Journal of Multivariate Analysis **72**, 132–148 (2000)
Article ID jmva.1999.1866, available online at http://www.idealibrary.com on IDEAL®

Empirical Likelihood for Partially Linear Models

Jian Shi[1]

Institute of Systems Science, Academia Sinica, and Department of Statistics, Chinese University of Hong Kong,

and

Tai-Shing Lau

Department of Statistics, Chinese University of Hong Kong
E-mail: tslau@sparc2.sta.cuhk.edu.hk

Received July 9, 1998

In this paper, we consider the application of the empirical likelihood method to partially linear model. Unlike the usual cases, we first propose an approximation to the residual of the model to deal with the nonparametric part so that Owen's (1990) empirical likelihood approach can be applied. Then, under quite general conditions, we prove that the empirical log-likelihood ratio statistic is asymptotically chi-squared distributed. Therefore, the empirical likelihood confidence regions can be constructed accordingly. © 2000 Academic Press

AMS 1991 subject classifications: 62J99, 62G20, 62G15.
Key words and phrases: partially linear model, empirical likelihood, nonparametric likelihood ratio, sieve approximation, weight functions.

1. INTRODUCTION

Consider the following partially linear model

$$y = x'\beta + g(t) + \varepsilon, \tag{1.1}$$

where y and t are uni-dimensional real numbers, x and β are k-dimensional real vectors, k is a positive integer, x and t are non-random design variables, $t \in [0, 1]$, y is an observation, β is an unknown parameter, g is an unknown real-valued function on $[0, 1]$, ε is an unobservable random error variable with mean zero, "$'$" stands for matrix transposition. Let $(x_1, t_1)', ..., (x_n, t_n)'$ be the design vectors and $y_1, ..., y_n$ be the corresponding

[1] Research partially supported by the Youth Foundation of China.

0047-259X/00 $35.00

observations. Let $\varepsilon_1, ..., \varepsilon_n$ be the i.i.d. random error variables corresponding to (1.1).

The partially linear model originated from Engle *et al.* (1986). Afterwards, it has received extensive studies in the literature. For example, see Heckman (1986), Chen (1988), Speckman (1988), Cuzick (1992a,b), among the others. Here, what we are concerned with is the statistical inference on the parametric part of the model (1.1). Then more specifically, we try to construct confidence region for the parameter β. A typical non-parametric approach to this problem generally includes the following steps: (1) derive an β_n to estimate β, (2) construct an estimate of the asymptotic variance of β_n, and (3) invert confidence region by the limiting normal distribution. However, in semiparametric and nonparametric settings, variance estimation is often complicated. In addition, confidence regions derived from the limiting normal distribution is predetermined to be symmetric which may not be adequate when the underlying distribution is typically asymmetric.

In this paper, we propose to use the empirical likelihood method to construct confidence region for β. One of the motivations is that the empirical likelihood does not involve any variance estimation and the other is that the empirical likelihood based confidence region does not have the predetermined symmetry. The method of empirical likelihood was introduced by Owen (1988) and its general property was studied by Owen (1990). Hall (1990) discussed the pseudo-likelihood theory for the method. DiCiccio, Hall and Romano (1991) proved that the empirical likelihood is Bartlett correctable and thus it has an advantage over the bootstrap. Qin and Lawless (1994) gave a general account on the equivalence between the empirical likelihood and the method of estimating equations. In addition to the general theory of the method, its various applications have also been studied by many authors, e.g., linear models (Owen (1991)), quantiles (Chen and Hall (1993)), generalized linear models (Kolaczyk (1994)), incomplete data (Li (1995) and Li *et al.* (1996)), among the others.

It can be seen that a key point in the existing papers concerning the empirical likelihood method is that the support points of some class of distributions are fixed once the parameter is given. For example, in the linear regression model $y = x'\beta + \varepsilon$, the set of support points can be chosen as $\{y_i - x_i'\beta\}_{i=1}^n$ if the parameter β is given. However, in the partially linear model, even the parametric part β is given, the usual support points $\{y_i - x_i'\beta - g(t_i)\}_{i=1}^n$ are still unknown because the nonparametric part g is unknown. Therefore, it will not be a trivial extension of Owen (1990, 1991) to establish the nonparametric likelihood ratio statistics in this model. Note that, in the model (1.1), g is an infinite dimensional nuisance parameter and can be approximated by various sieves. In this paper, the weight function method such as the nearest neighbor and the kernel will be employed to approximate g. Therefore, the approximated random error sequence can be used as the set of support points and the empirical likelihood procedure goes accordingly.

2. THE METHODOLOGY AND THE RESULT

To apply the empirical likelihood method for the partially linear model, we have to give an approximate random error sequence. Our basic idea is: suppose β is known, then the model (1.1) is reduced to a nonparametric regression model $y - x'\beta = g(t) + \varepsilon$, hence g can be estimated by using $y - x'\beta$ and t as usual. Here, we adopt the weight function method to estimate the nonparametric part g. More precisely, g can be estimated by

$$\hat{g}(t, \beta) = \sum_{i=1}^{n} W_{ni}(t)(y_i - x_i'\beta),$$

where $\{W_{ni}(t): 1 \leqslant i \leqslant n\}$ is a group of non-negative weight functions. Let $\tilde{x}_i = x_i - \sum_{j=1}^{n} W_{nj}(t_i)\, x_j$ and $\tilde{y}_i = y_i - \sum_{j=1}^{n} W_{nj}(t_i)\, y_j$ for $1 \leqslant i \leqslant n$. Let β_0 be the true parameter of the model. Then, we have an approximate residual as the following:

$$\tilde{\varepsilon}_i(\beta_0) = y_i - x_i'\beta_0 - \hat{g}(t_i, \beta_0) = \tilde{y}_i - \tilde{x}_i'\beta_0, \qquad 1 \leqslant i \leqslant n.$$

An important feature of $\tilde{\varepsilon}_i(\beta_0)$ is that $E(\tilde{\varepsilon}_i(\beta_0)) = 0$ because $\tilde{\varepsilon}_i(\beta_0) = \varepsilon_i - \sum_{j=1}^{n} W_{nj}(t_j)\, \varepsilon_j$ for $1 \leqslant i \leqslant n$. Therefore, it can be treated as a random sieve approximation of the random error sequence of $\{\varepsilon_i\}_{i=1}^{n}$. Note that, $\tilde{x}$, $\tilde{y}$ and $\tilde{\varepsilon}$ are locally centerized quantities of x, y and ε, respectively.

