Statistica Sinica **15**(2005), 767-779

RANK ESTIMATION IN PARTIAL LINEAR MODEL WITH CENSORED DATA

Kani Chen, Jia Shen and Zhiliang Ying

Hong Kong University of Science and Technology, Fudan University and Columbia University

Abstract: We propose a simple estimation method for regression parameter in a partial linear model when the response variable is subject to random right censorship. It is based on suitably stratifying a Gehan-type extension of the Wilcoxon-Mann-Whitney estimating function. The stratification is rational, flexible and natural. The resulting estimate is shown to be consistent and asymptotically normal, even with the size of each stratum being as small as 2. In some special situations, the estimate is asymptotically as accurate as the analogous estimate with the function of the nonparametric component being completely known, implying that the stratification poses little loss of information. Inference is easily obtained through a resampling scheme which is valid with small or moderate sizes of strata. Both the parameter estimation and the resampling scheme can be carried out by linear programming and are easy to implement numerically. Extensive simulations are carried out and the results show strong support of the theory.

Key words and phrases: Accelerated failure time model, asymptotic normality, consistency, efficiency, linear programming, resampling.

1. Introduction

Partial linear models have proved useful, especially when the dependence of the response on one of the covariates is not certain and is not of main interest. For example, in many clinical trials and biomedical studies, the main interest is to identify the effect of a treatment in the presence of a confounding factor such as age. In these studies, a partial linear model is often employed and the confounding factor is treated as the nonparametric component. By so doing, the confounding effect, which is less certain and of less interest, is included properly in the model and the estimation accuracy of the treatment effect interpreted through the key slope parameter can be maintained. We propose the use of ranks to estimate the slope parameter in a partial linear model with censored data. Our method is rigorous, yet computationally simple.

The partial linear model is a simple and direct generalization of the linear regression model. With a log-transformation of the response variable, partial linear models give rise to a more general form of accelerated failure time models. In survival analysis, accelerated failure time linear regression models are important alternatives to the Cox proportional hazards model (Cox (1972)), and sometimes more appealing because of more direct interpretation (Reid (1994)). For linear regression models with censored data, the least squares estimation (LSE) and rank estimation methods were proposed and analyzed in Buckley and James (1979), Ritov (1990), Tsiatis (1990), Lai and Ying (1991a, b), Ying (1993) and Jin, Lin, Wei and Ying (2003), among many others. In particular, Jin, Lin, Wei and Ying (2003) used linear programming and a resampling scheme (see also Parzen, Wei and Ying (1994)) to overcome the computational complexity of the rank estimation method. However, it is unclear whether and how these computational advantages can be carried over to the partial linear model with censored data. With the presence of censorship, it is not clear how to eliminate the infinite dimensional nuisance parameter to obtain the estimate of the slope parameter. In particular, the conventional least squares method for the partial linear model without censorship, developed in Chen (1988) and Speckman (1988) for example becomes increasingly difficult to carry over to censored data. In addition, one of main problems with linear or partial linear models for censored data is variance estimation. Wang (1996), Wang and Zheng (1997), Liang and Zhou (1998) and Wang and Li (2002) tried to tackle the problem using synthetic data, however some of their estimation and inference procedures are also rather complex. Moreover, the method of synthetic data relies on the assumption of the independence of the censoring variable and the covariates, which is often a too restrictive assumption; see also Koul, Susarla and van Ryzin (1981). In summary, the analysis of censored data with partial linear model is quite challenging both theoretically and numerically.

In this paper we propose a rank estimation procedure, a direct generalization of Wilcoxon-Mann-Whitney estimation, to partial linear model with censored data. A key step is to eliminate the nuisance parameter by a proper stratification. Using counting process martingale theory, we prove that the resulting slope estimate is consistent and asymptotically normal. The two computational problems commonly associated with rank estimation, namely computing the slope estimate and obtaining its inference, are solved by linear programming and a resampling scheme and are easy to implement numerically; see Remark 1. One of the major advantages of this estimation method is the flexibility of the stratification involved. There is essentially no restriction on the stratification−consistency, asymptotic normality and inference via a resampling scheme are all valid as long as the size of each stratum is greater than 1 and is of order $o(n^{1/2})$; see Remarks

1 and 3. Moreover, the stratification does not depend on whether the nonparametric component is discrete/categorical or continuous. In certain situations the estimate can be as accurate as its linear model analogue when the function $h(\cdot)$ as the nuisance parameter of infinite dimension is completely known. This implies that the proposed estimation method via stratification indeed successfully eliminates the nuisance parameter $h(\cdot)$ with little loss of estimation accuracy; see Remark 1.

The rank estimate is proposed and proved to be consistent and asymptotic normal in Section 2. Section 3 contains simulation results in support of the theory and an application of it.

