Statistica Sinica **9**(1999), 199-210

RANK ESTIMATING EQUATIONS FOR PARTIAL SPLINE MODELS WITH MONOTONICITY

Chu-chih Chen and Fushing Hsieh

Tamkang University and National Taiwan University

Abstract: For partial spline models with a monotone nonlinear component, a class of monotone estimating equations is proposed for estimating the slope parameters of the vector of covariables Z of the linear component, while adjusting for the corresponding ranks of the vector of covariables X of the nonlinear component. This approach avoids the technical complications due to the smoothing of estimators for the nonlinear component with monotonicity, as well as the *curse of dimensionality*. Also, computationally, our inferences do not involve the unknown error probability density function. As an R-estimator taking into account the rank correlation between Y and X, the asymptotic relative efficiency with respect to other estimators ignoring X is proportional to the Spearman correlation coefficient between them.

Key words and phrases: Asymptotic relative efficiency, censored regression, rank analysis of covariance, rank transformation, semiparametric regression model.

1. Introduction

In this paper we consider regression models consisting of linear as well as non-linear components. These so-called partial spline models, or partly linear models, can be found in economics and biometrics (Engle, Granger, Rice and Weiss (1986), Gray (1994), Sleeper and Harrington (1990)), and have been covered extensively in statistics literature (see, e.g., Chen (1988), (1995), Heckman (1986), Shiller (1984), Speckman (1988)). Most often the primary interest lies in making inferences of the linear coefficients with adjustments, rather than make the strong assumption of linear associations between the response and remaining covariable(s). For example, in a randomized clinical trial for a comparison of two treatments, the experimenter may be unsure of the effects of age on the response, but may want to estimate the treatment differences which are believed to be constant and independent of age (Heckman (1986)). In the same setting, the research interest is focused on the regression coefficient(s) of the linear term(s).

Consider a partial spline model of the form

$$Y_i = \gamma^T Z_i + m(X_i) + \epsilon_i, \qquad i = 1, \ldots, n, \tag{1}$$

where Z is a d-dimensional vector of covariables, $m(\cdot)$ is an unknown component-wise monotone (possibly nonlinear) function of a p-dimensional vector of covariables X, and ϵ is the residual with an unknown distribution function $F(\cdot)$.

Given model (1) and γ, the parameter of interest, the common approach is to estimate γ and the function $m(\cdot)$ or its finite approximation simultaneously, using the spline method with a penalty function for over-fitting (see, e.g., Wahba (1990), Chen (1988)). These procedures, however, seem complicated and unstable due to the sensitivity of the smoothing technique, as well as the *curse of dimensionality.* Further technical difficulty can be incurred with monotonicity assumed in each component of $m(X)$, such as the age effect in two-sample treatments.

In this paper, we note that rank transformations are invariant under monotone transformations and regress the ranks of $Y_i - \gamma^T Z_i$ on the component-wise ranks of X to form estimating equations for γ. By doing so we avoid the nonparametric estimation of $m(X)$. For expository purposes we first derive the estimating equation when Z and X are stochastically independent. It is noted that, when Z are group indicators and $\gamma = 0$, as a test statistic this is the rank analysis of covariance (Quade (1967)). The linear component part of the corresponding estimating equation reduces asymptotically to the statistics proposed by Sen (1968) for linear models, and Fygenson and Ritov (1994) for censored regression. We then make some adjustment when Z and X are dependent.

The paper is organized as follows. First, we apply the above approach to the two-sample case in Section 2, followed by a one-sample regression case in Section 3. For model (1) with right censored data we adopt Gehen's (1965) generalized Wilcoxon score for the response Y in Section 4. The results from previous sections are generalized to the case where Z and X are dependent in Section 5. Small simulation studies were conducted to compare with monotone estimating equations (Fygenson and Ritov (1994)) and other spline methods, and they are discussed in Section 6. Concluding remarks are made in the final section.

2. Treatment Effect in the Two-Sample Case

Consider a set of two-sample data consisting of vectors (Y_i, Z_i, X_i), where Y is the response and $X_i = (X_{i,1}, \ldots, X_{i,p})^T$ is the p-dimensional vector of the coveriables. Here $Z_i = 0$ labels the control group of size $n_0, Z_i = 1$ the treatment group of size n_1, $n = n_1 + n_0$. Model (1) imples that $F(Y - \gamma|Z = 1, X) = F(Y|Z = 0, X)$. And we are interested in estimating the treatment effect γ. By ignoring $m(X)$, or the vector of the covariables X, one finds the classical two-sample location model. In this setting the Hodges-Lemann estimator is a popular estimator for γ.

