Ann. Inst. Statist. Math.
Vol. 55, No. 3, 585–595 (2003)


# EMPIRICAL LIKELIHOOD FOR PARTIAL LINEAR MODELS

QI-HUA WANG[1]* AND BING-YI JING[2]**

[1]*Academy of Mathematics and System Science, Chinese Academy of Science, Beijing 100080, China and Heilongjiang University, Harbin, China*
[2]*Department of Mathematics, Hong Kong University of Science and Technology, Kowloon, Hong Kong, China,* e-mail: majing@ust.hk

(Received December 12, 2001; revised October 15, 2002)

**Abstract.** In this paper the empirical likelihood method due to Owen (1988, *Biometrika*, **75**, 237–249) is applied to partial linear random models. A nonparametric version of Wilks' theorem is derived. The theorem is then used to construct confidence regions of the parameter vector in the partial linear models, which has correct asymptotic coverage. A simulation study is conducted to compare the empirical likelihood and normal approximation based method.

*Key words and phrases*: Empirical likelihood, partial linear model, Wilks' theorem.

## 1. Introduction

Suppose that $\{(X_i, T_i, Y_i), 1 \le i \le n\}$ is a random sample generated from the following partial linear random model

$$Y_i = X_i^\tau \beta + g(T_i) + \epsilon_i, \qquad i = 1, 2, \dots, n, \tag{1.1}$$

where $Y_i$'s are scalar response variates, $X_i$'s are $p$-variate covariates and $T_i$'s are scalar covariates taking values in $[0,1]$, and where $\beta$ is a $p \times 1$ column vector of unknown regression parameter, $g(\cdot)$ is an unknown measurable function on $[0,1]$ and $\epsilon_i$'s are random statistical errors. It is assumed that the errors $\epsilon_i$'s are independent and identically distributed random variables with zero mean and variance $\sigma^2 = E\epsilon_1^2$, and $\epsilon_i$'s are independent of $(X_i, T_i)$'s.

The partial linear model was introduced by Engle *et al.* (1986) to study the effect of weather on electricity demand, and further studied by Heckman (1986), Rice (1986), Speckman (1988), Chen (1988), Robinson (1988), Chen and Shiau (1991), Gao *et al.* (1994), Schick (1996) and Hamilton and Truong (1997). Various estimators for $\beta$ 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. These estimators of $\beta$ are proved to be asymptotically normal with zero mean and covariance $\sigma^2\Sigma^{-1}$ under different conditions, where $\Sigma = E(X_1 - E[X_1 \mid T_1])(X_1 - E[X_1 \mid T_1])^\tau$. For the model (1.1), a basic problem of statistical inference is the construction of confidence region for $\beta$. Clearly, the normal

*Research supported by NNSF of China and a grant to the first author for his excellent Ph.D. dissertation work in China.

**Research supported by Hong Kong RGC CERG No. HKUST6162/97P.

approximation based method can be used for this purpose, where an estimate of the asymptotic variance of $\widehat{\beta}$ can be obtained by the usual plug-in estimator. In this paper, we shall explore some other ways to construct confidence regions for $\beta$.

Owen (1988) introduced a nonparametric method of inference—an empirical likelihood method. This method defines an empirical likelihood ratio function, and use its maximum subject to a hypothesis that place restrictions on the parameter (or parametric vector) to construct confidence intervals. Hence, the method uses only the data to determine the shape and orientation of a confidence region and does not use the estimator of the asymptotic covariance. Hence, empirical likelihood is indeed appealing for the construction of confidence region of $\beta$. Actually, empirical likelihood has been studied under various models, including smooth function models, linear regression models, generalized linear and projection regression, etc. For more details, the reader is referred to Owen (1988, 1990, 1991), Qin and Lawless (1994), Hall (1990), DiCiccio *et al.* (1991), Chen (1993, 1994), Kolaczyk (1994), Qin and Wong (1996), Wang and Jing (1999) among others. In particular, Wang and Jing (1999) extended empirical likelihood method to the partial linear model with fixed designs and proved that the empirical log-likelihood function is asymptotically standard chi-square.