Now, we can define a likelihood function according to the empirical likelihood principle. For any $\beta \in R^k$, let $\tilde{\varepsilon}_i(\beta) = \tilde{y}_i - \tilde{x}_i'\beta$ for $1 \leqslant i \leqslant n$. Denote a class of distribution functions $\mathscr{F}(\beta) = \{F: F$ is a distribution function which is supported only on $\tilde{\varepsilon}_i(\beta)$ with mass p_i such that $\sum_{i=1}^{n} p_i \tilde{x}_i \tilde{\varepsilon}_i(\beta) = 0\}$. Define the empirical likelihood function of β as

$$L_n(\beta) = \max_{F \in \mathscr{F}(\beta)} \prod_{i=1}^{n} p_i. \tag{2.1}$$

Here, set $L_n(\beta) \equiv 0$ if $\mathscr{F}(\beta)$ is empty. As a consequence, a maximum empirical likelihood estimator (MELE) can be defined by $\hat{\beta}_n = \arg\max_{\beta \in R^k} L_n(\beta)$. Therefore, a nonparametric log-likelihood ratio statistic based on (2.1) is given by

$$LR(\beta_0) = \log L_n(\beta_0) - \log L_n(\hat{\beta}).$$

Let $F_{n,\beta} = \{\hat{p}_i(\beta)\}_{i=1}^n \in \mathscr{F}(\beta)$ such that $L_n(\beta) = \prod_{i=1}^n \hat{p}_i(\beta)$ for $\beta \in R^k$. It can be shown later that $L_n(\beta)$ is maximized at $\hat{\beta}$ with $\hat{p}_i(\hat{\beta}) = n^{-1}$ for $1 \leqslant i \leqslant n$ and $\hat{\beta} = (\sum_{i=1}^n \tilde{x}_i \tilde{x}_i')^{-1} \sum_{i=1}^n \tilde{x}_i \tilde{y}_i$.

It is known that Owen's empirical log-likelihood ratio statistic has a chi-squared limiting distribution which is analogous to the well known Wilk's theorem for parametric settings. So, we can expect that $LR(\beta_0)$ will also be asymptotically chi-squared distributed. To establish a theory for $LR(\beta_0)$, some necessary assumptions have to be imposed for the model. In this paper, the following assumptions are made.

Assumptions. To begin with, the following weight functions will be assumed.

A.1. Weight functions $\{W_{ni}(t) \geqslant 0 : 1 \leqslant i \leqslant n\}$ satisfy

$$\sum_{j=1}^n W_{nj}(t_i) = 1 \qquad \text{holds for all} \quad 1 \leqslant i \leqslant n,$$

$$\max_{1 \leqslant i,\, j \leqslant n} W_{nj}(t_i) = O(n^{-(1-a)}),$$

$$\max_{1 \leqslant i \leqslant n} \sum_{j=1}^n W_{ni}(t_j) = O(1),$$

$$\max_{1 \leqslant i \leqslant n} \sum_{j=1}^n W_{nj}(t_i)\, I(|t_i - t_j| > C_0 n^{-a}) = O(n^{-a}),$$

for some $0 < a < \frac{1}{2}$ and $C_0 > 0$.

Remark. The feasibility of the above condition is discussed by Shi (1998). In fact, assuming that $0 = t_0 \leqslant t_1 \leqslant t_2 \leqslant \cdots \leqslant t_n \leqslant t_{n+1} = 1$, if $\max_{1 \leqslant i \leqslant n+1} |t_i - t_{i-1}| = O(n^{-1})$, then the following k_n nearest neighbor type weight functions can be employed, that is,

$$W_{ni}(t) = \begin{cases} k_n^{-1}, & \text{if } t_i \text{ belongs to the } k_n \text{ nearest neighbor of } t, \\ 0, & \text{otherwise,} \end{cases}$$

for $1 \leqslant i \leqslant n$ and $k_n = n^{1-a}$.

A.2. $\sum_{i=1}^n \|x_i\|^2 = O(n)$ and $\max_{1 \leqslant i \leqslant n} \|\sum_{j=1}^n W_{ni}(t_j)\, \tilde{x}_j\| = o(1)$, where $\|\cdot\|$ denotes the Euclidean norm in R^k.

Remark. Note that, $\sum_{j=1}^n W_{ni}(t_j)\, \tilde{x}_j$ is a weighted average of the locally centerized quantities $\{\tilde{x}_j\}_{j=1}^n$. A.2 is a mild condition.

A.3. There exist some functions $\{h_j(\cdot)\}_{j=1}^k$ on $[0, 1]$, such that $x_i = h(t_i) + u_i$ for $1 \leqslant i \leqslant n$, where $h = (h_1, ..., h_k)'$. Moreover,

$\max_{1\leqslant i\leqslant n}\|\sum_{j=1}^n W_{nj}(t_i)\,u_j\|=o(1)$ and $\text{mineig}(n^{-1}\sum_{i=1}^n u_i u_i')$ is bounded away from 0. Hereafter, $\text{mineig}(\cdot)$ and $\text{maxeig}(\cdot)$ represent the minimum and maximum eigenvalues of a symmetric matrix, respectively. Note that, here the variable u can be viewed as the residual of the regression of x on t.

A.4. g and $\{h_j\}_{j=1}^k$ satisfy the first order Lipschitz condition on $[0, 1]$.

A.5.

$$Pr(0\in ch\{\tilde{x}_1\tilde{\varepsilon}_1(\beta_0), ..., \tilde{x}_n\tilde{\varepsilon}_n(\beta_0)\})\rightarrow 1,$$

here "ch" denotes the convex hull of a set in R^k.

Remark. Owen (1991) imposes essentially the same assumptions for application of the empirical likelihood method for linear models, see Theorem 2 of Owen (1991). A simple sufficient condition is also given by Owen (1991). Let $P=\{\tilde{x}_i \mid \tilde{y}_i-\tilde{x}_i'\beta_0>0\}$ and $N=\{\tilde{x}_i \mid \tilde{y}_i-\tilde{x}_i'\beta_0<0\}$. If $ch(P)\cap ch(N)\neq\varnothing$, then 0 is in the convex hull of $\{\tilde{x}_i\tilde{\varepsilon}_i(\beta_0)\}_{i=1}^n$. See Corollary 2 of Owen (1991). In fact, asymptotically, we have $n^{-1}\sum_{i=1}^n \tilde{x}_i\tilde{\varepsilon}_i(\beta_0)\xrightarrow{a.s.}0$ which reveals that eventually 0 will fall into the convex hull of $\{\tilde{x}_i\tilde{\varepsilon}_i(\beta_0)\}_{i=1}^n$.

Now, we can give the main result of the paper.

THEOREM. *Assume that conditions* A.1–A.5 *hold,* $E(\varepsilon^2)=\sigma^2>0$ *and* $E(|\varepsilon|^{2+\gamma})<\infty$ *for some* $\gamma>0$. *If* $\max_{1\leqslant i\leqslant n}\|x_i\|^2=o(n^{\gamma/(2+\gamma)}(\log n)^{-1})$, *then*

$$-2LR(\beta_0)\xrightarrow{d}\chi_k^2, \tag{2.2}$$

where χ_k^2 *is a chi-squared distribution with k degrees of freedom.*

Remark. In order to derive the asymptotic distribution, the existence of the fourth moment of the residuals is needed in Owen (1991) for linear models. To our knowledge, our conditions here are so far the weakest one in the derivation of the asymptotic distribution for partially linear models.