2. The Estimate And Inference

Let $(\tilde{Y}, X, W, \epsilon)$ be random variables satisfying the partial linear model

$$\tilde{Y} = \beta' X + h(W) + \epsilon, \tag{1}$$

where $\tilde{Y}$ is the response, X and W are covariates of p and 1 dimensions respectively, ϵ is the error term which is independent of (X, W), β is the slope parameter of p dimensions, and $h(\cdot)$ is a function which can be viewed as an infinite dimensional nuisance parameter. Let C be a censoring variable which is conditionally independent of $\tilde{Y}$ given (X, W). Let $(\tilde{Y}_i, X_i, W_i, \epsilon_i, C_i), 1 \leq i \leq n$, be n independent and identically distributed copies of $(\tilde{Y}, X, W, \epsilon, C)$. With the presence of censorship, we observe $(Y_i, X_i, W_i, \delta_i), 1 \leq i \leq n$, where $\delta_i = I(\tilde{Y}_i \leq C_i)$ and $Y_i = \min(\tilde{Y}_i, C_i)$.

The main feature of a partial linear regression model is the presence of the infinite dimensional nuisance parameter $h(\cdot)$. To eliminate it, we propose the following method. Stratify the observations into K_n groups, $J_1, \ldots, J_{K_n}$, according to the values of the nonparametric component W such that, for $1 \leq k0$. Since M is chosen to be large, the minimization of $L(\beta)$ is equivalent to the minimization of

$$\sum_{k=1}^{K_n} \sum_{i,j \in J_k} |u_{ij} - \beta' v_{ij}| + \Big| M + \beta' \sum_{k=1}^{K_n} \sum_{i,j \in J_k} v_{ij} \Big|.$$

This can be implemented easily with statistical software. For example, in Splus or R, use l1fit or rq.fit. In addition, the minimization of $L(\beta)$ can also be carried out in linear programming: minimize the linear function $\sum_{k=1}^{K_n} \sum_{i,j \in J_k} w_{ij}$ subject to the linear constraints $w_{ij} \geq 0$ and $w_{ij} \geq -(u_{ij} - \beta' v_{ij})$.

For clarity of presentation, let β_0 be the true value of β. Denote by f, F and Λ the density, distribution and cumulative hazard functions of ϵ. Let λ and $\dot{\lambda}$ be the first and second derivative functions of Λ. To show the consistency and asymptotic normality of $\hat{\beta}$, we need the following regularity conditions.

C1. The covariates X and W have bounded supports, and the support of C is bounded above.

C2. On the support of W, $h(\cdot)$ satisfies the Lipschitz condition $|h(t) - h(s)| \leq D|t - s|$ for some constant $D > 0$.

C3. As $n \to \infty$, $\max\{|W_k - W_l| : k, l \in J_i, 1 \leq i \leq K_n\} = o(n^{-1/2})$. For all $n \geq 1$, $\max\{|J_i|/a_n : 1 \leq i \leq K_n\}$ is bounded, where $|J_i|$ is the size of the set J_i and $a_n = n/K_n$.

C4. As $n \to \infty$, $a_n \to \infty$ and $G_n(w)/a_n \to g(w)$ as $n \to \infty$, where $g(\cdot)$ is a deterministic function and $G_n(\cdot)$ is a step function such that $G_n(w) = |J_k|$ if $w \in [\min(W_i, i \in J_k), \min(W_i, i \in J_{k+1}))$.

C5. The matrices Σ_1 and Σ_2 defined in (A.9) and (A.10) are finite and nondegenerate, and $\int_{-\infty}^{c} |\dot{\lambda}(t)|dt$ is finite for any constant c.

Throughout the paper, we show no limits for integration over the whole line, and $b^{\otimes 2} = bb'$ for any vector b.

Proposition 1. *Assume* C1−C3 *and* C5 *hold. Then, $\hat{\beta}$ is consistent and asymptotically normal. If in addition* C4 *holds,*

$$n^{1/2}(\hat{\beta} - \beta_0) \to N(0, \Sigma_2^{-1}\Sigma_1\Sigma_2^{-1}), \tag{3}$$

where Σ_1 and Σ_2 are defined in (A.9) *and* (A.10) ıin the Appendix.