In this paper, we extend the empirical likelihood to the model (1.1) with fully random designs and prove that the nonparametric versions of Wilks' theorem also hold true. This can be used to construct confidence region for $\beta$. Details are given in Section 2. In Section 3, a simulation study is conducted to compare the empirical likelihood method with the normal approximation based method. Finally, the proofs of the main results are given in Section 4. Though this paper considers the same model as Wang and Jing (1999), some details and techniques to obtain the asymptotic results are different from those in Wang and Jing (1999) since the designs here are fully random. In the proof of theorems here, we give only the different details from those in Wang and Jing (1999) and omit some overlapping parts.

## 2. Description of methods and main results

The model (1.1) can be rewritten as

$$Y_i - E[Y_i \mid T_i] = (X_i - E[X_i \mid T_i])^\tau \beta + \epsilon_i, \qquad i = 1, 2, \ldots, n. \tag{2.1}$$

Let $g_1(t) = E[X_1 \mid T_1 = t]$, $g_2(t) = E[Y_1 \mid T_1 = t]$. Clearly, (2.1) could be considered as a linear model, and then empirical likelihood method could be applied to (2.1) and hence (1.1) when $g_1(t)$ and $g_2(t)$ were known.

Let

$$Z_i = (X_i - E[X_i \mid T_i])(Y_i - E[Y_i \mid T_i] - (X_i - E[X_i \mid T_i])^\tau \beta), \qquad i = 1, \ldots, n.$$

It is easy to see that $EZ_i = 0$, $i = 1, \ldots, n$, when $\beta$ is the true parameter. Hence, the problem of testing whether $\beta$ is the true parameter is equivalent to testing whether $EZ_i = 0$, for $i = 1, \ldots, n$. By Owen (1991), this may be done using empirical likelihood. Let $p_1, \ldots, p_n$ be nonnegative numbers summing to unity. Then, the empirical log-likelihood ratio, evaluated at true parameter value $\beta$, is defined by

$$l(\beta) = -2 \max_{\sum p_i Z_i = 0} \sum \log(np_i). \tag{2.2}$$

When both $g_1(t)$ and $g_2(t)$ were known and $\beta$ is the true parameter, $l(\beta)$ could be proved to be asymptotically $\chi_p^2$ distribution. And one could reject the hypothesis that $\beta$ is the true parameter when $l(\beta)$ is greater than some critical value. But, both $g_1(t)$ and $g_2(t)$ are usually unknown. Hence, the empirical log-likelihood $l(\beta)$ for $\beta$ can not be used directly to make inference on $\beta$ since it contains unknown functions $g_1(t)$ and $g_2(t)$ and hence $\beta$ is not identifiable. To solve the problem, a natural way is to replace $g_1(t)$ and $g_2(t)$ in $l(\beta)$ by their estimators, respectively, and define an estimated empirical log-likelihood. To do this, let

$$W_{nj}(t) = \frac{K\left(\dfrac{t-T_j}{h_n}\right)}{\sum_{j=1}^n K\left(\dfrac{t-T_j}{h_n}\right)}$$

where $K(\cdot)$ is a kernel function and $h_n$ a bandwidth tending to zero. Then, the estimators of $g_1(t)$ and $g_2(t)$ can be defined by

$$\widehat{g}_{1n}(t) = \sum_{j=1}^n W_{nj}(t)X_j, \qquad \widehat{g}_{2n}(t) = \sum_{j=1}^n W_{nj}(t)Y_j.$$

Now let us define

$$\widetilde{X}_i = X_i - \widehat{g}_{1n}(T_i), \qquad \widetilde{Y}_i = Y_i - \widehat{g}_{2n}(T_i), \qquad \widetilde{Z}_i = \widetilde{X}_i(\widetilde{Y}_i - \widetilde{X}_i^\tau \beta).$$

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

$$\widetilde{l}(\beta) = -2 \min_{\sum p_i \widetilde{Z}_i = 0} \sum_{i=1}^n \log(np_i). \tag{2.3}$$

By using the Lagrange multiplier method, the optimal value for $p_i$ satisfying (2.3) may be shown to be