As a consequence of the theorem, confidence regions for the parameter β can be constructed by (2.2). More precisely, for any $0<\alpha<1$, let c_α be such that $Pr(\chi_k^2>c_\alpha)\leqslant 1-\alpha$. Then, $\mathscr{C}(\alpha)=^{\text{def}}\{\beta\in R^k: -2LR_n(\beta)\leqslant c_\alpha\}$ constitutes a confidence region for β with asymptotic coverage α because the event that β_0 belongs to $\mathscr{C}(\alpha)$ is equivalent to the event that $-2LR_n(\beta_0)\leqslant c_\alpha$.

There are two advantages of the above nonparametric likelihood ratio inference over the asymptotic normality approach. The first is that $\mathscr{C}(\alpha)$ is not predetermined to be symmetric so that it can better correspond with the true shape of the underlying distribution. The second is that there is no need to estimate the asymptotic variance which is rather complicated in nonparametric or semiparametric settings.

It can be seen that once the weights $\{W_{ni}(t_j): 1 \leqslant i, j \leqslant n\}$ are given and the locally centerized quantities $\tilde{x}$, $\tilde{y}$ and $\tilde{\varepsilon}(\beta)$ are computed, the computation of the empirical likelihood confidence region for this partially linear model follows the same procedure as that of the linear regression model. See Owen (1991) on this aspect.

On the issue of Bartlett correction, it is not difficult to see from the following derivation that $LR(\beta_0)$ is linear in natural, and therefore is Bartlett correctable under some regularity conditions as in usual linear models. (See Chen (1994)).

3. PROOF OF THE THEOREM

Denote locally centerized quantities $\tilde{g}_i = g(t_1) - \sum_{j=1}^n W_{nj}(t_i)\, g(t_j)$, $\tilde{h}_i = h(t_i) - \sum_{j=1}^n W_{nj}(t_i)\, h(t_j)$, $\tilde{u}_i = u_i - \sum_{j=1}^n W_{nj}(t_i)\, u_j$, $\tilde{\varepsilon}_i = \varepsilon_i - \sum_{j=1}^n W_{nj}(t_i)\, \varepsilon_j$ for $1 \leqslant i \leqslant n$. Let $\hat{x}_i = \tilde{x}_i - \sum_{j=1}^n W_{ni}(t_j)\, \tilde{x}_j$ for $1 \leqslant i \leqslant n$. Let $\tilde{A}_n = n^{-1} \sum_{i=1}^n \tilde{x}_i \tilde{x}_i'$, $\hat{A}_n = n^{-1} \sum_{i=1}^n \hat{x}_i \hat{x}_i'$, $U_n = n^{-1} \sum_{i=1}^n u_i u_i'$. Let $Z_i = \tilde{x}_i \tilde{\varepsilon}_i$ for $1 \leqslant i \leqslant n$, $\bar{Z}_n = n^{-1} \sum_{i=1}^n Z_i$, $Z_n^* = \max_{1 \leqslant i \leqslant n} \|Z_i\|$, $V_n = n^{-1} \sum_{i=1}^n Z_i Z_i'$. In this paper, we say an $k \times k$ matrix converges to zero if all its k^2 entries converge to zero uniformly, or equivalently, its maximum absolute entry converges to zero.

Before giving the proof of the theorem, some preliminary lemmas are listed in what follows. Their proofs are put into the Appendix.

LEMMA 1. *Let $\{e_1, ..., e_n\}$ be i.i.d. random variables which satisfy $E(e_1) = 0$ and $E(|e_1|^\delta) < \infty$ for some $\delta > 1$. Let $\{a_{ni}^{(j)}, 1 \leqslant i, j \leqslant n\}$ be a series of real numbers such that $\max_{1 \leqslant j \leqslant n} \sum_{i=1}^n |a_{ni}^{(j)}| \leqslant C_1 < \infty$. If $d_n = \max_{1 \leqslant i, j \leqslant n} |a_{ni}^{(j)}|$, then*

$$\max_{1 \leqslant j \leqslant n} \left| \sum_{i=1}^n a_{ni}^{(j)} e_i \right| \overset{a.s.}{=} O((n^{1/\delta} d_n \vee d_n^{1/2}) \log n).$$

LEMMA 2. *Under the same assumptions of the theorem, we have*

(i) $\sum_{i=1}^n \|\tilde{x}_i\|^2 = O(n)$ *and* $\sum_{i=1}^n \|\hat{x}_i\|^2 = O(n)$;

(ii) $\tilde{A}_n - U_n \to 0$ *and* $\hat{A}_n - \tilde{A}_n \to 0$;

(iii) $V_n - \sigma^2 U_n \xrightarrow{a.s.} 0$;

(iv) $Z_n^* \overset{a.s.}{=} o(n^{1/2} (\log n)^{-1/2})$.

Proof of the theorem. To begin with, we study the form and the property of $F_{n,\beta_0}=\{\hat{p}_i(\beta_0)\}_{i=1}^n$ as defined in Section 2 because $LR(\beta_0)$ is relevant to it directly. Write $\hat{p}_i$ for $\hat{p}_i(\beta_0)$ in the sequel.

From A.5, we know that 0 belongs to the convex hull of $\{\tilde{x}_1\tilde{\varepsilon}_1, ..., \tilde{x}_n\tilde{\varepsilon}_n\}$ with probability tending to 1 and $\mathscr{F}(\beta_0)$ is not empty. Therefore, by analogous to Owen (1990), using the Lagrange multiplier method, we have

$$\hat{p}_i=\frac{1}{n}\,\frac{1}{1+\lambda'\tilde{x}_i\tilde{\varepsilon}_i} \tag{3.1}$$

for $1\leqslant i\leqslant n$, where the Lagrange multiplier $\lambda\in R^k$ is a solution of the following equation

$$\sum_{i=1}^n \frac{\tilde{x}_i\tilde{\varepsilon}_i}{1+\lambda'\tilde{x}_i\tilde{\varepsilon}_i}=0. \tag{3.2}$$

It can be noticed that for any $\beta\in R^k$, if $\mathscr{F}(\beta)$ is not empty, the solution $\hat{p}_i(\beta)$ has the same form as (3.1). Therefore, the likelihood function $L_n(\beta)$ is maximized globally at $\hat{\beta}=A_n^{-1}\sum_{i=1}^n \tilde{x}_i\tilde{y}_i$ with $\lambda=0$ and $\hat{p}_i=n^{-1}$. Thus, we have

$$LR(\beta_0)=\log L_n(\beta_0)+n\log n=-\sum_{i=1}^n \log(1+\lambda'\tilde{x}_i\tilde{\varepsilon}_i).$$

We will study the magnitude of λ in what follows because it plays an important role in analyzing the asymptotic behaviour of $LR(\beta_0)$.