Let $R_i(\gamma) = 1/2 \sum_{j=1}^n \mathrm{sgn}[(Y_i - \gamma Z_i) - (Y_j - \gamma Z_j)]$ be the (mean adjusted) rank of $Y_i - \gamma Z_i$, and $C_{i,l} = \frac{1}{2} \sum_{j=1}^n \mathrm{sgn}(X_{i,l} - X_{j,l})$ be the rank of $X_{i,l}$ in the pooled sample. Then the resulting residual scores

$$D_i(\gamma) = R_i(\gamma) - \sum_{l=1}^{p} \lambda_l C_{i,l} \tag{2}$$

are interchangeable, where $C_i = (C_{i,1}, \ldots, C_i, p)^T$, $i = 1, \ldots, n$, and standard U-statistic theory applies. Heuristically, $\lambda = (\lambda_1, \ldots, \lambda_p)^T$ can be taken as the vector of the regression coefficients for R_i on C_i. A test criterion, in the spirit of the Wilcoxon rank-sum statistic, is the sum of the residual scores of the treated observations. For a given λ, the null hypothesis should be rejected for large values of the test statistic

$$Q(\gamma; \lambda) = \sum_{i=1}^{n} Z_i(R_i(\gamma) - \sum_{l} \lambda_l C_{i,l})/n_1 n_0. \tag{3}$$

The statistic $Q(\gamma; \lambda)$ is centered at zero under the null hypothesis $H_0 : \gamma = \gamma_0$. In addition, the larger the value of γ, the smaller the statistic Q, and vice versa. Therefore the test statistic can be used as a monotone estimating equation. The proposed statistic $Q(\gamma; \lambda)$ is not continuous in γ, so one may not be able to solve $Q(\gamma; \lambda) = 0$. The generalized solution will be that γ for which a slight perturbation of any of its components changes the sign of Q.

Given λ, let $\hat{\gamma}_\lambda^{(L)} = \sup\{\gamma : Q(\gamma; \lambda) > 0\}$, and $\hat{\gamma}_\lambda^{(U)} = \inf\{\gamma : Q(\gamma; \lambda < 0\}$; we take

$$\hat{\gamma}_\lambda = \frac{\hat{\gamma}_\lambda^{(U)} + \hat{\gamma}_\lambda^{(L)}}{2}. \tag{4}$$

Note that when $\lambda = 0$, $Q(\gamma; 0)$ reduces to the well-known Wilcoxon rank-sum statistic, and $\hat{\gamma}_0$ is the corresponding Hodges-Lemann estimator in the two-sample case (the median of $\{(Y_{1j} - Y_{0j'})\}$, where Y_{1j} and $Y_{0j'}$ are observations from the treatment group and the control group respectively, $j = 1, \ldots, n_1$ and $j' = 1, \ldots, n_0$, Randles and Wolfe (1991) pp. 213-23). By a two-sample U-statistic theorem, $n^{1/2} Q(\gamma_0; \lambda)$ is asymptotically distributed as $N(0, {\sigma_{Q,\lambda}}^2)$, where

$${\sigma_{Q,\lambda}}^2 = {\sigma_0}^2 - 2\lambda^T \eta + \lambda^T \Lambda \lambda,$$

$(n^2 - 1){\sigma_0}^2/12$ is the variance for $R_i(\gamma_0)$, $(n^2 - 1)\Lambda/12$ and $(n^2 - 1)\eta/12$ are the covariance matrix of $C_i = (C_{i,1}, \ldots, C_{i,p})^T$ and the covariance between C_i and $R_i(\gamma_0)$ respectively. The following inequality (5) gives the asymptotic distribution of $\hat{\gamma}_\lambda$ based on the estimating equation $Q(\gamma; \lambda)$. For any constant a,

$$\mathrm{pr}_\gamma\{Q(a; \lambda) < 0\} \le \mathrm{pr}_\gamma(\hat{\gamma}_\lambda < a) \le \mathrm{pr}_\gamma\{Q(a; \lambda) \le 0\}, \tag{5}$$

where $\mathrm{pr}_\gamma(\cdot)$ denotes the probability under γ. Let $\xi(\gamma) = E(Q(\gamma; \lambda))$. Then based on (5), we have that $n^{1/2}(\hat{\gamma}_\lambda - \gamma_0)$ is asymptotically normal with mean zero and variance $1/{K_{Q,\lambda}}^2$, where ${K_{Q,\lambda}}^2 = \xi'(\gamma_0)^2/{\sigma_{Q,\lambda}}^2$ is the efficacy of the test based on $Q(\gamma; \lambda)$. We provide the proof for (5) and the above result in the Appendix.