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

where $\lambda$ is the solution of the equation

$$\frac{1}{n}\sum_{i=1}^n \frac{\widetilde{Z}_i}{1+\lambda^\tau \widetilde{Z}_i} = 0. \tag{2.4}$$

The corresponding empirical log-likelihood ratio is then

$$\widetilde{l}(\beta) = 2\sum_{i=1}^n \log\{1 + \lambda^\tau \widetilde{Z}_i\}. \tag{2.5}$$

The asymptotic distribution of $\widetilde{l}(\beta)$ is mainly decided by the asymptotic distribution of $\frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{Z}_i$ and the consistency of $\frac{1}{n}\sum_{i=1}^n \widetilde{Z}_i\widetilde{Z}_i^\tau$. Hence, $\widetilde{l}(\beta)$ has the same asymptotic distribution as $l(\beta)$ by proving $\frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{Z}_i = \frac{1}{\sqrt{n}}\sum_{i=1}^n Z_i + o_p(1)$ and

$\frac{1}{n}\sum_{i=1}^n \widetilde{Z}_i\widetilde{Z}_i^\tau = \frac{1}{n}\sum_{i=1}^n Z_iZ_i^\tau + o_p(1)$. That is, the nonparametric estimation does not affect the asymptotic result of the empirical log-likelihood ratio.

Before stating the main results, we first list the following conditions.

(A1) $g_1(t)$, $g_2(t)$ and $g(t)$ satisfy Lipschitz conditions of order 1.

(A2) $0 < \Sigma := E(X_1 - E[X_1 \mid T_1])(X_1 - E[X_1 \mid T_1])^\tau < \infty$.

(A3) The density of $T$, say $r(t)$, exists and satisfies

$$0 < \inf_{0\le t\le 1} r(t) \le \sup_{0\le t\le 1} r(t) < \infty. \tag{2.6}$$

(A4) There exists absolute constants $M_1$, $M_2$ and $\rho > 0$ such that

$$M_1 I[|t| \le \rho] \le K(t) \le M_2 I[|t| \le \rho]. \tag{2.7}$$

(A5) $\sup_t E[\|X_1\|^4 \mid T_1 = t] < \infty$, where $\|\cdot\|$ denotes the Euclidean norm.

(A6) (i) $nh_n \to \infty$, (ii) $nh_n^3 \to 0$.

(A7) $E\epsilon_1^4 < \infty$.

THEOREM 2.1. *Under the above conditions* (A1)–(A7), *we have*

$$\widetilde{l}(\beta) = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{Z}_i\right)^\tau \left(\frac{1}{n}\sum_{i=1}^n \widetilde{Z}_i\widetilde{Z}_i^\tau\right)^{-1} \left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{Z}_i\right) + o_p(1).$$

Theorem 2.1 can be used to prove the following theorem by proving $\frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{Z}_i$ is asymptotically normal and $\frac{1}{n}\sum_{i=1}^n \widetilde{Z}_i\widetilde{Z}_i^\tau$ estimates the asymptotic variance of $\frac{1}{\sqrt{n}}\widetilde{Z}_i$ consistently (see Lemma 4.1(i) and Lemma 4.2(i)).

THEOREM 2.2. *Under all the conditions listed above, if $\beta_0$ is the true value of $\beta$, then $\widetilde{l}(\beta_0)$ has an asymptotic $\chi_p^2$ distribution. That is*

$$P(\widetilde{l}(\beta_0) < c_\alpha) = 1 - \alpha + o(1)$$

*with* $P(\chi_p^2 \le c_\alpha) = 1 - \alpha$.