Write $\lambda=\rho\theta$, where $\rho\geqslant 0$, $\theta\in R^k$ and $\|\theta\|=1$. Note that, from (3.2) there holds

$$|\theta'\bar{Z}_n|\geqslant\frac{\rho}{1+\rho Z_n^*}\,\theta' V_n\theta\geqslant\frac{\rho}{1+\rho Z_n^*}\,\text{mineig}(V_n).$$

Because $\hat{p}_i$ is a probability mass, from (3.1) we know that $0<1+\lambda' Z_i\leqslant 1+\rho Z_n^*$. Therefore,

$$\rho[\text{mineig}(V_n)-\theta'\bar{Z}_n Z_n^*]\leqslant|\theta'\bar{Z}_n|. \tag{3.3}$$

It is needed to study the rate of convergence of $\bar{Z}_n$. By reformulation, we can see that $\bar{Z}_n=n^{-1}\sum_{i=1}^n \tilde{x}_i\tilde{\varepsilon}_i=n^{-1}\sum_{i=1}^n \hat{x}_i\varepsilon_i$. Hence $\bar{Z}_n$ is a weighted sum of independent random variables. To investigate its asymptotic distribution, we know that, by (ii) of Lemma 2, $\text{mineig}(\hat{A}_n)$ is bounded away from 0. Second, there holds

$$\max_{1 \leqslant i \leqslant n} \|\hat{x}_i\| \leqslant \max_{1 \leqslant i \leqslant n} \|\tilde{x}_i\| \left[1 + k \max_{1 \leqslant i \leqslant n} \sum_{j=1}^{n} W_{ni}(t_j) \right]$$

$$\leqslant \left[1 + k \max_{1 \leqslant i \leqslant n} \sum_{j=1}^{n} W_{ni}(t_j) \right] (k+1) \max_{1 \leqslant i \leqslant n} \|x_i\|,$$

and moreover by (2.1) and the assumption we have $\max_{1 \leqslant i \leqslant n} \|\hat{x}_i\| = o(n^{\gamma/2(2+\gamma)}(\log n)^{-1/2})$. Hence, applying the classical central limit theorem (i.e., verifying the Lindeberg–Feller conditions), we can derive that

$$\sqrt{n}\, \hat{A}_n^{-1/2} \bar{Z}_n = (n\hat{A}_n)^{-1/2} \sum_{i=1}^{n} \hat{x}_i \varepsilon_i \xrightarrow{d} N(0, \sigma^2 I_k), \tag{3.4}$$

where I_k is the $k \times k$ identity matrix, $N(0, I_k)$ is the k-dimensional standard normal distribution. As a consequence, from (i) of Lemma 2, we have

$$\begin{aligned} \|\bar{Z}_n\| &\leqslant n^{-1/2} \,\text{maxeig}(\hat{A}_n^{1/2}) \,\|\sqrt{n}\, \hat{A}_n^{-1/2} \bar{Z}_n\| \\ &= n^{-1/2} (\text{maxeig}(\hat{A}_n))^{1/2} \,\|\sqrt{n}\, \hat{A}_n^{-1/2} \bar{Z}_n\| \\ &\leqslant \left(n^{-1} \sum_{i=1}^{n} \|\hat{x}_i\|^2 \right)^{1/2} n^{-1/2} \,\|\sqrt{n}\, \hat{A}_n^{-1/2} \bar{Z}_n\| = O_p(n^{-1/2}). \end{aligned} \tag{3.5}$$

From (3.3), (iii) and (iv) of Lemma 2, and (3.5), we have $\rho = O_p(n^{-1/2})$ and $\|\lambda\| = O_p(n^{-1/2})$. Hence,

$$\max_{1 \leqslant i \leqslant n} |\lambda' Z_i| \leqslant \|\lambda\| \, Z_n^* = o_p((\log n)^{-1/2}) = o_p(1). \tag{3.6}$$

Let $R_{n1} = n^{-1} \sum_{i=1}^{n} (\lambda' Z_i)^2 Z_i (1 + \lambda' Z_i)^{-1}$. From the constraint (3.2), we have

$$\bar{Z}_n - V_n \lambda + R_{n1} = 0.$$

Because $V_n - \sigma^2 U_n \xrightarrow{a.s.} 0$, it follows, by summing the diagonal components of V_n and U_n, respectively, that $n^{-1} \sum_{i=1}^{n} \|Z_i\|^2 - n^{-1} \sum_{i=1}^{n} \|u_i\|^2 \sigma^2 \xrightarrow{a.s.} 0$. Hence, $n^{-1} \sum_{i=1}^{n} \|Z_i\|^2$ is bounded from above almost surely. Consequently,

$$\|R_{n1}\| \leqslant (\|\lambda\|^2 Z_n^*) \, n^{-1} \sum_{i=1}^{n} \frac{\|Z_i\|^2}{1 + \lambda' Z_i} = o_p((n \log n)^{-1/2}).$$

At this time, we can have

$$\lambda = V_n^{-1}\bar{Z}_n + V_n^{-1}R_{n1} = V_n^{-1}\bar{Z}_n + o_p((n\log n)^{-1/2}).$$

We now turn to study the empirical likelihood ratio statistic $LR(\beta_0) = -\sum_{i=1}^n \log(1+\lambda' Z_i)$. Under (3.6), applying the Taylor expansion, there holds

$$LR(\beta_0) = -\sum_{i=1}^n \left[\lambda' Z_i - \frac{(\lambda' Z_i)^2}{2}\right] + R_{n2}, \qquad (say)$$

where $R_{n2} \leqslant \sum_{i=1}^n |\lambda' Z_i|^3 \leqslant Z_n^* \|\lambda\|^3 \sum_{i=1}^n \|Z_i\|^2 = o_p((\log n)^{-1/2}) = o_p(1)$. We have $LR(\beta_0) = n(\lambda'\bar{Z}_n - 2^{-1}\lambda' V_n \lambda) + o_p(1)$. Note that $\lambda = V_n^{-1}\bar{Z}_n + o_p(n^{-1/2})$, $V_n - \sigma^2 U_n \overset{a.s.}{=} 0$, $\tilde{A}_n - U_n \to 0$ and $\hat{A}_n - \tilde{A}_n \to 0$, then

$$\begin{aligned} LR(\beta_0) &= -2^{-1} n\bar{Z}_n' V_n^{-1}\bar{Z}_n + o_p(1) \\ &= -2^{-1}\sigma^{-2} n^2 \bar{Z}_n' \hat{A}_n^{-1}\bar{Z}_n + o_p(1). \end{aligned}$$

Finally, by (3.4), we have

$$-2LR(\beta_0) \xrightarrow{d} \chi_k^2.$$

This completes the proof of the theorem. ▮

4. APPENDIX

In this part, we prove Lemma 2 in the first place.