Remark 1. Conditions C1, C2 and C5 are simply regularity conditions while Conditions C3 and C4 are related to the stratification rule. Stratification is a commonly used method in biostatistics, especially in studying the effect of certain risk factors (e.g., X) when a confounding factor (e.g., W) is also present. It is seen from C3 and C4 that the proposed stratification rule is quite flexible. In fact, C4 is only for obtaining an analytic expression for Σ_1 and Σ_2, and is not necessary for consistency, asymptotic normality or inference via resampling; see Remark 2 for more discussion. Moreover, the stratification also allows for discrete/categorical or continuous W. When W is categorical, stratification may reduce to the conventional one: each stratum is the group of subjects with the same value of W. When W is continuous, the restriction that the nonparametric components within each stratum do not vary too much (of the order $o(n^{-1/2})$), is still mild. This restriction ensures that the bias arising from stratification is negligible. The rationale for the proposed stratification can be understood in the following two scenarios. The first is when $h(\cdot)$ is assumed completely known. Then β can be estimated by the $\hat{\beta}^{\circ}$ which minimizes

$$L^{\circ}(\beta) = \sum_{i,j=1}^{n} \delta_i\{Y_i^{\circ} - Y_j^{\circ} - \beta'(X_i - X_j)\}^{-},$$

where $Y_i^\circ = Y_i - h(W_i)$. The second is when $h(\cdot)$ is assumed to be constant (known or unknown). Then the partial linear model reduces to a linear model and $Y_i^\circ - Y_j^\circ = Y_i - Y_j$. The estimate which minimizes $L^\circ(\cdot)$ is studied in Jin, Lin, Wei and Ying (2003). In general, it can be shown along the line of proof presented in the Appendix that $n^{1/2}(\hat\beta^\circ - \beta_0) \to N(0, \Sigma_2^{\circ -1}\Sigma_1^\circ\Sigma_2^{\circ -1})$, where Σ_1° and Σ_2° are similarly defined as Σ_1 and Σ_2 in (A.9) and (A.10), except with $g(W)$ replaced by 1 and all the conditional expections/probabilities replaced by those without conditioning on W. Suppose now $g(W) = 1$, i.e., each stratum has the same size, and W is independent of $\tilde C$ and X. Then $\Sigma_1 = \Sigma_1^\circ$, $\Sigma_2 = \Sigma_2^\circ$ and $\hat\beta$ and $\hat\beta^\circ$ have the same asymptotic distribution. In particular, in the absence of censorship, the independence of W and X and equal size of stratum imply the equal asymptotic accuracy of $\hat\beta$ and $\hat\beta^\circ$. When the partial linear model reduces to a linear model, the equal size of strata and independence of W and $(\tilde C, X)$ imply that the proposed estimate is as accurate as the estimate studied in Jin, Lin, Wei and Ying (2003).

Remark 2. The synthetic estimates proposed and analyzed in Wang (1996), Wang and Zheng (1997) and Wang and Li (2002) are motivated quite differently from the estimates in this paper. A key step in constructing synthetic estimates is to estimate the distribution of the censoring variable by the Kaplan-Meier estimate and thereby to synthesize the response variable. As with other synthetic estimates in the literature, such as the Koul-Susarla-van Ryzin estimate, validity requires the independence of the censoring variable and the covariates, which is often too restrictive. For example, it excludes competing risk models and cannot be naturally extended to dependent censoring. These can be accomplished with the assumption of conditional independence of the censoring variable and the response variable given the covariates. The proposed estimate does not rely on any structural assumption. It is properly motivated by the classical Wilcoxon-Mann-Whitney estimates and has a valid and easy inference procedure described below.

Remark 3. Assume C1$-$C3 and C5 hold. Suppose each stratum contains more than one subject but the size of each stratum may not converge to infinity. Here C4 is not imposed, and the size of each stratum may be as small as 2. The consistency and asymptotic normality of $\hat\beta$, as expressed in (A.8), still hold even though $\Sigma_{n,2}^{-1}\Sigma_{n,1}\Sigma_{n,2}^{-1}$ as the approximate variance of $\hat\beta$ may not have a limit. Most importantly, the following resampling scheme is also valid without C4.

Although the expressions of $\Sigma_{n,1}$, $\Sigma_{n,2}$, Σ_1 and Σ_2 regarding the asymptotic variance of $\hat\beta$ are available in closed forms in (A.4), (A.5), (A.9) and (A.10), a direct estimation of the variance of $\hat\beta$ via the plug-in rule is in fact quite difficult since these expressions contain the derivative of the density of ϵ, which cannot be accurately estimated. For ease of implementation, we propose to use a resampling scheme that is computationally straightforward and does not involve estimating any density or its derivative. Define

$$L^*(\beta) = \sum_{k=1}^{K_n} \sum_{i,j \in J_k} \delta_i \{Y_i - Y_j - \beta'(X_i - X_j)\}^- \xi_i,$$

where $\xi_i, i \geq 1$, are independent and identically distributed positive random variables, with $E(\xi_i) = var(\xi_i) = 1$, that are independent of $\{(Y_i, X_i, W_i, \delta_i), i \geq 1\}$. Denote the minimizor of $L^*(\beta)$ by $\hat{\beta}^*$.