Since $Q(\gamma;\lambda)$ is monotone in γ, a 95% confidence interval for γ_0 can be constructed by finding the roots of $Q(\gamma;\lambda)$ such that $Q(\hat{\lambda}^L;\lambda) = -1.96\hat{\sigma}_{Q,\lambda}$ and $Q(\hat{\gamma}^U;\lambda) = 1.96\hat{\sigma}_{Q,\lambda}$ respectively. Further, the efficacy of $Q(\gamma;0)$ is $\xi'(\gamma_0)^2/\sigma_0^2$. Theorefore, taking the optimal weight $\tau = \Lambda^{-1}\eta$ of λ, we see that the asymptotic relative efficiency of $\hat{\gamma}_\tau$ with respect to $\hat{\gamma}_0$ is

$$\text{ARE}(\hat{\gamma}_\tau, \hat{\gamma}_0) = \frac{\text{eff}(Q(\hat{\gamma}_\tau;\tau))}{\text{eff}(Q(\hat{\gamma}_0;0))} = \frac{{\sigma_0}^2}{{\sigma_0}^2 - \eta^T\Lambda^{-1}\eta} = \frac{1}{1 - {R_S}^2}, \tag{6}$$

where R_s is the multiple Spearman correlation coefficient between Y and X (Quade (1967)). In summary, by adjusting for the ranks of X, the propsed estimator improves the efficiency over the Hodges-Lemann estimator by a factor of $1/(1 - {R_S}^2)$ when Z and X are independent.

3. Regression Coefficient for the Partial Spline Model with Monotonicity

We extend the results of the previous section to the one-sample case where Z is a d-dimensional vector of covariables. That is, we estimate the regression parameters γ of the partial spline model (1).

For a correctly specified coefficient γ_0, the adjusted ranks of $Y_i - \gamma_0^T Z_i$ will be independent of Z_i, $i = 1, \ldots, n$. Then, a reasonable estimate for γ_0 would be the solution of the following estimating equation:

$$Q_n(\gamma;\lambda) = n^{-5/2}\sum_i\sum_j(Z_i - Z_j)\{(R_j(\gamma) - R_i(\gamma)) - \lambda^T(C_j - C_i)\} = 0, \tag{7}$$

where $R_i(\gamma)$ is the rank of $Y_i - \gamma^T Z_i$, and C_i is the vector of the component-wise ranks of X_i, $i = 1, \ldots, n$.

As in the two-sample case, the statistic $Q_n(\gamma;\lambda)$ is not continuous in γ and a generalized solution is required. In view of Ritov (1987), $Q_n(\gamma;\lambda)$ is a monotone nondecreasing field, since for any γ, $\xi \in R^d$, $\xi^T Q_n(\gamma + v\xi;\lambda)$ is a monotone nondecreasing function of the real variable v.

Let $Q_n = (Q_{n,1}, \ldots, Q_{n,d})^T$. For the kth component γ_k of γ, the value of $Q_{n,k}(\gamma;\lambda)$ is invariant to the other components γ_l, $k \neq l$. Therefore we can take $\hat{\gamma}_{\lambda,k}$ to be the average of $\sup\{\gamma_k : Q_{n,k}(\gamma;\lambda) > 0\}$ and $\inf\{\gamma_k : Q_{n,k}(\gamma;\lambda) < 0\}$. Then $\hat{\gamma}_\lambda = (\hat{\gamma}_{\lambda,1}, \ldots, \hat{\gamma}_{\lambda,d})^T$ is a generalized solution of $Q(\gamma;\lambda) = 0$. Furthermore, $n^{-1/2}Q_n(\gamma;\lambda)$ is asymptotically equivalent to a U-statistic $U(\gamma;\lambda)$ of order 2 with a symmetric kernel

$$\begin{aligned} &h\{(Y_1, Z_1, X_1), (Y_2, Z_2, X_2)\} \\ &= (Z_1 - Z_2)[\text{sgn}\{Y_2 - Y_1 - \gamma^T(Z_2 - Z_1)\} - \sum_l \lambda_l \text{sgn}(X_{2,l} - X_{1,l})]. \end{aligned} \tag{8}$$

Let $A(\gamma) = \frac{1}{2}E[h\{(Y_1, Z_1, X_1), (Y_2, Z_2, X_2)\}]$ and let $V(\gamma; \lambda)$ be the variance for $Q_n(\gamma; \lambda)$. Our primary result is the following theorem (the proof is given in the Appendix).

Theorem 1. *Suppose $E(\|Z\|^2) < \infty$ and the density $f(\cdot)$ of ϵ has a finite first derivative. Then*

(i) *For any fixed γ, $Q_n(\gamma; \lambda)$ is asymptotically $N(A(\gamma), V(\gamma; \lambda))$, where $A(\gamma) : R^d \to R^d$ is deterministic and monotone. In particular $A(\gamma_0) = 0$, and $Q_n(\gamma_0; \lambda)$ is asymptotically distributed as a $N(0, V(\gamma; \lambda))$. The derivative A of $A(\cdot)$ at $\gamma_0 = 0$ is invertible.*

(ii) $n^{1/2}(\hat{\gamma}_\lambda - \gamma_0)$ *is asymptotically* $N(0, \dot{A}(\gamma_0)^{-1}V(\gamma_0; \lambda)\dot{A}(\gamma_0)^{-1^T})$.