Proof of Lemma 2. In the beginning, by assumptions A.1 and A.4, we have

$$\begin{aligned} \max_{1 \leqslant i \leqslant n} \|\tilde{h}_i\| &= \max_{1 \leqslant i \leqslant n} \Bigg\| \sum_{j=1}^n W_{nj}(t_i)(h(t_i) - h(t_j))(I(|t_i - t_j| \\ &\qquad > C_0 n^{-a}) + I(|t_i - t_j| \leqslant C_0 n^{-a}))\Bigg\| \\ &= O(n^{-a}). \end{aligned}$$

Now, we prove part (i) of Lemma 2. Denote $\Delta_n = \max_{1 \leqslant i \leqslant n} \|\sum_{j=1}^n W_{nj}(t_i)\, u_j\|$. From A.3, we know that $\Delta_n = o(1)$. Note that $\|\tilde{x}_i\|^2 = \|x_i - \sum_{j=1}^n W_{nj}(t_i)\, x_j\|^2 \leqslant 2\,\|x_i\|^2 + 2\max_{1 \leqslant i \leqslant n} \|\sum_{j=1}^n W_{nj}(t_i)\, x_j\|^2$. While,

$$\max_{1 \leqslant i \leqslant n} \left\| \sum_{j=1}^n W_{nj}(t_i)\, x_j \right\| \leqslant \sum_{i=1}^k \sup_{t \in [0,1]} |h_i(t)| + \Delta_n = O(1).$$

From A.2, we therefore have $\sum_{i=1}^{n} \|\tilde{x}_i\|^2 = O(n)$ which is the first assertion of (i) of the lemma. Similarly, by A.2, there holds

$$\|\hat{x}_i\|^2 = \left\| \tilde{x}_i - \sum_{j=1}^{n} W_{ni}(t_j)\, \tilde{x}_j \right\|^2$$

$$\leqslant 2\, \|\tilde{x}_i\|^2 + 2 \max_{1 \leqslant i \leqslant n} \left\| \sum_{j=1}^{n} W_{ni}(t_j)\, \tilde{x}_j \right\|^2 = 2\, \|\tilde{x}_i\|^2 + o(1).$$

The second assertion of (i) of the lemma is therefore obvious.

We now turn to prove part (ii) of Lemma 2. Because $\max_{1 \leqslant i \leqslant n} \|\tilde{h}_i\| = O(n^{-a})$, there holds

$$\tilde{A}_n = n^{-1} \sum_{i=1}^{n} (\tilde{h}_i + \tilde{u}_i)(\tilde{h}_i + \tilde{u}_i)' = n^{-1} \sum_{i=1}^{n} (\tilde{u}_i \tilde{u}_i' + \tilde{h}_i \tilde{u}_i' + \tilde{u}_i \tilde{h}_i') + O(n^{-2a}).$$

While $\|u_i\| \leqslant \|x_i\| + \sup_{t \in [0,1]} \|h(t)\|$, we have

$$\left\| n^{-1} \sum_{i=1}^{n} \tilde{h}_i \tilde{u}_i' \right\|_M = \left\| n^{-1} \sum_{i=1}^{n} \tilde{h}_i \left(u_i - \sum_{j=1}^{n} W_{nj}(t_i)\, u_j \right)' \right\|_M$$

$$\leqslant C \max_{1 \leqslant i \leqslant n} \|\tilde{h}_i\| \left(n^{-1} \sum_{i=1}^{n} \|u_i\| + \Delta_n \right)$$

$$= O(n^{-a})(O(1) + o(1)) = o(1),$$

where $\|\cdot\|_M$ is such a matrix norm that takes the maximum absolute value of its k^2 entries, C is a positive constant which characterizes the equivalence between the Euclidean norm and this matrix norm. Similarly, we have

$$\sum_{i=1}^{n} \tilde{u}_i \tilde{u}_i' = \sum_{i=1}^{n} u_i u_i' + \sum_{i=1}^{n} \left(\sum_{j=1}^{n} W_{nj}(t_i)\, u_j \right) \left(\sum_{j=1}^{n} W_{nj}(t_i)\, u_j \right)'$$

$$- \sum_{i=1}^{n} \sum_{j=1}^{n} W_{nj}(t_i)(u_i u_j' + u_j u_i'),$$

and therefore, $\|n^{-1} \sum_{i=1}^{n} \tilde{u}_i \tilde{u}' - n^{-1} \sum_{i=1}^{n} u_i u_i'\|_M \leqslant C \Delta_n^2 + 2C \Delta_n n^{-1} \sum_{i=1}^{n} \|u_i\| = o(1)$. As a consequence, we have $\|A_n - U_n\|_M \to 0$ which is the first assertion of (ii) of the lemma. We now prove the second assertion. Note that, for $1 \leqslant i \leqslant n$,

$$\hat{x}_i \hat{x}_i' = \tilde{x}_i \tilde{x}_i' + \left(\sum_{j=1}^{n} W_{ni}(t_j)\, \tilde{x}_j \right) \left(\sum_{j=1}^{n} W_{ni}(t_j)\, \tilde{x}_j \right)' - \sum_{j=1}^{n} W_{ni}(t_j)(\tilde{x}_j \tilde{x}_i' + \tilde{x}_i \tilde{x}_j')$$

By A.2, part (i) of the lemma and the relationship between the Euclidean norm and the matrix norm, we have

$$\|\hat{A}_n - \tilde{A}_n\|_M = \left\| n^{-1} \sum_{i=1}^n \hat{x}_i \hat{x}_i' - n^{-1} \sum_{i=1}^n \tilde{x}_i \tilde{x}_i' \right\|_M \leqslant o(1) + o\left(n^{-1} \sum_{i=1}^n \|\tilde{x}_i\| \right) = o(1).$$

This completes the proof of (ii) of the lemma. ▮

We now proceed to prove part (iii) of Lemma 2. It can be seen that $Z_i Z_i' = (\tilde{u}_i \tilde{u}_i' + \tilde{h}_i \tilde{h}_i' + \tilde{h}_i \tilde{u}_i' + \tilde{u}_i \tilde{h}_i') \, \tilde{\varepsilon}_i^2$. Because $\max_{1 \leqslant i \leqslant n} \|\tilde{h}_i\| = O(n^{-a})$, we have

$$\begin{aligned}
&\left\| n^{-1} \sum_{i=1}^n (\tilde{h}_i \tilde{h}_i' + \tilde{h}_i \tilde{u}_i' + \tilde{u}_i \tilde{h}_i') \, \tilde{\varepsilon}_i^2 \right\|_M \\
&= n^{-1} \sum_{i=1}^n (\|\tilde{h}_i \tilde{h}_i' + \tilde{h}_i \tilde{u}_i' + \tilde{u}_i \tilde{h}_i'\|_M) \, \tilde{\varepsilon}_i^2 \\
&\leqslant C n^{-1} \sum_{i=1}^n (\|\tilde{h}_i\|^2 + 2 \, \|\tilde{h}_i\| \, \|\tilde{u}_i\|) \, \tilde{\varepsilon}_i^2 \\
&= \left(n^{-1} \sum_{i=1}^n \tilde{\varepsilon}_i^2 \right) O(n^{-2a}) + \left(n^{-1} \sum_{i=1}^n \|\tilde{u}_i\| \, \tilde{\varepsilon}_i^2 \right) O(n^{-a}) \\
&\leqslant \left(n^{-1} \sum_{i=1}^n \tilde{\varepsilon}_i^2 \right) O(n^{-2a} + n^{-a}) + \left(n^{-1} \sum_{i=1}^n \|u_i\| \tilde{\varepsilon}_i^2 \right) O(n^{-a}).
\end{aligned}$$