Clearly, Theorem 2.2 can be used not only to test the hypothesis $H_0 : \beta = \beta_0$, but also to construct confidence regions for $\beta$. Let

$$I_\alpha(\beta) = \{\beta : \widetilde{l}(\beta) \le c_\alpha\}. \tag{2.8}$$

Then, by Theorem 2.2, $I_\alpha(\beta)$ gives an approximate confidence region for $\beta$ with asymptotically correct coverage probability $1 - \alpha$, i.e.,

$$P(\beta_0 \in I_\alpha(\beta)) = 1 - \alpha + o(1). \tag{2.9}$$

## 3. Some simulation results

In this section, we shall conduct some simulation studies to compare the performance of our empirical likelihood method with the normal approximation based method. For simplicity, we shall only consider the case where $\beta$ is a scalar. It is known that under appropriate conditions the least-square type estimator of $\beta$, $\widehat{\beta}_n = \sum_{i=1}^n \widetilde{X}_i\widetilde{Y}_i/\sum_{i=1}^n \widetilde{X}_i^2$, has an asymptotically normal distribution, i.e.,

$$\sqrt{n}\widehat{\Sigma}^{1/2}(\widehat{\beta}_n - \beta)/\widehat{\sigma}_n \to N(0,1),$$

where $\widehat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n(\widetilde{Y}_i - \widetilde{X}_i\widehat{\beta}_n)^2$ and $\widehat{\Sigma} = n^{-1}\sum_{i=1}^n \widetilde{X}_i^2$. See, for instance, Speckman (1988) or Chen (1988). Therefore a two-sided confidence interval for $\beta$ at level $1-\alpha$ is given by

$$\widehat{\beta}_n \pm z_{1-\alpha/2}\widehat{\sigma}_n/(\sqrt{n}\widehat{\Sigma}^{1/2}),$$

where $z_{1-\alpha/2}$ satisfies $\Phi(z_{1-\alpha/2}) = 1-\alpha/2$ with $\Phi(\cdot)$ being the standard normal distribution.

In our simulation studies, we generate $X_i$'s and $T_i$'s from the standard normal distribution $N(0,1)$ and uniform distribution $U[0,1]$, respectively. The function $g(\cdot)$ is chosen to be $g(t) = t^2$. The kernel function $K(t)$ is the biweight kernel function

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

Also, three different bandwidths of $h_n$ are selected to be $(n\log n)^{-1/2}$, $(n\log n)^{-1/3}$ and $5(n\log n)^{-1/3}$, respectively. It is easy to check that all the conditions (A1)–(A7) in the paper are satisfied. Furthermore, the error distribution is taken to be standard normal distribution. The sample sizes have been chosen to be 10, 20 and 50, respectively. The coverage probabilities are calculated for the empirical likelihood and normal approximation methods based on 500 pairs of simulated data. The nominal levels are taken to be $\alpha = 0.10$ and 0.05, respectively. The results are presented in Tables 1, 2 and 3.

From these three tables, we see that the coverage accuracies for both the empirical likelihood method and the studentized-$t$ method generally increase as the sample size $n$ increases. Furthermore, the empirical likelihood method outperforms the studentized-$t$ method in general, particularly for small sample sizes (say, $n \le 20$). When $n$ gets large (say, $n = 50$), the difference between the two methods seems to diminish. We choose three different bandwidths $h_n = (n\log n)^{-1/2}$, $(n\log n)^{-1/3}$ and $5(n\log n)^{-1/3}$. It seems that the first choice gives the worst performance (see Table 1) while the third choice offers the best (see Table 3). Clearly, the bandwidth plays an important role here, however, we shall not address the problem of how to find the optimal bandwidth.

Table 1. Coverage probabilities for $\beta$. $h_n = (n\log n)^{-1/2}$.

| $n$ | Nominal level $\alpha = 0.10$ | | Nominal level $\alpha = 0.05$ | |
|---|---|---|---|---|
| | Studentize-$t$ | Empirical likelihood | Studentize-$t$ | Empirical likelihood |
| 10 | 0.710 | 0.730 | 0.788 | 0.864 |
| 20 | 0.788 | 0.794 | 0.848 | 0.886 |
| 50 | 0.875 | 0.855 | 0.915 | 0.920 |

Table 2. Coverage probabilities for $\beta$. $h_n = (n \log n)^{-1/3}$.