In view of Theorem 1 (ii), the asymptotic variance of $\hat{\gamma}_\lambda$ is minimized by taking $\lambda = \tau$ such that $\min_\lambda V(\gamma_0; \lambda) = V(\gamma_0; \tau)$. Similarly, if Λ^{-1} exists, we have $\tau = \Lambda^{-1}\eta$. Therefore a reasonable estimate for τ would be its sample estimate $\hat{\tau} = (C^TC)^{-1}C^TR(\hat{\gamma}_0)$, where $R(\hat{\gamma}_0)$ is the vector of the ranks of $Y_i - \hat{\gamma}_0{}^TZ_i$. Also, it is clear that the solution γ for $Q_n(\gamma; \lambda) = o_p(1)$ is $n^{1/2}$-consistent. As in the two-sample case, a $(1-\alpha)$ confidence interval for γ_0 can be constructed from the test statistic $Q_n(\gamma_0; \lambda)$ without estimating $m(\cdot)$ and the error density $f(\cdot)$.

4. Partial Spline Model for Censored Data

When the response Y is the minimum of the failure time T and an independent censoring time C, take $\Delta_i = I\{T_i \le C_i\}$. Then model (1) becomes

$$T_i = \gamma^T Z_i + m(X_i) + \epsilon_i, \tag{9}$$

and the test statistic becomes

$$Q_n{}^c(\gamma; \lambda) = n^{-5/2}\sum_i\sum_j (Z_i - Z_j)\{(R_j{}^c(\gamma) - R_i{}^c(\gamma)) - \lambda^T(C_j - C_i)\}, \tag{10}$$

where

$$R_i{}^c(\gamma) = \frac{1}{2}\sum_k [I\{Y_i - \gamma^TZ_i - (Y_k - \gamma^TZ_k) > 0\}\Delta_k - I\{Y_i - \gamma^TZ_i - (Y_k - \gamma^TZ_k) < 0\}\Delta_i] \tag{11}$$

is the rank of $Y_i - \gamma^TZ_i$ for censored data, $i = 1, \ldots, n$. The subscript c of the test statistic and the ranks of the response denote that these are for censored data. This is the Gehan's (1965) generalized Wilcoxon rank-sum score in the pooled sample. Note that if $\Delta_i = 1$ for all i, then $R_i{}^c(\gamma)$ reduces to $R_i(\gamma)$, and the test statistic $Q_n{}^c(\gamma; \lambda)$ is the $Q_n(\gamma; \lambda)$ of the previous section. Furthermore, we have

$$Q_n{}^c(\gamma; \lambda) = W_n(\gamma) - n^{-3/2}\sum\sum_{i<j}(Z_i - Z_j)\sum_l \lambda_l \mathrm{sgn}(X_j^{(l)} - X_i^{(l)}) + o_p(n^{-1/2}), \tag{12}$$

where $W_n(\gamma)$ is the monotone estimating equation considered by Fygenson and Ritov (1994). Therefore, ${Q_n}^c(\gamma;\lambda)$ is a U-statistic with a symmetrical kernel of order 2

$$h^c\{(Y_1,Z_1,X_1),(Y_2,Z_2,X_2)\} = (Z_1-Z_2)[\Delta_1 I\{Y_2-\gamma^T Z_2 > (Y_1-\gamma^T Z_1)\} \\ -\Delta_2 I\{Y_1-\gamma^T Z_1 > (Y_2-\gamma^T Z_2)\}) - \sum_l \lambda_l \text{sgn}(X_{2,l}-X_{1,l})]. \quad (13)$$

The expectation for ${Q_n}^c$ is $n^{1/2}A(\gamma)$, the same as that of Fygenson and Ritov (1994), and $A(\gamma_0)=0$. However the variance of ${Q_n}^c$ is now

$$V(\gamma_0;\lambda) = {\sigma_c}^2(1-2\lambda^T\eta_c{\sigma_c}^{-2}+\lambda^T\Lambda\lambda{\sigma_c}^{-2})\text{Var}\,(Z), \quad (14)$$

where $\text{Var}\,({R_i}^c(\gamma_0)) = (n^2-1){\sigma_c}^2/12$, and $\text{Cov}\,({R_i}^c(\gamma_0),C_i) = (n^2-1)-\eta_c/12$. Therefore the asymptotic relative efficiency of the estimator with respect to $W_n(\gamma)$ of Fygenson and Ritov (1994) has the same form as equation (6), with the Spearman correlation coefficient ${R_S}^c$ based on Gehan's score instead of the usual rank of Y.