Proposition 2. *Suppose conditions* C1−C3 *and* C5 *hold. Then, with probability* 1 *as* $n \to \infty$,

$$\sup_{t \in R^p} |F_n(t) - F_n^*(t)| \to 0, \tag{4}$$

where $F_n(\cdot)$ *is the distribution of* $n^{1/2}(\hat{\beta} - \beta_0)$ *and* $F_n^*(\cdot)$ *is the conditional distribution of* $n^{1/2}(\hat{\beta}^* - \hat{\beta})$ *given the data* $(Y_i, X_i, W_i, \delta_i)$, $1 \leq i \leq n$. *If in addition* C4 *holds, then with probability* 1, F_n^* *has limiting distribution* $N(0, \Sigma_2^{-1}\Sigma_1\Sigma_2^{-1})$, *as does* F_n.

Since (4) holds without C4, it implies that the resampling procedure for inference is valid without requiring the size of each stratum to be large. The above proposition suggests, for example, that a confidence interval for β can be constructed as follows. Let $\hat{\sigma}$ be the conditional variance of $\hat{\beta}^*$ given the data $(Y_i, X_i, W_i, \delta_i), 1 \leq i \leq n$. This can be computed straightforwardly through resampling. For any given $0 < \alpha < 1$, one can use $(\hat{\beta} - z(\alpha/2)\hat{\sigma}, \hat{\beta} + z(\alpha/2)\hat{\sigma})$ as the confidence interval for β at confidence level $1 - \alpha$. Here $z(\alpha/2)$ is the $1 - \alpha/2$ percentile of the standard normal distribution. As the minimization of $L^*(\beta)$ is as simple as the minimization of $L(\beta)$, the inference is easily obtained.

3. Numerical Studies

Extensive numerical studies were conducted. The following simulation examples were done with the software R, a key part was running L_1 regression using the *quantreg* package. The sample size n was 100; ϵ_i followed the standard normal distribution $N(0,1)$; X_i followed the uniform distribution $U(0,5)$ and W_i was chosen to be independent of or dependent on X_i specified below. We took β to be 1 or 2 and $h(\cdot)$ to be constant, linear or quadratic. The censoring variable C_i was generated from $C = \beta' X + h(W) + \varepsilon^*$, where ε^* follows the uniform distribution $U(0,r)$ independent of the rest of the variables. Here r was chosen such that the proportion of censoring was approximately 80%; see, e.g., Jin, Lin, Wei and Ying (2003). We chose equal size of strata, denoted by m, in each simulation, and m was 5, 10 or 20. The results presented in the following are based on 1,000 simulations. The size of each resampling was 500. For the resampling, ξ_i followed the exponential distribution with $E(\xi_i) = \text{var}(\xi_i) = 1$. Table 1 contains simulation results for the independent case, i.e., W_i is $U(0,1)$ and independent of X_i. Table 2 contains simulation results for the dependent case in which $W_i = 0.1(X_i + \eta_i)$, where η_i is $U(0,5)$ and independent of the rest of the variables.

In the following tables, we present the average of the estimates $\hat{\beta}$, the empirical standard error ("EMPSE"), the average of the estimated standard errors ("ESTSE") and coverage probabilities ("CP") of 95% confidence intervals based on the resampling. For the purpose of comparison, we also present results based on the analogous estimation method with known $h(\cdot)$, as discussed in Remark 1.

Table 1. Simulation results with W independent of X.

			Bias	EMPSE	ESTSE	CP
$\beta_0 = 1$	$m = 2$	$h(w) = 0.5$	0.0079	0.133	0.137	94.4
		$h(w) = w$	-0.0085	0.131	0.136	93.0
		$h(w) = w^2$	-0.0050	0.133	0.135	94.3
	$m = 5$	$h(w) = 0.5$	-0.0026	0.091	0.091	93.6
		$h(w) = w$	-0.0074	0.090	0.092	93.7
		$h(w) = w^2$	-0.0008	0.087	0.091	94.6
	$m = 10$	$h(w) = 0.5$	0.0024	0.084	0.081	93.3
		$h(w) = w$	-0.0041	0.082	0.082	93.9
		$h(w) = w^2$	-0.0022	0.086	0.081	93.3
	$m = 20$	$h(w) = 0.5$	0.0017	0.078	0.077	93.5
		$h(w) = w$	-0.0005	0.077	0.077	93.8
		$h(w) = w^2$	-0.0020	0.079	0.078	94.3
		$h(\cdot)$ is known	-0.0033	0.077	0.074	93.0
$\beta_0 = 2$	$m = 2$	$h(w) = 0.5$	0.0006	0.129	0.136	94.3
		$h(w) = w$	0.0020	0.128	0.136	94.7
		$h(w) = w^2$	0.0008	0.126	0.135	94.5
	$m = 5$	$h(w) = 0.5$	0.0006	0.090	0.091	94.3
		$h(w) = w$	0.0011	0.091	0.092	94.3
		$h(w) = w^2$	0.0049	0.089	0.090	93.5
	$m = 10$	$h(w) = 0.5$	-0.0037	0.080	0.081	93.5
		$h(w) = w$	0.0043	0.080	0.080	95.0
		$h(w) = w^2$	0.0012	0.085	0.081	93.3
	$m = 20$	$h(w) = 0.5$	-0.0009	0.074	0.077	95.5
		$h(w) = w$	-0.0008	0.077	0.078	94.7
		$h(w) = w^2$	-0.0008	0.077	0.077	93.3
		$h(\cdot)$ is known	-0.0010	0.074	0.074	94.9

Table 2. Simulation results with W dependent on X.