While, $\tilde{\varepsilon}_i^2 = \varepsilon_i^2 - 2 \sum_{j=1}^n W_{nj}(t_i) \, \varepsilon_j \varepsilon_i + (\sum_{j=1}^n W_{nj}(t_i) \, \varepsilon_j)^2$. Because $0 < a < \frac{1}{2}$, by Lemma 1, we have

$$\max_{1 \leqslant i \leqslant n} \left| \sum_{j=1}^n W_{nj}(t_i) \, \varepsilon_j \right| \overset{a.s.}{=} O(n^{-(1-2a)/2} \log n) = o(1). \tag{4.1}$$

Therefore, by (4.1) and the strong law of large numbers, we have

$$\begin{aligned}
n^{-1} \sum_{i=1}^n \tilde{\varepsilon}_i^2 \leqslant n^{-1} \sum_{i=1}^n \varepsilon_i^2 + 2 \max_{1 \leqslant i \leqslant n} \left| \sum_{j=1}^n W_{nj}(t_i) \, \varepsilon_j \right| \left(n^{-1} \sum_{i=1}^n |\varepsilon_i| \right) \\
+ \left(\max_{1 \leqslant i \leqslant n} \left| \sum_{j=1}^n W_{nj}(t_i) \, \varepsilon_j \right| \right)^2 < \infty. \qquad (a.s.).
\end{aligned}$$

Moreover, by (4.1), we have

$$n^{-1}\sum_{i=1}^{n}\|u_i\|\,\tilde{\varepsilon}_i^2 \overset{a.s.}{=} n^{-1}\sum_{i=1}^{n}\|u_i\|\,\varepsilon_i^2 + o\left(n^{-1}\sum_{i=1}^{n}\|u_i\varepsilon_i\|\right) + o\left(n^{-1}\sum_{i=1}^{n}\|u_i\|\right).$$

From the assumption of the theorem, we know $\max_{1\leqslant i\leqslant n}\|u_i\| = o(n^{\gamma/2(2+\gamma)}(\log n)^{-1/2})$. Thus, by Lemma 1, there hold

$$n^{-1}\sum_{i=1}^{n}\|u_i\|\,(|\varepsilon_i| - E(|\varepsilon_i|)) \overset{a.s.}{=} o(n^{-1/(2+\gamma)}(\log n)^{1/2}) = o(1), \qquad \text{and}$$

$$n^{-1}\sum_{i=1}^{n}\|u_i\|\,(\varepsilon_i^2 - \sigma^2) \overset{a.s.}{=} o(n^{-(4+\gamma)/4(2+\gamma)}(\log n)^{3/4} \vee n^{-\gamma/2(2+\gamma)}(\log n)^{1/2}) = o(1).$$

Hence, we have $n^{-1}\sum_{i=1}^{n}\|u_i\|\,\tilde{\varepsilon}_i^2 < \infty$ (a.s.). Furthermore, $\|n^{-1}\sum_{i=1}^{n}(\tilde{h}_i\tilde{h}_i' + \tilde{h}_i\tilde{u}_i' + \tilde{u}_i\tilde{h}_i')\,\tilde{\varepsilon}_i^2\|_M \overset{a.s.}{\longrightarrow} 0$ and $V_n \overset{a.s.}{=} n^{-1}\sum_{i=1}^{n}\tilde{u}_i\tilde{u}_i'\tilde{\varepsilon}_i^2 + o(1)$. Note that

$$\begin{aligned} n^{-1}\sum_{i=1}^{n}\tilde{u}_i\tilde{u}_i'\tilde{\varepsilon}_i^2 &= n^{-1}\sum_{i=1}^{n}\Bigg[u_iu_i' - \sum_{j=1}^{n}W_{nj}(t_i)(u_iu_j' + u_ju_i') \\ &\qquad + \left(\textstyle\sum_{j=1}^{n}W_{nj}(t_i)\,u_j\right)\left(\sum_{j=1}^{n}W_{nj}(t_i)\,u_j\right)'\Bigg]\,\tilde{\varepsilon}_i^2 \\ &= n^{-1}\sum_{i=1}^{n}u_iu_i'\tilde{\varepsilon}_i^2 + I_{1n} + I_{2n}, \qquad (say). \end{aligned}$$

It is easy to see that $\|I_{1n}\|_M \leqslant 2C\,\Delta_n(n^{-1}\sum_{i=1}^{n}\|u_i\|\,\tilde{\varepsilon}_i^2)$ and $\|I_{2n}\|_M \leqslant C\,\Delta_n^2(n^{-1}\sum_{i=1}^{n}\tilde{\varepsilon}_i^2)$. Both I_{1n} and I_{2n} converge to zero almost surely. Consequently, $V_n - n^{-1}\sum_{i=1}^{n}u_iu_i'\tilde{\varepsilon}_i^2 \overset{a.s.}{\longrightarrow} 0$. We will show that $n^{-1}\sum_{i=1}^{n}u_iu_i'(\tilde{\varepsilon}_i^2 - \sigma^2) \overset{a.s.}{\longrightarrow} 0$ in what follows. Similarly, by (4.1), there holds

$$\begin{aligned} \left\|n^{-1}\sum_{i=1}^{n}u_iu_i'(\tilde{\varepsilon}_i^2 - \sigma^2)\right\|_M &\leqslant \left\|n^{-1}\sum_{i=1}^{n}u_iu_i'(\varepsilon_i^2 - \sigma^2)\right\|_M \\ &\quad + o\left(n^{-1}\sum_{i=1}^{n}\|u_i\|^2\,|\varepsilon_i|\right) + o\left(n^{-1}\sum_{i=1}^{n}\|u_i\|^2\right). \end{aligned}$$

We only consider the first term in the right side of the above inequality. Let $u_i=(u_{i1}, ..., u_{ik})'$ for $1 \leqslant i \leqslant n$. For any $1 \leqslant j \leqslant k$, because $\max_{1 \leqslant i \leqslant n} \|u_i\|^2 = o(n^{\gamma/(2+\gamma)}(\log n)^{-1})$, it is easy to derive from Lemma 1 that $n^{-1} \sum_{i=1}^n u_{ij}^2(\varepsilon_i^2 - \sigma^2) \overset{a.s.}{=} o(1)$. Furthermore, from the finiteness of the dimension, we know $\|n^{-1} \sum_{i=1}^n u_i u_i'(\varepsilon_i^2 - \sigma^2)\|_M \xrightarrow{a.s.} 0$.