5. The Case When Z and X are Dependent

So far we have been assuming that the covariables Z and X are independent. Although the assumption is reasonable when Z is the group indicator in a randomized trial, this in general does not hold when Z is continuous in the one-sample case. When Z and X are dependent, we may adjust Z using the conditional expectation $E(Z|X)$, and treat the resultant covariable $\tilde{Z}=Z-E(Z|X)$ as independent of X, see Chen (1995). Therefore the technique in the previous sections will apply by replacing Z with $\tilde{Z}$, if the resultant model still has the form of model (1). As a special case assume X is univariate and positively correlated with each component of Z_i, and that each component of γ is positive. Then we have

$$Y_i = \gamma^T Z_i + m(Xi) + \epsilon_i = \gamma^T\tilde{Z}_i + m(X_i) + \gamma^T E(Z_i|X_i) + \epsilon_i = \gamma^T\tilde{Z}_i + \tilde{m}(X_i) + \epsilon_i, \quad (15)$$

$i=1,\ldots,n$. In fact when X is a p-dimensional vector, as long as $\gamma^T E(Z|X)$ has the same direction of change as that of $m(X)$ or, more generally, $\tilde{m}(X) = m(X)+\gamma^T E(Z|X)$ remains component-wise monotone in X, the results in the previous sections apply.

6. Simulation Studies

We performed two simulation studies. The first compared our method with that of the semiparametric regression model obtained by ignoring the nonlinear function $m(\cdot)$ of model (1). The other compared the performance of our estimate with those obtained using kernel smoothing (Speckman (1988)) and regression splines (see, e.g., He and Shi (1996)). The standard deviation (std) is reported together with the confidence intervals of significance level $\alpha = 0.05$. One thousand repetitions with a sample size of 100 were generated using SAS/IML. The average length of the confidence intervals ($E|I|$) and the coverage percentages (cova) of the true parameter that fell in the estimated confidence intervals are also listed.

Case (i). Consider the model

$$Yi = \gamma Z_i + \beta e^{X_i} + \epsilon_i,$$

where Z_i, X_i, and ϵ_i are all independent and identically distributed as $N(0,1), i = 1, \ldots, n$. We summarize the results in Table 1.

Table 1. Comparisons of the proposed estimating method with that of the semiparametric regression model without censoring and the nonlinear term.

	Q_n				W_n			
(γ, β)	$E(\hat{\gamma} - \gamma_0)$	std	$E\|I\|$	cova	$E(\hat{\gamma} - \gamma_0)$	std	$E\|I\|$	cova
(1, 1)	-0.003	0.117	0.456	93.4%	-0.008	0.166	0.649	94.2%
(1, 4)	0.007	0.120	0.480	95.5%	-0.005	0.406	1.599	94.7%
(1, 10)	-0.002	0.128	0.510	95.1%	-0.019	0.872	3.700	95.2%

Note. $W_n = n^{-3/2} \sum_i \sum_j (Z_i - Z_j) I\{Y_j(\gamma) > Y_i(\gamma)\}$ is the statistic used by Fygenson and Ritov (1994) without censored data, $Y_i(\gamma) = Y_i - \gamma^T Z_i$.

It is clear from Table 1 that the estimating equation Q_n adjusting for the ranks of the covariable X works much better than the equation W_n in terms of standard deviation and the length of the confidence interval. The improved relative efficiency is seen to be proportional to the coefficient β (and thus the variance for $m(X)$). Similar results should be expected for the right-censored data case.

Case (ii). Consider the same model as in Case (i), but with $Z_i = e^{X_i} + e_i$, $X_i \sim U(-2, 2)$, $\epsilon_i \sim \chi^2(1)$, and $e_i \sim N(0,1)$ independent of X_i, $i = 1, \ldots, n$. We replaced Z with an estimate $\tilde{Z}$ in the statistic Q_n, where $E(Z|X)$ was estimated with nonparametric smoothing using the Epanchinokov kernel with bandwidth $h = 0.2$. The estimates were compared with those obtained by kernel smoothing and with B-splines having knots at $X = -1$, 0, and 1. The results are summarized in Table 2.

In Table 2, the standard deviations of our estimates were slightly smaller than those using kernel and regression splines. This may be because rank estimates are insensitive to skewed error distributions (Chen (1997)), whereas the other two estimates considered here were not. Also, note that the bias of our estimate does not increase with β. Thus, even when X is univariate the rank estimator can work better. The superior performance would become clearer were the dimensionality of X increased.

Table 2. Comparisions of estimates from Q_n (ranks) with those by kernel smoothing and B-splines.

	kernel		B-splines		ranks	
(γ, β)	bias	std	bias	std	bias	std
(1, 1)	0.011	0.147	0.013	0.138	0.030	0.128
(1, 4)	0.025	0.158	0.037	0.156	0.028	0.142
(1, 10)	0.069	0.178	0.086	0.198	0.026	0.169

A referee has kindly pointed out that we used a fixed set of knots for the regression spline in the simulation, which could be sub-optimal as compared to adaptively chosen knots. This is certainly true. The main purpose here is, however, to demonstrate that our proposed estimator is competitive with the existing methods is such cases.