			Bias	EMPSE	ESTSE	CP
$\beta_0 = 1$	$m = 2$	$h(w) = 0.5$	0.0103	0.190	0.194	93.0
		$h(w) = w$	0.0029	0.181	0.193	93.7
		$h(w) = w^2$	0.0043	0.185	0.193	93.8
	$m = 5$	$h(w) = 0.5$	-0.0021	0.127	0.128	94.0
		$h(w) = w$	0.0054	0.131	0.129	93.4
		$h(w) = w^2$	0.0037	0.129	0.129	94.1
	$m = 10$	$h(w) = 0.5$	0.0049	0.115	0.114	94.7
		$h(w) = w$	0.0049	0.110	0.113	94.8
		$h(w) = w^2$	0.0077	0.115	0.114	93.3
	m=20	$h(w) = 0.5$	0.0006	0.106	0.105	93.7
		$h(w) = w$	0.0147	0.109	0.105	93.4
		$h(w) = w^2$	0.0081	0.105	0.104	93.4
		$h(\cdot)$ is known	-0.0046	0.076	0.074	93.7
$\beta_0 = 2$	$m = 2$	$h(w) = 0.5$	-0.0018	0.186	0.193	93.7
		$h(w) = w$	-0.0010	0.187	0.193	93.5
		$h(w) = w^2$	0.0020	0.176	0.189	94.2
	$m = 5$	$h(w) = 0.5$	-0.0086	0.129	0.127	93.0
		$h(w) = w$	0.0015	0.129	0.128	92.5
		$h(w) = w^2$	-0.0033	0.133	0.129	93.6
	$m = 10$	$h(w) = 0.5$	0.0055	0.114	0.113	93.7
		$h(w) = w$	0.0048	0.110	0.113	94.8
		$h(w) = w^2$	0.0032	0.115	0.112	92.9
	$m = 20$	$h(w) = 0.5$	-0.0006	0.104	0.105	95.0
		$h(w) = w$	0.0130	0.101	0.104	94.6
		$h(w) = w^2$	0.0138	0.105	0.105	94.5
		$h(\cdot)$ is known	-0.0027	0.075	0.074	94.0

It is seen from the tables that the proposed slope estimates $\hat{\beta}$ are quite accurate. The empirical standard error and the estimated standard error are in general close to each other, implying that the resampling method for inference is appropriate. Moreover, the coverage probabilities are close to the nominal level 95%. In Table 1, the standard errors of the estimates, even for $m = 5$, are quite close to those of the estimates with known $h(\cdot)$. This provides empirical evidence to the claim that in some cases the proposed estimate can be as accurate as the one with known $h(\cdot)$; see Remark 1 and Jin, Lin, Wei and Ying (2003) for more discussion.

We apply the estimation procedure to a study on multiple myeloma; see Krall, Uthoff and Harley (1975). There were 65 patients in the study. Among them, there were 48 failures/deaths and 17 survivals/censoring. The partial linear model we consider is

$$\tilde{Y} = \beta \times \log(\text{BUN}) + h(\text{age}) + \epsilon,$$

where $\tilde{Y}$ is lifetime and log(BUN) is the logarithm of blood urea nitrogen. Age might be a confounding factor and it is treated as the nonparametric component in the model. For the stratification we chose $m = 2$ or 5 as the (equal) size of strata. For $m = 2$, $\hat{\beta} = -1.955$, its estimated standard error obtained through resampling (with 500 replications) is $\hat{\sigma} = 0.807$, and the 95% confidence interval for β is $(-3.538, -0.372)$. For $m = 5$, $\hat{\beta} = -1.863$, $\hat{\sigma} = 0.396$ and the 95% confidence interval for β is $(-2.641, -1.086)$. The two slope estimates are both negative and the two confidence intervals are both in the negative part of the real line. Blood urea nitrogen is negatively related to lifetime.

Acknowledgement

This research was supported by Hong Kong RGC Earmark grant HKUST 6134/01P, the National Institutes of Health and the National Science Foundation.