| $n$ | Nominal level $\alpha = 0.10$ | | Nominal level $\alpha = 0.05$ | |
|---|---|---|---|---|
| | Studentize-$t$ | Empirical likelihood | Studentize-$t$ | Empirical likelihood |
| 10 | 0.764 | 0.774 | 0.832 | 0.868 |
| 20 | 0.810 | 0.806 | 0.874 | 0.890 |
| 50 | 0.880 | 0.882 | 0.936 | 0.938 |

Table 3. Coverage probabilities for $\beta$. $h_n = 5(n \log n)^{-1/3}$.

| $n$ | Nominal level $\alpha = 0.10$ | | Nominal level $\alpha = 0.05$ | |
|---|---|---|---|---|
| | Studentize-$t$ | Empirical likelihood | Studentize-$t$ | Empirical likelihood |
| 10 | 0.830 | 0.852 | 0.894 | 0.914 |
| 20 | 0.862 | 0.865 | 0.924 | 0.928 |
| 50 | 0.894 | 0.902 | 0.940 | 0.942 |

## 4. Proof of theorems

LEMMA 4.1. *Under* (A1)-(A4) *and* (A6), *if* $\sup_t E[\|X_1\|^2 \mid T_1 = t] < \infty$ *and* $E\epsilon_1^2 < \infty$, *we have*

(i) $n^{-1/2}\sum_{i=1}^n \widetilde{Z}_i \xrightarrow{\mathcal{L}} N(0, \sigma^2\Sigma)$,

(ii) $n^{-1}\sum_{i=1}^n \widetilde{Z}_i = O_p(n^{-1/2})$,

*where* $\Sigma$ *is defined in* (A2) *and* $\sigma^2 = E\epsilon_1^2$.

PROOF. For $n^{-1/2}\sum_{i=1}^n \widetilde{Z}_i$, we have the following decomposition:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \widetilde{Z}_i = \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - E[X_i \mid T_i])\epsilon_i + \sum_{i=1}^8 r_{ni}, \tag{4.1}$$

where

$$r_{n1} = -\frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - E[X_i \mid T_i])\sum_{j=1}^n W_{nj}(T_i)\epsilon_j,$$

$$r_{n2} = \frac{1}{\sqrt{n}}\sum_{i=1}^n (X_i - E[X_i \mid T_i])\sum_{j=1}^n W_{nj}(T_i)(g(T_i) - g(T_j)),$$

$$r_{n3} = \frac{1}{\sqrt{n}}\sum_{i=1}^n \left[\sum_{j=1}^n W_{nj}(T_i)(g_1(T_i) - g_1(T_j))\right]\epsilon_i,$$

$$r_{n4} = -\frac{1}{\sqrt{n}}\sum_{i=1}^n \left[\sum_{j=1}^n W_{nj}(T_i)(X_j - E[X_j \mid T_j])\right]\epsilon_i,$$

$$r_{n5} = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \sum_{j=1}^{n} W_{nj}(T_i)(g_1(T_i) - g_1(T_j)) \sum_{j=1}^{n} W_{nj}(T_i)\epsilon_j \right],$$

$$r_{n6} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \sum_{j=1}^{n} W_{nj}(T_i)(X_j - E[X_j \mid T_j]) \sum_{j=1}^{n} W_{nj}(T_i)\epsilon_j \right],$$

$$r_{n7} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \sum_{j=1}^{n} W_{nj}(T_i)(g_1(T_i) - g_1(T_j)) \sum_{j=1}^{n} W_{nj}(T_i)(g(T_i) - g(T_j)) \right],$$

$$r_{n8} = -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ \sum_{j=1}^{n} W_{nj}(T_i)(X_j - E[X_j \mid T_j]) \sum_{j=1}^{n} W_{nj}(T_i)(g(T_i) - g(T_j)) \right].$$

By condition (A2) and the independence of $\epsilon_i$ and $(X_i, T_i)$, the central limit theorem can be used to prove that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - E[X_i \mid T_i])\epsilon_i \xrightarrow{\mathcal{L}} N(0, \sigma^2\Sigma). \tag{4.2}$$