Therefore, we can conclude from the above derivation that $V_n - \sigma^2 U_n \xrightarrow{a.s.} 0$ which is part (iii) of the lemma. As a consequence, mineig(V_n) is also bounded away from 0 almost surely.

Finally, we prove part (iv) of Lemma 2. Because $Z_i = \tilde{x}_i \tilde{\varepsilon}_i = (\tilde{h}_i + u_i - \sum_{j=1}^n W_{nj}(t_i)\, u_j)\, \tilde{\varepsilon}_i$, there hold

$$\|Z_i\| \leqslant (\max_{1 \leqslant i \leqslant n} \|\tilde{h}_i\| + \max_{1 \leqslant i \leqslant n} \|u_i\| + \Delta_{n1}) \max_{1 \leqslant i \leqslant n} |\tilde{\varepsilon}_i|, \qquad \text{and}$$

$$\max_{1 \leqslant i \leqslant n} |\tilde{\varepsilon}_i| \leqslant \max_{1 \leqslant i \leqslant n} |\varepsilon_i| + \max_{1 \leqslant i \leqslant n} \left| \sum_{j=1}^n W_{nj}(t_i)\, \varepsilon_j \right|.$$

Because $E(|\varepsilon_1|^{2+\gamma}) < \infty$, by Lemma 3 of Ghosh *et al.* (1984) we know that $\max_{1 \leqslant i \leqslant n} |\varepsilon_i| \overset{a.s.}{=} o(n^{1/(2+\gamma)})$. Hence, from Lemma 2, (4.1) and the assumptions, we have

$$Z_n^* \leqslant (\max_{1 \leqslant i \leqslant n} \|u_i\| + 1)(\max_{1 \leqslant i \leqslant n} |\varepsilon_i| + 1) \overset{a.s.}{=} o(n^{1/2}(\log n)^{-1/2}).$$

Thus, the proof of Lemma 2 is completed. ∎

Proof of Lemma 1. The proof will be divided into 5 facts.

Fact 1. Suppose Y is a random variable, $|Y| \leqslant M < \infty$, $E(Y)=0$, $E(Y^2)=\tau^2$. Then, for $0 \leqslant tM \leqslant 1$ there holds

$$E(\exp\{tY\}) \leqslant \exp\left\{\frac{t^2\tau^2}{2}\left(1+\frac{tM}{2}\right)\right\}.$$

See Petrov (1975).

Fact 2. For any $0<a<1$, we can show that there holds $\sum_{i=1}^n i^{-a} \leqslant (2^{a+1}(n+1)^{1-a})/(1-a)$.

Let $e_i' = e_i I(|e_i| \leqslant i^{1/\delta})$ and $e_i'' = e_i - e_i' = e_i I(|e_i| > i^{1/\delta})$ for $1 \leqslant i \leqslant n$. Here, $'$ is not matrix transpose. Then, for any $1 \leqslant j \leqslant n$, we have

$$\left|\sum_{i=1}^n a_{ni}^{(j)}\, e_i\right| \leqslant \left|\sum_{i=1}^n a_{ni}^{(j)}(e_i - e_i')\right| + \left|\sum_{i=1}^n a_{ni}^{(j)}(e_i' - E(e_i'))\right| + \left|\sum_{i=1}^n a_{ni}^{(j)}\, E(e_i' - e_i)\right|. \tag{4.2}$$

Fact 3.

$$\max_{1\leqslant j\leqslant n}\left|\sum_{i=1}^{n} a_{ni}^{(j)}(e_i-e_i')\right| \overset{a.s.}{=} O(d_n).$$

Because $E(|e_1|^{\delta})<\infty$, we have $\sum_{i=1}^{\infty} P(|e_1|^{\delta}>i)<\infty$. By the Borel–Cantelli lemma, we have $P(\{|e_i|>i^{1/\delta}\}\ i.o.)=0$, i.e., the event $\{|e_i|>i^{1/\delta}\}\ i\geqslant 1$ can only happen for finite many times. Therefore $\sum_{i=1}^{\infty}|e_i''|=\sum_{i=1}^{\infty}|e_i|\ I(|e_i|>i^{1/\delta})<\infty$ (*a.s.*). Moreover, we have

$$\max_{1\leqslant j\leqslant n}\left|\sum_{i=1}^{n} a_{ni}^{(j)}(e_i-e_i')\right| \leqslant d_n\sum_{i=1}^{\infty}|e_i''| \overset{a.s.}{=} O(d_n).$$

This derives the Fact 3.

Fact 4.

$$\max_{1\leqslant j\leqslant n}\left|\sum_{i=1}^{n} a_{ni}^{(j)}E(e_i-e_i')\right| = O(n^{1/\delta}\,d_n).$$

By the Fact 2, it can be seen that

$$\begin{aligned}\max_{1\leqslant j\leqslant n}\left|\sum_{i=1}^{n} a_{ni}^{(j)}E(e_i-e_i')\right| &\leqslant \max_{1\leqslant j\leqslant n}\left|\sum_{i=1}^{n}|a_{ni}^{(j)}|\,E(|e_i|\,I(|e_i|>i^{1/\delta}))\right|\\ &\leqslant d_n\sum_{i=1}^{n} i^{-(\delta-1)/\delta}E(|e_1|^{\delta})\\ &= O(n^{1/\delta}\,d_n).\end{aligned}$$

the Fact 4 is thus derived.

In what follows, we analyze the rate of convergence of $\max_{1\leqslant j\leqslant n}|\sum_{i=1}^{n} a_{ni}^{(j)}(e_i'-E(e_i'))|$. For any fixed $1\leqslant j\leqslant n$, let $Y_i=a_{ni}^{(j)}(e_i'-E(e_i'))$, $1\leqslant i\leqslant n$, $M_n=2n^{1/\delta}\,d_n$, $t=(2^{-1}n^{-1/\delta}\,d_n^{-1})\wedge d_n^{-1/2}$. It is obvious that $0\leqslant tM_n\leqslant 1$. Then, for some $\tau_n>0$, we have

$$\begin{aligned}P\left(\sum_{i=1}^{n} a_{ni}^{(j)}(e_i'-E(e_i'))>\tau_n\right) &= P\left(\sum_{i=1}^{n} Y_i>\tau_n\right)\\ &\leqslant \exp\{-t\tau_n\}\cdot E\left[\exp\left\{t\sum_{i=1}^{n} Y_i\right\}\right]. \quad (4.3)\end{aligned}$$