Appendix

1. Proof of Proposition 1. More notation is needed. Let $\tilde{C}_i = C_i - \beta_0' X_i - h(W_i)$. Let $N_i(\beta; t) = \delta_i I(Y_i - \beta' X_i \le t)$, $M_i(\beta; t) = N_i(\beta; t) - \int_{-\infty}^{t} I(Y_i - \beta' X_i \ge s) d\Lambda\{s - h(W_i)\}$. Then $M_i(\beta_0; t)$ is a counting process martingale. Set $R_k(\beta; t) = \sum_{i \in J_k} I(Y_i - \beta' X_i \ge t)$ and $\bar{X}_k(\beta; t) = \sum_{i \in J_k} X_i I(Y_i - \beta' X_i \ge t) / R_k(\beta; t)$.

Notice that, for all large n, $L(\beta)$ is a strictly convex function of β and therefore the minimizor $\hat{\beta}$ is unique. It follows from the Law of Large Numbers and the empirical approximation that, with probability 1, $|L(\beta) - E\{L(\beta)\}|/(na_n)$ converges to 0 uniformly over any bounded region of β. Moreover, it can be shown that β_0 is the unique minimizor of the strictly convex function $E\{L(\beta)\}/(na_n)$. It then follows that $\hat{\beta}$ converges to β_0 with probability 1. Let B_n be a sequence of balls centered at β_0 with radius $r_n \downarrow 0$ slowly enough. We have $P(\hat{\beta} \in B_n) \to 1$.

Consider β in a fixed neighborhood of β_0. Observe that $\sum_{i \in J_k} R_k(\beta; t)\{X_i - \bar{X}_k(\beta; t)\} I(Y_i - \beta' X_i \ge t) = 0$. Let $b_{k,1} = \min\{h(W_i); i \in J_k\}$ and $b_{k,2} = \max\{h(W_i); i \in J_k\}$. Then

$$\sum_{i \in J_k} \int R_k(\beta; t)\{X_i - \bar{X}_k(\beta; t)\} I(Y_i - \beta' X_i \ge t) d\Lambda(t + b_{k,1}) = 0.$$

Using C2−C3, one can verify that $\max\{|b_{k,2} - b_{k,1}| : 1 \le k \le K_n\} = \max\{|h(W_i) - h(W_j)| : i, j \in J_k; 1 \le k \le K_n\} = o(n^{-1/2})$. It then follows from C1−C3 and

C5 that

$$\begin{aligned}
&\sum_{k=1}^{K_n}\sum_{i\in J_k}\int R_k(\beta;t)\{X_i-\bar{X}_k(\beta;t)\}I(Y_i-\beta'X_i\geq t)d\Lambda\{t+h(W_i)\}\\
&=\sum_{k=1}^{K_n}\sum_{i\in J_k}\int R_k(\beta;t)\{X_i-\bar{X}_k(\beta;t)\}I(Y_i-\beta'X_i\geq t)[d\Lambda\{t+h(W_i)\}-d\Lambda(t+b_{k,1})]\\
&=O_P(a_n)\sum_{k=1}^{K_n}\sum_{i\in J_k}\int_{-\infty}^{c}|\lambda(t+h(W_i))-\lambda(t+b_{k,1})|dt\\
&=O_P(a_n^2)\sum_{k=1}^{K_n}\int_{-\infty}^{c}\int_{t+b_{k,1}}^{t+b_{k,2}}|\dot{\lambda}(s)|ds\,dt\\
&=O_P(a_n^2)K_n o_P(n^{-1/2})=o_P(n^{1/2}a_n), \qquad\qquad (A.1)
\end{aligned}$$

where c is a large but fixed constant. Formally differentiate $L(\beta)$ with respect to β and let $U(\beta)$ be the gradient. Then $\hat{\beta}$ is a zero crossing of $U(\beta)$. Write

$$\begin{aligned}
U(\beta)=\frac{\partial}{\partial\beta}L(\beta)&=\sum_{k=1}^{K_n}\sum_{i,j\in J_k}\delta_i(X_i-X_j)I(Y_i-\beta'X_i\leq Y_j-\beta'X_j)\\
&=\sum_{k=1}^{K_n}\sum_{i\in J_k}\int R_k(\beta;t)\{X_i-\bar{X}_k(\beta;t)\}dN_i(\beta;t). \qquad\qquad (A.2)
\end{aligned}$$

It follows from (A.1) and (A.2) that

$$U(\beta)=\sum_{k=1}^{K_n}\sum_{i\in J_k}\int R_k(\beta;t)\{X_i-\bar{X}_k(\beta;t)\}dM_i(\beta;t)+o_P(n^{1/2}a_n). \qquad (A.3)$$