From $E\epsilon_1^2 < \infty$, $\sup_t E[\|X_1\|^2 \mid T_1 = t] < \infty$ and the fact that $EW_{nj}^2(T_i) \le C(n^2h_n)^{-1}$ (see Lemma 3.1 of Wang and Zheng (1997)), we have

$$\begin{aligned} E\|r_{n1}\|^2 &= \frac{1}{n} E \left\| \sum_{j=1}^{n} \sum_{i=1}^{n} W_{nj}(T_i)(X_i - E[X_i \mid T_i])\epsilon_j \right\|^2 \\ &= \frac{\sigma^2}{n} \sum_{j=1}^{n} E \left\| \sum_{i=1}^{n} W_{nj}(T_i)(X_i - E[X_i \mid T_i]) \right\|^2 \\ &= \frac{\sigma^2}{n} \sum_{j=1}^{n} \sum_{i=1}^{n} E\{W_{nj}^2(T_i) E[\|X_i - E[X_i \mid T_i]\|^2 \mid T_i]\} \\ &\le \frac{C\sigma^2}{n} \sum_{j=1}^{n} \sum_{i=1}^{n} EW_{nj}^2(T_i) \le C(nh_n)^{-1}. \end{aligned} \tag{4.3}$$

This implies that

$$\|r_{n1}\| = o_p(1), \qquad \text{as} \quad nh_n \to \infty. \tag{4.4}$$

By (A1) and the condition: $\sup_t E[\|X_1\|^2 \mid T_1 = t] < \infty$, we have

$$\begin{aligned} &E\|r_{n2}\|^2 \\ &= \frac{1}{n} \sum_{i=1}^{n} E \left\{ E[\|X_i - E[X_i \mid T_i]\|^2 \mid T_i] \left( \sum_{j=1}^{n} W_{nj}(T_i)(g(T_i) - g(T_j)) \right)^2 \right\} \\ &\le Ch_n^2 \sum_{i=1}^{n} \sum_{j=1}^{n} E\left[ W_{nj}^2(T_i) \left| \frac{T_i - T_j}{h_n} \right|^2 \right]. \end{aligned} \tag{4.5}$$

From (A4), it is obtained that

$$W_{nj}^2(T_i)\left|\frac{T_i-T_j}{h_n}\right|^2 \leq W_{nj}^2(T_i)\left|\frac{T_i-T_j}{h_n}\right|^2 I\left[\left|\frac{T_i-T_j}{h_n}\right|\leq\rho\right]$$
$$+W_{nj}^2(T_i)\left|\frac{T_i-T_j}{h_n}\right|^2 I\left[\left|\frac{T_i-T_j}{h_n}\right|>\rho\right] \leq \rho^2 W_{nj}^2(T_i)$$

by noting that $W_{nj}(T_i)=0$ if $|\frac{T_i-T_j}{h_n}|>\rho$ for $i,j=1,2,\ldots,n$. This together with (4.5) and the fact $EW_{nj}^2(T_i)\leq C(n^2h_n)^{-1}$ yields

$$E\|r_{n2}\|^2 \leq Ch_n, \tag{4.6}$$

and hence

$$\|r_{n2}\| = o_p(1), \quad \text{as} \quad h_n\to 0. \tag{4.7}$$

Similarly to (4.3) and (4.6), we can show that

$$\begin{cases} E\|r_{ni}\|^2 \leq Ch_n, & i=3,5,8,\\ E\|r_{nj}\|^2 \leq C(nh_n)^{-1}, & j=4,6,\\ E\|r_{n7}\|^2 \leq Cnh_n^3. \end{cases} \tag{4.8}$$

Now (4.8) implies that

$$\|r_{ni}\| = o_p(1), \quad i=3,4,5,6,7,8, \tag{4.9}$$

under the assumption (A6). Therefore, Lemma 4.1(i) follows from (4.1), (4.2), (4.4), (4.7) and (4.9). Lemma 4.1(ii) is a direct consequence of Lemma 4.1(i).