7. Concluding Remarks

We have proposed a class of rank estimating equations for a partial spline model with a monotonicity constraint in the nonlinear component $m(X)$. We think this approach has some advantages over the existing spline methods. First, using the monotonicity of our estimating equations, one can find the corresponding confidence interval of γ by converting the equations without estimating $m(X)$ or the error distribution $F(X)$. Hence it is computationally simple and stable. Second, it ensures model monotonicity which can cause technical problems. Finally, it bypasses the *curse of dimensionality* problem as well.

As far as efficiency is concerned, we can compare with the Hodges-Lemann estimator in the two-sample case and with Fygenson and Ritov's estimator for a censored regression case by ignoring $m(X)$. The ARE of our estimator is given in equation (6). Improvement can be substantial when the variance for $m(X)$ is relatively large, as shown in the simulation studies.

The proposed method still works as long as the term $\tilde{m}(X)$ of equation (15) remains monotone or approximately monotone when Z and X are dependent. Further research may be required in different situations.

Acknowledgements

The authors thank the associate editor for pointing out an error in the earlier manuscript and for constructive comments, the editor for many corrections in writing. The work was partially supported by the National Science Council of Taiwan.

Appendix

Proof of inequality (5) and the normality of $\hat{\gamma}_\lambda$:
Since

$$\hat{\gamma}_\lambda^{(L)} > a \Rightarrow Q(a;\lambda) \Rightarrow \hat{\gamma}_\lambda^{(L)} \geq a,$$

we have

$$\hat{\gamma}_\lambda^{(L)} < a \Rightarrow Q(a;\lambda) \leq 0.$$

Similarly,

$$\hat{\gamma}_\lambda^{(U)} < a \Rightarrow Q(a;\lambda) < 0 \Rightarrow \hat{\gamma}_\lambda^{(U)} \leq a.$$

Therefore,

$$\text{pr}_\gamma(\hat{\gamma}_\lambda^{(L)} < a) \leq \text{pr}_\gamma\{Q(a;\lambda) \leq 0\};\ \text{pr}_\gamma\{Q(a;\lambda) < 0\} \leq \text{pr}_\gamma(\hat{\gamma}_\lambda^{(U)} \leq a).$$

Now $\hat{\gamma}_\lambda^{(L)} \leq \hat{\gamma}_\lambda^{(U)}$ and both are continuous, so

$$\hat{\gamma}_\lambda^{(U)} \leq a \Rightarrow \hat{\gamma}_\lambda < a; \hat{\gamma}\lambda < a \Rightarrow \hat{\gamma}_\lambda^{(L)} < a.$$

Thus inequality (5) follows. Based on (5), we have for a sequence of alternatives, $\gamma_n = \gamma_0 + a/n^{1/2}$, and any given constant a,

$$\lim_{n\to\infty} \text{pr}_{\gamma_0}\{n^{1/2}(\hat{\gamma}_\lambda - \gamma_0) \leq a\} = \lim_{n\to\infty} \text{pr}_{\gamma_0}\{\hat{\gamma}_\lambda \leq \gamma_0 + a/n^{1/2}\} = \lim_{n\to\infty} \text{pr}_{\gamma_n}\{Q(\gamma_0;\lambda) \leq 0\},$$

and

$$n^{1/2}Q(\gamma_0;\lambda) = n^{1/2}[Q(\gamma_0;\lambda) - E_{\gamma_n}\{Q(\gamma_0;\lambda)\}] + n^{1/2}[E_{\gamma_n}\{Q(\gamma_0;\lambda)\} - E_{\gamma_n}\{Q(\gamma_n;\lambda)\}],$$

where $E_{\gamma_n}(\cdot)$ is the expectation under γ_n. Now, $n^{1/2}[Q(\gamma_0;\lambda) - E_{\gamma_n}\{Q(\gamma_0;\lambda)\}]$ is asymptotically normally distributed with mean zero and variance $\text{Var}(Q(\gamma_0;\lambda))$, and

$$\lim_{n\to\infty} n^{1/2}[E_{\gamma_n}\{Q(\gamma_0;\lambda)\} - E_{\gamma_n}\{Q(\gamma_n;\lambda)\}] = -a\xi'(\gamma_0).$$

Therefore ${K_{Q,\lambda}}^2 = \xi'(\gamma_0)^2/\text{Var}\{Q(\gamma_0;\lambda)\}$, and the result follows.