Define

$$\Sigma_{n,1}=\frac{1}{n}\sum_{k=1}^{K_n}\sum_{i\in J_k}\int\{R_k(\beta_0;t)/a_n\}^2\{X_i-\bar{X}_k(\beta_0;t)\}^{\otimes 2}dN_i(\beta_0;t), \qquad (A.4)$$

$$\Sigma_{n,2}=\frac{1}{n}\sum_{k=1}^{K_n}\sum_{i\in J_k}\int\frac{R_k(\beta_0;t)}{a_n}\{X_i-\bar{X}_k(\beta_0;t)\}^{\otimes 2}I(Y_i-\beta_0'X_i\geq t)\dot{\lambda}\{t-h(W_i)\}dt. \qquad (A.5)$$

It follows from the Martingale Central Limit Theorem and (A.3) that

$$(n^{-1/2}/a_n)\Sigma_{n,1}^{-1/2}U(\beta_0)\to N(0,I_p), \qquad (A.6)$$

where I_p is the $p \times p$ identity matrix. By a change of variable in t, (A.3) can be written as

$$U(\beta) = \sum_{k=1}^{K_n} \sum_{i \in J_k} \int R_k(\beta; t - \beta' X_i)\{X_i - \bar{X}_k(\beta; t - \beta' X_i)\}$$
$$\times \Big[d\delta_i I(Y_i \leq t) - I(Y_i \geq t) d\Lambda\{t - \beta' X_i - h(W_i)\}\Big] + o_P(n^{1/2} a_n).$$

Then, for $\beta \in B_n$, using the above expression, one can write

$$\frac{1}{na_n}\{U(\beta) - U(\beta_0)\} = o_P(n^{-1/2})$$
$$+ \frac{1}{na_n} \sum_{k=1}^{K_n} \sum_{i \in J_k} \int R_k(\beta_0; t - \beta_0' X_i)\{X_i - \bar{X}_k(\beta_0; t - \beta_0' X_i)\} X_i'$$
$$(\beta - \beta_0) I(Y_i \geq t) \dot{\lambda}\{t - \beta_0' X_i - h(W_i)\} dt$$
$$= o_P(n^{-1/2}) + \frac{1}{n}\Big[\sum_{k=1}^{K_n} \sum_{i \in J_k} \int \frac{R_k(\beta_0; t)}{a_n}\{X_i - \bar{X}_k(\beta_0; t)\}^{\otimes 2}$$
$$\times I(Y_i - \beta_0' X_i \geq t) \dot{\lambda}\{t - h(W_i)\} dt\Big](\beta - \beta_0)$$
$$= \Sigma_{n,2}(\beta - \beta_0) + o_P(n^{-1/2}). \tag{A.7}$$

Combining (A.7) with (A.6), we have

$$n^{1/2} \Sigma_{n,1}^{-1/2} \Sigma_{n,2}(\hat{\beta} - \beta_0) \to N(0, I_p). \tag{A.8}$$

This proves the consistency and asymptotic normality of $\hat{\beta}$ under C1−C3 and C5. If, in addition C4 holds, it follows from the Law of Large Numbers that

$$\Sigma_{n,1} \to \int E\Big[\{g(W) P(\epsilon > t) P(\tilde{C} \geq t | W)\}^2 \{X - E(X | \tilde{C} \geq t; W)\}^{\otimes 2}$$
$$\times f(t) P(\tilde{C} \geq t | W)\Big] dt,$$
$$= \Sigma_1, \quad \text{say}, \tag{A.9}$$

$$\Sigma_{n,2} \to \int E\Big[\{g(W) P(\epsilon \geq t) P(\tilde{C} \geq t | W)\}\{X - E(X | \tilde{C} \geq t; W)\}^{\otimes 2}$$
$$\times \dot{\lambda}(t) P(\epsilon > t)(P(\tilde{C} \geq t | W)\Big] dt$$
$$= \int E\Big[g(W)\{X - E(X | \tilde{C} \geq t; W)\}^{\otimes 2}$$
$$\times [\dot{f}(t)\{1 - F(t)\} + f^2(t)]\{P(\tilde{C} \geq t | W)\}^2\Big] dt$$
$$= \Sigma_2, \quad \text{say}. \tag{A.10}$$

Hence (3) holds and the proof is complete.

2. Proof of Proposition 2. The proof follows that of Proposition 1. Observe that $L^*(\beta)$ is also a strictly convex function and the minimizor $\hat{\beta}^*$ is therefore unique for all large n. Let Ω be the σ-algebra of $(Y_i, X_i, W_i, \delta_i), i \geq 1$. The conditional mean of $L^*(\beta)$ given Ω is precisely $L(\beta)$. It follows that $|L^*(\beta) - L(\beta)|/(na_n)$ converges to 0 uniformly over any bounded region of β with probability 1. Hence $P(\hat{\beta}^* \in B_n|\Omega) \to 1$ with probability 1 by the strong consistency of $\hat{\beta}$.