LEMMA 4.2. *Under the same conditions of Theorem* 2.1, *we have*
(i) $\frac{1}{n}\sum_{i=1}^n \widetilde{Z}_i\widetilde{Z}_i^\tau = \frac{1}{n}\sum_{i=1}^n (X_i-E[X_i\mid T_i])(X_i-E[X_i\mid T_i])^\tau\epsilon_i^2+o_p(1)$.
(ii) $\frac{1}{n}\sum_{i=1}^n \widetilde{Z}_i\widetilde{Z}_i^\tau = O_p(1)$.

Using similar arguments to those employed in the proof of (4.1), (4.4), (4.7) and (4.9), we can prove Lemma 4.2(i). The detail of the proof is then omitted. Lemma 4.2(ii) is a direct result of Lemma 4.2(i) and the law of large number.

LEMMA 4.3. *Let* $\widehat{Z}_{(n)}=\max_{1\leq i\leq n}\|\widetilde{Z}_i\|$. *If* (A1), (A3), (A4), (A5), (A7) *and* (i) *of* (A6) *are satisfied, we have*

$$\widehat{Z}_{(n)} = o_p(n^{1/2}).$$

PROOF. Consider

$$\widehat{Z}_{(n)} \leq \max_{1\leq i\leq n}\left\|X_i-\sum_{j=1}^n W_{nj}(T_i)X_j\right\| \max_{1\leq i\leq n}\sum_{j=1}^n W_{nj}(T_i)|g(T_i)-g(T_j)| \tag{4.10}$$

$$
\begin{aligned}
&+ \max_{1\le i\le n}\|X_i\epsilon_i\| + \max_{1\le i\le n}\left\|X_i\sum_{j=1}^n W_{nj}(T_i)\epsilon_j\right\| \\
&+ \max_{1\le i\le n}\left\|\epsilon_i\sum_{j=1}^n W_{nj}(T_i)X_j\right\| \\
&+ \max_{1\le i\le n}\left\|\sum_{j_1=1}^n\sum_{j_2=1}^n W_{nj_1}(T_i)W_{nj_2}(T_i)X_{j_1}\epsilon_{j_2}\right\| \\
\le\ & C\max_{1\le i\le n}\|X_i\| + \max_{1\le i\le n}\|X_i\epsilon_i\| + \max_{1\le i\le n}\|X_i\|\max_{1\le i\le n}\left|\sum_{j=1}^n W_{nj}(T_i)\epsilon_j\right| \\
&+ \max_{1\le i\le n}|\epsilon_i|\max_{1\le i\le n}\left\|\sum_{j=1}^n W_{nj}(T_i)X_j\right\| \\
&+ \max_{1\le i\le n}\left\|\sum_{j_1=1}^n W_{nj_1}(T_i)X_{j_1}\right\|\max_{1\le i\le n}\left\|\sum_{j_2=1}^n W_{nj_2}(T_i)\epsilon_{j_2}\right\|.
\end{aligned}
$$

By Lemma 3 of Owen (1990), we have

$$\max_{1\le i\le n}\|X_i\| = o(n^{1/2}),\qquad \max_{1\le i\le n}\|X_i\epsilon_i\| = o(n^{1/2}),\qquad \max_{1\le i\le n}|\epsilon_i| = o(n^{1/2}). \tag{4.11}$$

For any $M > 0$, using the result $E[W_{nj}^4(T_i)] \le C(n^4h_n)^{-1}$ from Lemma 3.1 of Wang and Zheng (1997) and (A7), we have

$$
\begin{aligned}
&P\left(\max_{1\le i\le n}\left|\sum_{j=1}^n W_{nj}(T_i)\epsilon_j\right| > M(nh_n)^{-1/4}\right) \\
&\le nP\left(\left|\sum_{j=1}^n W_{nj}(T_i)\epsilon_j\right| > M(nh_n)^{-1/4}\right) \\
&\le \frac{n^2h_n}{M^4}E\left(\sum_{j=1}^n W_{nj}(T_i)\epsilon_j\right)^4 \\
&\le \frac{n^3h_n}{M^4}\sum_{j=1}^n E[W_{nj}^4(T_i)E\epsilon_j^4] \\
&\le \frac{C}{M^4}\to 0,\qquad \text{as}\quad M\to\infty.
\end{aligned} \tag{4.12}
$$

Hence,

$$\max_{1\le i\le n}\left|\sum_{j=1}^n W_{nj}(T_i)\epsilon_j\right| = O_p((nh_n)^{-1/4}). \tag{4.13}$$

Similarly, under the assumption (A5), we can show that

$$\max_{1\le i\le n}\left\|\sum_{j=1}^{n}W_{nj}(T_i)X_j\right\| = O_p(1). \tag{4.14}$$

Combining (4.10), (4.11), (4.13) and (4.14), Lemma 4.3 is then proved.

