i=1}^{n} \tilde{U}_i^{\otimes 2} \right)^{-1} \sum_{i=1}^{n} \tilde{U}_i + o_p(n^{-1/2}). \tag{5.29}$$

Again by (3.3), we get that

$$0 = \sum_{i=1}^{n} \frac{\tilde{\lambda}^\tau \tilde{U}_i}{1+\tilde{\lambda}^\tau \tilde{U}_i} = \sum_{i=1}^{n} (\tilde{\lambda}^\tau \tilde{U}_i) - \sum_{i=1}^{n} (\tilde{\lambda}^\tau \tilde{U}_i)^2 + \sum_{i=1}^{n} \frac{(\tilde{\lambda}^\tau \tilde{U}_i)^3}{1+\tilde{\lambda}^\tau \tilde{U}_i}. \tag{5.30}$$

By Lemma 3, (5.25), and (5.26), it follows that

$$\sum_{i=1}^{n} \frac{(\tilde{\lambda}^\tau \tilde{U}_i)^3}{1+\tilde{\lambda}^\tau \tilde{U}_i} = o_p(1). \tag{5.31}$$

From (5.30) and (5.31), we get

$$\sum_{i=1}^{n} \tilde{\lambda}^\tau \tilde{U}_i = \sum_{i=1}^{n} (\tilde{\lambda}^\tau \tilde{U}_i)^2 + o_p(1). \tag{5.32}$$

By (5.24), (5.27), (5.29), and (5.32), we get

$$\begin{aligned}\tilde{l}(\beta) &= \sum_{i=1}^{n} \tilde{\lambda}^\tau \tilde{U}_i \tilde{U}_i^\tau \tilde{\lambda} + o_p(1)\\ &= \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{U}_i\right)^\tau \left(\frac{1}{n} \sum_{i=1}^{n} \tilde{U}_i^{\otimes 2}\right)^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{U}_i\right) + o_p(1)\\ &= \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{U}_i\right)^\tau \Sigma^{-1} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \tilde{U}_i\right) + o_p(1)\\ &= \left(V^{-1/2} n^{-1/2} \sum_{i=1}^{n} \tilde{U}_i\right)^\tau (V^{1/2}\Sigma^{-1}V^{1/2}) \left(V^{-1/2} n^{-1/2} \sum_{i=1}^{n} \tilde{U}_i\right) + o_p(1).\end{aligned} \tag{5.33}$$

By Lemma 4(iii), we have $V^{-1/2}(n^{-1/2} \sum_{i=1}^{n} \tilde{U}_i) \xrightarrow{\mathscr{L}} N(0, I_p)$. Also note that $V^{1/2}\Sigma^{-1}V^{1/2}$ and $\Sigma^{-1}V$ have the same eigenvalues. Then Theorem 2 follows from Lemma 5 straightaway.

Proof of Corollary 3. From Lemma 4(iii), (5.33), and (5.22), it follows that

$$\hat{l}(\beta) = \hat{\Sigma} \cdot \left(\frac{1}{n} \sum_{i=1}^{n} \tilde{U}_i^2\right)^{-1} \left(\frac{\sum_{i=1}^{n} \tilde{U}_i}{\sqrt{n\hat{V}}}\right)^2 + o_p(1) \xrightarrow{\mathscr{L}} \chi_1^2.$$

The proof of Corollary 3 is thus complete.

REFERENCES

1. G. Adimari, Empirical likelihood type confidence intervals under random censorship, *Ann. Inst. Statist. Math.* **49** (1997), 447–466.
2. J. Buckley and I. James, Linear regression with censored data, *Biometrika*, **66** (1979), 429–436.
3. H. Chen, Convergent rates for parametric components in a partly linear model, *Ann. Statist.* **16** (1988), 136–146.
4. H. Chen and H. J. Shiau, A two-stage spline smoothing method for partially linear model, *J. Statist. Plan. Inference* **27** (1991), 187–201.
5. H. Chen and J. H. Shiau, Data-driven efficient estimators for a partially linear model, *Ann. Statist.* **22** (1994), 211–237.
6. K. Chen and S.-H. Lo, On the rate of uniform convergence of the product-limit estimators: Strong and weak laws, *Ann. Statist.* **25** (1997), 1050–1087.
7. S. X. Chen, On the accuracy of empirical likelihood confidence regions for linear regression model, *Ann. Inst. Statist. Math.* **45** (1993), 621–637.
8. S. X. Chen, Empirical likelihood confidence intervals for linear regression coefficients, *J. Multivariate Anal.* **49** (1994), 24–40.
9. S. X. Chen and P. Hall, Smoothed empirical likelihood confidence intervals for quantiles, *Ann. Statist.* **21** (1993), 1166–1181.
10. D. M. Dabrowska, Smoothed Cox regression, *Ann. Statist.* **25** (1997), 1510–1540.