$E[\exp\{t\sum_{i=1}^n Y_i\}]=\prod_{i=1}^n E[\exp\{tY_i\}]$. For any $1\leqslant i\leqslant n$, because $|Y_i|\leqslant M_n$, $E(Y_i)=0$, and $0\leqslant tM_n\leqslant 1$, using the Fact 1 we have

$$E\left[\exp\left\{t\sum_{i=1}^n Y_i\right\}\right]\leqslant\prod_{i=1}^n\exp\left\{\frac{t^2}{2}E(Y_i^2)\left(1+\frac{tM_n}{2}\right)\right\}$$
$$=\exp\left\{\frac{t^2}{2}\left(1+\frac{tM_n}{2}\right)\sum_{i=1}^n E(Y_i^2)\right\}. \tag{4.4}$$

If $1<\delta<2$, then

$$E\left[\exp\left\{t\sum_{i=1}^n Y_i\right\}\right]\leqslant\exp\left\{4^{-1}n^{-2/\delta}d_n^{-2}\sum_{i=1}^n|a_{ni}^{(j)}|^2E(e_i'-E(e_i'))^2\right\}$$
$$\leqslant\exp\left\{4^{-1}n^{-2/\delta}\sum_{i=1}^n E(|e_i'|^\delta\, i^{(2-\delta)/\delta})\right\}$$
$$\leqslant\exp\left\{4^{-1}n^{-2/\delta}n^{(2-\delta)/\delta}\sum_{i=1}^n E(|e_i|^\delta)\right\}$$
$$\leqslant\exp\{4^{-1}E(|e_1|^\delta)\} \tag{4.5}$$

If $\delta\geqslant 2$, then

$$E\left[\exp\left\{t\sum_{i=1}^n Y_i\right\}\right]\leqslant\exp\left\{d_n^{-1}\sum_{i=1}^n|a_{ni}^{(j)}|^2E(e_i'-E(e_i'))^2\right\}$$
$$\leqslant\exp\left\{\sum_{i=1}^n|a_{ni}^{(j)}|\,E(e_i')^2\right\}\leqslant\exp\{C_1E(e_1^2)\}. \tag{4.6}$$

Therefore, it can be inferred from (4.3)–(4.6) that there exists some constant $C_2>0$ such that $P(\sum_{i=1}^n a_{ni}^{(j)}(e_i'-E(e_i'))>\tau_n)\leqslant\exp\{-t\tau_n+C_2\}$. Taking $t\tau_n=4\log n$, i.e., $\tau_n=4(2n^{1/\delta}d_n\vee d_n^{1/2})\log n$, there holds $P(\sum_{i=1}^n a_{ni}^{(j)}(e_i'-E(e_i'))>\tau_n)\leqslant\exp\{-3\log n\}=n^{-3}$ when n is large enough. Moreover, symmetrically we have $P(\sum_{i=1}^n a_{ni}^{(j)}(e_i'-E(e_i'))<-\tau_n)\leqslant n^{-3}$. Thus, $P(|\sum_{i=1}^n a_{ni}^{(j)}(e_i'-E(e_i'))|>\tau_n)\leqslant 2n^{-3}$. From the above derivation, when n is large enough, we have

$$P\left(\max_{1\leqslant j\leqslant n}\left|\sum_{i=1}^n a_{ni}^{(j)}(e_i'-E(e_i'))\right|>4(2^{1/2}n^{1/\delta}d_n\vee d_n^{1/2})\log n\right)$$
$$\leqslant\sum_{j=1}^n P\left(\left|\sum_{i=1}^n a_{ni}^{(j)}(e_i'-E(e_i'))\right|>4(2^{1/2}n^{1/\delta}d_n\vee d_n^{1/2})\log n\right)$$
$$\leqslant 2\sum_{j=1}^n n^{-3}=2n^{-2}.$$

By the Borel–Cantelli lemma, we have the following fact

Fact 5.

$$\max_{1 \leqslant j \leqslant n} \left| \sum_{i=1}^{n} a_{ni}^{(j)}(e_i' - E(e_i')) \right| \stackrel{a.s.}{=} O((n^{1/\delta}\, d_n \vee d_n^{1/2}) \log n).$$

Finally, the lemma is proved by combining (4.2) and the Facts 3, 4 and 5.

ACKNOWLEDGMENTS

Our sincere thanks are due to the Editor and the referees for their valuable comments and helpful suggestions which have led to substantial improvements in the manuscript.

REFERENCES

1. H. Chen, Convergence rates for parameter components in a partly linear model, *Ann. Statist.* **16** (1988), 136–146.
2. S. X. Chen, Empirical likelihood confidence intervals for linear regression coefficients, *J. Multivariate Anal.* **49** (1994), 24–90.
3. S. X. Chen and P. Hall, Smoothed empirical likelihood confidence intervals for quantiles, *Ann. Statist.* **21** (1993), 1166–1181.
4. J. Cuzick, Semiparametric additive regression, *JRSS* (*B*) **54** (1992a), 831–843.
5. J. Cuzick, Efficient estimates in semiparametric additive regression models with unknown error distribution, *Ann. Statist.* **20** (1992b), 1129–1136.
6. T. J. DiCiccio, P. J. Hall, and J. P. Romano, Empirical likelihood is Bartlett-correctable, *Ann. Statist.* **19** (1991), 1053–1061.
7. R. F. Engle *et al.*, Semiparametric estimates of the relation between weather and electricity sales, *JASA* **81** (1986), 310–320.
8. M. Ghosh, W. C. Parr, K. Singh, and J. Babu, A note on bootstrapping the sample median, *Ann. Statist.* **12** (1984), 1130–1135.
9. P. Hall, Pseudo-likelihood theory for empirical likelihood, *Ann. Statist.* **18** (1990), 121–140.
10. N. E. Heckman, Spline smoothing in partly linear models, *JRSS* (*B*) **48** (1986), 244–248.
11. E. D. Kolaczyk, Empirical likelihood for general linear models, *Statist. Sinica* **4** (1994), 199–218.
12. G. Li, Nonparametric likelihood ratio estimation of probabilities for truncated data, *JASA* **90** (1995), 997–1003.
13. G. Li, M. Hollander, I. W. Mckeague, and J. Yang, Nonparametric likelihood ratio confidence bands for quantile functions from incomplete survival data, *Ann. Statist.* **24** (1996), 628–640.
14. A. B. Owen, Empirical likelihood ratio confidence intervals for a single functional, *Biometrika* **75** (1988), 237–249.
15. A. B. Owen, Empirical likelihood confidence regions, *Ann. Statist.* **18** (1990), 90–120.
16. A. B. Owen, Empirical likelihood for linear models, *Ann. Statist.* **19** (1991), 1725–1747.
17. V. V. Petrov, "Sums of Independent Random Variables," Springer, New York, 1975.

18. J. Qin and J. F. Lawless, Empirical likelihood and general estimating equations, *Ann. Statist*. **22** (1994), 300–325.
19. J. Shi, Edgeworth expansion in partial linear models, *Acta Math. Sinica* **14** (1998), 303–314.
20. P. Speckman, Kernel smoothing in partial linear models, *JRSS* (*B*) **50** (1988), 413–436.