PROOF OF THEOREM 2.1. Based on Theorems 4.1–4.3, along the lines to prove Theorem 1 in Wang and Jing (1999) we can prove Theorem 2.1.

PROOF OF THEOREM 2.2. By Lemma 4.2(i) and the strong law of large numbers, it follows that

$$\frac{1}{n}\sum_{i=1}^{n}\widetilde{Z}_i\widetilde{Z}_i^{\tau} \xrightarrow{P} \sigma^2\Sigma. \tag{4.15}$$

Hence, Lemma 4.1(i), (4.15) and Theorem 2.1 together prove

$$\widetilde{l}(\beta) \xrightarrow{\mathcal{L}} \chi_p^2, \tag{4.16}$$

which completes the proof of Theorem 2.2.

## REFERENCES

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

Chen, H. and Shiau, H. J. (1991). A two-stage spline smoothing method for partially linear model, *J. Statist. Plann. Inference*, **27**, 187–201.

Chen, S. X. (1993). On the accuracy of empirical likelihood confidence regions for linear regression model, *Ann. Inst. Statist. Math.*, **45**, 621–637.

Chen, S. X. (1994). Empirical likelihood confidence intervals for linear regression coefficients, *J. Multivariate Anal.*, **49**, 24–40.

DiCiccio, T. J., Hall, P. and Romano, J. P. (1991). Bartlett adjustment for empirical likelihood, *Ann. Statist.*, **19**, 1053–1061.

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

Gao, J., Chen, X. and Zhao, L. (1994). Asymptotic normality of a class of estimators in partial linear models, *Acta Math. Sinica*, **37**, 256–268.

Hall, P. (1990). Pseudo-likelihood theory for empirical likelihood, *Ann. Statist.*, **18**, 121–140.

Hamilton, S. A. and Truong, Y. K. (1997). Local linear estimation in partly linear models, *J. Multivariate Anal.*, **60**, 1–19.

Heckman, N. (1986). Spline smoothing in partly linear models, *J. Roy. Statist. Soc. Ser. B*, **48**, 244–248.

Kolaczyk, E. D. (1994). Empirical likelihood for generalized linear models, *Statist. Sinica*, **4**, 199–218.

Owen, A. (1988). Empirical likelihood ratio confidence intervals for single functional, *Biometrika*, **75**, 237–249.

Owen, A. (1990). Empirical likelihood ratio confidence regions, *Ann. Statist.*, **18**, 90–120.

Owen, A. (1991). Empirical likelihood for linear models, *Ann. Statist.*, **19**, 1725–1747.

Qin, J. and Lawless, J. F. (1994). Empirical likelihood and general estimating equations, *Ann. Statist.*, **22**, 300–325.

Qin, J. and Wong, A. (1996). Empirical likelihood in a semiparametric models, *Scand. J. Statist.*, **23**, 209–219.

Rice, J. (1986). Concergence rates for partially splined models, *Statist. Probab. Lett.*, **4**, 203–208.

Robinson, P. M. (1988). Root $n$-consistent semiparametric regression, *Econometrica*, **56**, 931–954.

Schick, A. (1996). Root-$n$-consistent in partly linear regression models, *Statist. Probab. Lett.*, **28**, 353–358.

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

Wang, Q. H. and Jing, B. Y. (1999). Empirical likelihood for partial linear model with fixed design, *Statist. Probab. Lett.*, **41**, 425–433.

Wang, Q. H. and Zheng, Z. G. (1997). Some asymptotic properties for semiparametric regression models with censored data, *Sci. China*, **40**, 945–957.