Proof of Theorem 1. (i)

$$\begin{aligned} Q_n(\gamma;\lambda) &= 2n^{-5/2}\sum\sum_{i<j}\{Z_i(R_j(\gamma) - \lambda^T C_j) + Z_j(R_i(\gamma) - \lambda^T C_i)\} \\ &\quad -2n^{-5/2}(n-1)\sum_i Z_i\{R_i(\gamma) - \lambda^T C_i\} \\ &= n^{-5/2}\sum\sum_{i<j}\sum_k \Big(Z_i[\text{sgn}\{Y_j - Y_k - \gamma^T(Z_j - Z_k)\} - \sum_l \lambda_l \text{sgn}(X_{j,l} - X_{k,l})] \end{aligned}$$

$$
\begin{aligned}
&+Z_j[\text{sgn}\{Y_i - Y_k - \gamma^T(Z_i - Z_k)\} - \sum_l \lambda_l \text{sgn}(X_{i,l} - X_{k,l})]\Big) \\
&-n^{-5/2}(n-1)\sum_i\sum_j Z_i[\text{sgn}\{Y_i - Y_j - \gamma^T(Z_i - Z_j)\} \\
&-\sum_l \lambda_l \text{sgn}(X_{i,l} - X_{j,l})] \\
= {}& n^{-3/2}\binom{n}{2} U(\gamma;\lambda).
\end{aligned}
$$

By a standard U-statistic theorem, $Q_n(\gamma;\lambda)$ is asymptotically normally distributed. Now, in the neighborhood of γ_0,

$$
\begin{aligned}
A(\gamma) &= \frac{1}{2}E[(Z_1 - Z_2)\text{sgn}\{Y_2 - Y_1 + \gamma^T(Z_1 - Z_2)\}] \\
&= \frac{1}{2}E[(Z_1 - Z_2)(Z_1 - Z_2)^T \int f\{u - m(X_2) + m(X_1)\}f(u)du](\gamma - \gamma_0) + o(\|\gamma - \gamma_0\|) \\
&= E[\int f\{u - m(X_2) + m(X_1)\}f(u)du]\text{Var}\,(Z)(\gamma - \gamma_0), + o(\|\gamma - \gamma_0\|),
\end{aligned}
$$

where $f(\cdot)$ the marginal p.d.f. of the model error ϵ. When $\gamma = \gamma_0$, $Y_i(\gamma_0) = Y_i - \gamma_0^T Z_i$ is independent of $Z_i, i = 1, \ldots, n$. Therefore $A(\gamma_0) = 0$, and the asymptotic variance of $Q_n(\gamma;\lambda)$ at $\gamma = \gamma_0$ is

$$
\begin{aligned}
V(\gamma_0;\lambda) = {}& \Big(E[\text{sgn}\{Y_2(\gamma_0) - Y_1(\gamma_0)\}\text{sgn}\{Y_3(\gamma_0) - Y_1(\gamma_0)\}] \\
&-2\sum_l \lambda_l E[\text{sgn}(X_{2,l} - X_{1,l})\text{sgn}\{Y_2(\gamma_0) - Y_1(\gamma_0)\}] \\
&+\sum_k\sum_l \lambda_k\lambda_l E\{\text{sgn}(X_{2,k} - X_{1,k})\text{sgn}(X_{3,l} - X_{1,l})\}\Big)\text{Var}(Z) \\
= {}& \frac{1}{3}\sigma^2(1 - 2\lambda^T\eta\sigma^{-2} + \lambda^T\Lambda\lambda\sigma^{-2})\text{Var}(Z),
\end{aligned}
$$

where, as before, $\text{Var}(R_i(\gamma_0)) = (n^2-1)\sigma^2/12$, and $\text{Cov}(R_i(\gamma_0), C_i) = (n^2 - 1)\eta/12$. The asymptotic normality with mean $A(\gamma)$ and variance $V(\gamma;\lambda)$ of $Q_n(\gamma;\lambda)$ follows. Also, the matrix derivative

$$
\dot{A}(\gamma_0) = E[\int f\{u - m(X_2) + m(X_1)\}f(u)du]\text{Var}(Z)
$$

is positive definite, and thus is invertible.
(ii) From the monotonicity of $Q_n(\gamma;\lambda)$ and $A(\gamma)$, we have that for any $M < \infty$ and $v > 0$, there are C_1, $C_2 < \infty$ such that

$$
\text{pr}\Big\{\sup_{\|t\|<M} |Q_n(\gamma_0 + t/n^{1/2};\lambda) - \dot{A}(\gamma_0)t| > C_1\Big\} < v,
$$

and

$$\mathrm{pr}\Big\{ \sup_{\|t\|<M} Q_n(\gamma_0 + t;\lambda) - n^{1/2}A(\gamma_0 + t)| > C_2 \Big\} < v.$$

That is, the statistic $Q_n(\gamma;\lambda)$ and the linear approximation of $A(\gamma)$ are uniformly convergent on bounded subsets. In addition, since $Q_n(\gamma;\lambda)$ has finite variance and $\dot{A}(\gamma_0)$ is invertible, the asymptotic normality of $\hat{\gamma}$ of the generalized solution for $Q_n(\gamma;\lambda) = 0$ and the variance formula follows from linear approximation (Brown (1985), Ritov (1987)).