11. R. Engle, C. Granger, J. Rice, and A. Weiss, Nonparametric estimates of the relation between weather and electricity sales, *J. Amer. Statist. Assoc.* **81** (1986), 310–320.
12. J. Fan and I. Gijbels, Censored regression: Local linear approximations and their applications, *J. Amer. Statist. Assoc.* **89** (1994), 560–571.
13. R. Gill, "Censoring and Stochastic Integrals," Math. Centre Tract, Vol. 124., Math. Centrum, Amsterdam, 1980.
14. M. Gu and T. L. Lai, Functional laws of the iterated logarithm for the product-limit estimator of a distribution function under random censorship or truncation, *Ann. Probab.* **18** (1990), 160–189.
15. S. A. Hamilton and Y. K. Truong, Local linear estimation in partly linear models, *J. Multivariate Anal.* **60** (1997), 1–19.
16. N. Heckman, Spline smoothing in partly linear models, *J. Roy. Statist. Soc. B* **48** (1986), 244–248.
17. S. Y. Hong and P. Cheng, Bootstrap approximation for semiparametric regression models, *Sci. China Ser. A* **23** (1993), 239–251.
18. S. Y. Hong and P. Cheng, The convergence rate of estimation for parameter in a semiparametric model, *Chinese J. Appl. Probab. Statist.* **10** (1994), 62–71.
19. Y. Kitamura, Empirical likelihood methods with weakly dependent processes, *Ann. Statist.* **25** (1997), 2084–2102.
20. E. D. Kolaczyk, Empirical likelihood for generalized linear models, *Statistica Sinica* **4** (1994), 199–218.
21. H. Koul, V. Susarla, and J. Van Ryzin, Regression analysis with randomly right-censored data, *Ann. Statist.* **9** (1981), 1276–1288.
22. T. L. Lai, Z. Ying, and Z. K. Zheng, Asymptotic normality of a class of adaptive statistics with applications to synthetic data methods for censored regression, *J. Multivariate Anal.* **52** (1995), 259–279.
23. S. Leurgans, Linear models, random censoring and synthetic data, *Biometrika* **74** (1987), 301–309.
24. E. Mammen and Van de Geer, Penalized quasi-likelihood estimation in partial linear models, *Ann. Statist.* **25** (1997), 1014–1035.
25. A. Owen, Empirical likelihood ratio confidence intervals for single functional, *Biometrika* **75** (1988), 237–249.
26. A. Owen, Empirical likelihood ratio confidence regions, *Ann. Statist.* **18** (1990), 90–120.
27. A. Owen, Empirical likelihood for linear models, *Ann. Statist.* **19** (1991), 1725–1747.
28. G. S. Qin, Kernel smoothing in partial linear model with censored data, *Chinese Ann. Math.* **16A**, No. 4 (1995), 441–453.
29. J. Qin and J. F. Lawless, Empirical likelihood and general estimating equations, *Ann. Statist.* **22** (1994), 300–325.
30. J. Qin and A. Wong, Empirical likelihood in a semiparametric models, *Scand. J. Statist.* **23** (1996), 209–219.
31. P. Speckman, Kernel smoothing in partial linear models, *J. Roy. Statist. Soc. B* **50** (1988), 413–436.
32. C. Srinivasan and M. Zhou, Linear regression with censoring, *J. Multivariate Anal.* **49** (1991), 179–201.
33. Z. K. Zheng, "Regression Analysis with Censored Data," Ph.D. dissertation, Columbia University, 1984.
34. Z. K. Zheng, A class of estimator for the parameter in linear regression with censored data, *Acta Math. Appl. Sinica* **3** (1987), 231–241.