Let $U^*(\cdot)$ be the gradient of $L^*(\cdot)$. Then,

$$U^*(\beta) = \frac{\partial}{\partial\beta} L^*(\beta) = \sum_{k=1}^{K_n} \sum_{i,j \in J_k} \delta_i \xi_i (X_i - X_j) I(Y_i - \beta' X_i \leq Y_j - \beta' X_j).$$

Recall that $\hat{\beta}^*$ and $\hat{\beta}$ are zero crossings of $U^*(\cdot)$ and $U(\cdot)$, respectively. Write

$$\begin{aligned} & U^*(\hat{\beta}^*) - U^*(\hat{\beta}) = -U^*(\hat{\beta}) = -U^*(\hat{\beta}) + U(\hat{\beta}) \\ = & -\sum_{k=1}^{K_n} \sum_{i \in J_k} \int R_k(\beta; t) \{X_i - \bar{X}_k(\beta; t)\} dN_i(\beta; t)(\xi_i - 1). \end{aligned}$$

Since $\xi_i, i \geq 1$, are independent, identically distributed with mean 1 and variance 1 and are independent of Ω, the Central Limit Theorem implies that, conditioning on Ω, the conditional distribution of $n^{-1/2}/a_n\{U^*(\hat{\beta}^*) - U^*(\hat{\beta})\}$ is approximately $N(0, \Sigma_{n,1})$. On the other hand, analogous to (A.7), it can be shown that, with probability 1,

$$\frac{1}{na_n}\{U^*(\hat{\beta}^*) - U^*(\hat{\beta})\} = \Sigma_{n,2}(\hat{\beta}^* - \hat{\beta}) + o_{P^*}(n^{-1/2}),$$

where P^* is the conditional probability given Ω. Consequently, the conditional distribution of $n^{1/2}\Sigma_{n,1}^{-1/2}\Sigma_{n,2}(\hat{\beta}^* - \hat{\beta})$ given Ω is approximately $N(0, I_p)$. The proof is complete on observing (A.8) and the arguments following (A.8).

References

Buckley, J. and James, I. (1979). Linear regression with censored data. *Biometrika* **66**, 429-436.

Chen, H. (1988). Convergence rates for parametric components in a partly linear model. *Ann. Statist.* **16**, 136-146.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *J. Roy. Statist. Soc. Ser. B* **34**, 187-220.

Jin, Z., Lin, D., Wei, L. J. and Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika* **90**, 341-353.

Koul, H., Susarla, V. and van Ryzin, J. (1981). Regression analysis with randomly right-censored data. *Ann. Statist.* **9**, 1276-1288.

Krall, J. M., Uthoff, V. A. and Harley, J. B. (1975). A step-up procedure for selecting variables associated with survival. *Biometrics* **31,** 49-57.

Lai, T. L. and Ying, Z. (1991a). Large sample theory of a modified Buckley-James estimator for regression analysis with censored data. *Ann. Statist.* **19**, 1370-1402.

Lai, T. L. and Ying, Z. (1991b). Rank regression methods for left-truncated and right-censored data. *Ann. Statist.* **19**, 531-556.

Liang, H. and Zhou, Y. (1998). Asymptotic normality in a semiparametric partial linear model with right-censored data. *Comm. Statist. Theory Method* **27**, 2895-2907.

Lin, D. Y., Wei, L. J. and Ying, Z. (1998). Accelerated failure time models for counting processes. *Biometrika* **85**, 605-618.

Parzen, M. I., Wei, L. J. and Ying, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341-350.

Reid, N. (1994). A conversation with Sir David Cox. *Statist. Sci.* **9**, 439-455.

Speckman, P. (1988). Kernel smoothing in partial linear models. *J. R. Statist. Soc. B* **50**, 413-436.

Tsiatis, A. A. (1990). Estimating regression parameters using linear rank tests for censored data. *Ann. Statist.* **18**, 354-372.

Wang, Q. (1996). Consistent estimators in random censorship semiparametric models. *Science in China* **39**, 163-176.

Wang, Q. and Zheng, Z. (1997). Asymptotic properties for the semiparametric regression model with randomly censored data. *Sci. China* **40**, 945-957.

Wang, Q. and Li, G. (2002). Empirical likelihood semiparametric regression analysis under random censorship. *J. Multivariate Anal.* **83**, 469-486.

Ying, Z. (1993). A large sample study of rank estimation for censored data. *Ann. Statist.* **21**, 76-99.

Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong.

E-mail: makchen@ust.hk

Department of Statistics, Fudan University, Shanghai, China.

E-mail: jiashen@fudan.edu.cn

Department of Statistics, Columbia University, NY, NY 10027, U.S.A.

E-mail: zying@stat.columbia.edu

(Received April 2003; accepted June 2004)