The derivation of (12):

$$\begin{aligned}
Q_n{}^c(\gamma;\lambda) = {} & \frac{1}{2}n^{-5/2}\sum_i\sum_j\sum_k (Z_i - Z_j)\Delta_k[I\{Y_j - \gamma^T Z_j - (Y_k - \gamma^T Z_k) > 0\} \\
& -I\{Y_i - \gamma^T Z_i - (Y_k - \gamma^T Z_k) > 0\}] \\
& -\frac{1}{2}n^{-5/2}\sum_i\sum_j\sum_k (Z_i - Z_j)[\Delta_j I\{Y_j - \gamma^T Z_j - (Y_k - \gamma^T Z_k) < 0\} \\
& -\Delta_i I\{Y_i - \gamma^T Z_i - (Y_k - \gamma^T Z_k) < 0\}] \\
& -\frac{1}{2}n^{-5/2}\sum_i\sum_j\sum_k (Z_i - Z_j)\sum_l \lambda_l\{\mathrm{sgn}(X_{j,l} - X_{k,l}) - \mathrm{sgn}(X_{i,l} - X_{k,l})\} \\
= {} & n^{-3/2}\sum\sum_{i<j}(Z_j - Z_j)[\Delta_i I\{Y_j - \gamma^T Z_j > (Y_i - \gamma^T Z_i)\} \\
& -\Delta_j I\{Y_i - \gamma^T Z_i > (Y_j - \gamma^T Z_j)\}] \\
& -n^{-3/2}\sum\sum_{i<j}(Z_i - Z_j)\sum_l \lambda_l \mathrm{sgn}(X_{j,l} - X_{i,l}) + o_p(n^{-1/2}),
\end{aligned}$$

and (12) follows.

References

Brown, B. M. (1985). Multiparameter linearization theorems. *J. Roy. Statist. Soc. Ser. B* **47**, 323-331.

Chen, C. C. (1997). Rank transformations when covariables are measured with error or mismodelled. *Commu. Statist. Theory Method* **26**, 2967-2982.

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

Chen, H. (1995). Asymptotically efficient estimation in semiparametric generalized linear models. *Ann. Statist.* **23**, 1102-1129.

Engle, R. F., Granger, W. J., Rice, J. and Weiss, A. (1986). Semiparametric estimates of the relation between weather and electricity sales. *J. Amer. Statist. Assoc.* **81**, 310-320.

Fygenson, M. and Ritov, Y. (1994). Monotone estimating equations for censored data. *Ann. Statist.* **22**, 732-746.

Gehan, E. A. (1965). A generalized Wilcoxon test for comparing arbitrarily singly-censored samples. *Biometrika* **52**, 203-223.

Gray, R. J. (1994). Spline-based tests in survival analysis. *Biometrics* **50**, 640-652.

He, X. and Shi, P. D. (1996). Bivariate tensor-product splines in a partly linear model. *J. Multivariate Anal.* **58**, 162-181.

Heckman, N. E. (1986). Spline smoothing in a partly linear model. *J. Roy. Statist. Soc. Ser. B* **48**, 244-248.

Quade, D. (1967). Rank analysis of covariance. *J. Amer. Statist. Assoc.* **62**, 1187-1120.

Randles, R. H. and Wolfe, D. A. (1991). *Introduction to the Theory of Nonparametric Statistics.* Krieger Publishing Company, Florida.

Ritov, Y. (1987). Tightness of monotone random fields. *J. Roy. Statist. Soc. Ser. B* **49**, 331-333.

Sen, P. K. (1968). Estimates of the regression coefficient based on Kendall's tau. *J. Amer. Statist. Assoc.* **63**, 1379-1389.

Shiller, R. J. (1984). Smoothness priors and nonlinear regression. *J. Amer. Statist. Assoc.* **79**, 609-615.

Sleeper, L. A. and Harrington, D. P. (1990). Regression splines in the Cox model with application to covariate effects in the liver disease. *J. Amer. Statist. Assoc.* **85**, 941-949.

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

Wahba, G. (1990). Spline Methods for Observational Data. CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics, Philadelphia.

Department of Mathematics, Tamkang University, Taipei, Taiwan.

E-mail: ccchen@mail.tku.edu.tw

Institute of Statistical Science, Academia Sinica, Taipei, Taiwan.

E-mail: fushing@stat.sinica.edu.tw

(Received April 1997; accepted April 1998)