Ann. Inst. Statist. Math.
Vol. 45, No. 4, 621–637 (1993)

ON THE ACCURACY OF EMPIRICAL LIKELIHOOD CONFIDENCE REGIONS FOR LINEAR REGRESSION MODEL

SONG XI CHEN*

Centre for Mathematics and Its Applications, Australian National University, GPO Box 4, Canberra, ACT 2601, Australia

(Received October 7, 1991; revised June 22, 1992)

Abstract. The coverage errors of the empirical likelihood confidence regions for β in a linear regression model, $Y_i = x_i\beta + \epsilon_i$, $1 \le i \le n$, are of order n^{-1}. Bartlett corrections may be employed to reduce the order of magnitude of the coverage errors to n^{-2}. For practical implementation of Bartlett correction, an empirical Bartlett correction is given.

Key words and phrases: Bartlett correction, confidence regions, coverage, empirical likelihood, linear regression model.

1. Introduction

Considering a linear regression model of the form,

$$Y_i = x_i\beta + \epsilon_i, \qquad 1 \le i \le n, \tag{1.1}$$

where β is a $p \times 1$ vector of unknown parameters and x_i is a $1 \times p$ vector of the i-th fixed design point, for which scalar Y_i is the response. We allow the ϵ_i's to be heteroscedastic, that is, the ϵ_i's are independent random variables with mean zero and variances $\sigma^2(x_i)$. The data are observed in the form $\{(x_i, Y_i) \mid 1 \le i \le n\}$.

A classical problem for linear regression model is how to construct confidence regions for β, when the distribution functions of ϵ_i's are unknown. In these kinds of nonparametric settings, the bootstrap has been used to construct confidence regions for β. But one drawback of the bootstrap is that it needs some subjective instructions on the shapes and orientations of the confidence regions. Empirical likelihood methods, as alternatives to the bootstrap method for constructing confidence regions nonparametrically, were introduced by Owen (1988, 1990). An important feature of empirical likelihood is that it uses only the data to determine the shape and orientation of a confidence region. Furthermore in certain regular cases, empirical likelihood confidence regions are Bartlett correctable, meaning

* Now at Division of Fisheries and Biometrics Unit, CSIRO Australia, GPO Box 1538, Hobart, Tasmania 7001, Australia.

that simple empirical adjustments for scale can reduce coverage error from $O(n^{-1})$ to $O(n^{-2})$; for the case of smoothed function of means see DiCiccio *et al.* (1991), and for that of a quantile see Chen and Hall (1993).

Empirical likelihood methods were proposed by Owen (1991) to construct confidence regions for β in the model (1.1). He derived a nonparametric version of Wilks' theorem, ensuring that empirical likelihood confidence regions for β have correct asymptotic coverages. There are, however, two questions to be answered. They are "How accurate are the empirical likelihood confidence regions?" and "Are the empirical likelihood confidence regions Bartlett correctable?"

This paper aims to answer these two questions. We demonstrate in Section 2 that the coverage errors of the empirical likelihood confidence regions for β are of order n^{-1}. In Section 3 we show that Bartlett correction may be used to reduce the order of magnitude of the coverage errors to n^{-2}. An empirical Bartlett correction is given, which allow one to practically implement the Bartlett correction. A simulation study is presented in Section 4.

We close this section with the following notations. Let X be an $n \times p$ matrix with x_i as the i-th row; and β_{LS} denote the least squares estimator of β, $\beta_{\text{LS}} = (X^T X)^{-1} \sum x_i Y_i$; and $\hat{\epsilon}_i = Y_i - x_i \beta_{\text{LS}}$.

2. Wilks' theorem and coverage accuracy

As mentioned in Section 1, Owen (1991) proved a nonparametric version of Wilks' theorem for the empirical likelihood of β, which enables us to construct confidence regions with correct asymptotic coverages. In this section we investigate the second order property of those confidence regions. We first give a Taylor expansion for empirical log likelihood ratio, which is denoted by $\ell(\beta)$. Then we set up an Edgeworth expansion for the distribution function of $\ell(\beta)$, which allows us to evaluate coverage accuracy of the empirical likelihood confidence regions.

For the linear regression model (1.1), we know that

$$E(Y_i \mid x_i) = x_i \beta, \qquad E(\epsilon_i) = 0, \qquad \text{Var}(\epsilon_i) = \sigma^2(x_i).$$

We define auxilliary variables $z_i = x_i^T (Y_i - x_i \beta)$, for $1 \le i \le n$, and

$$V_n = n^{-1} \sum_{i=1}^{n} \text{Cov}(z_i) = n^{-1} \sum_{i=1}^{n} x_i^T x_i \sigma^2(x_i),$$

and let v_{1n} and v_{pn} denote the largest and smallest eigenvalues of V_n, respectively.

The problem of testing if β is the true parameter is equivalent to testing if $E\{z_i\} = 0$, for $1 \le i \le n$. Let $p_1, \ldots, p_n$ be nonnegative numbers adding to unity. Then the empirical log likelihood ratio, evaluated at true parameter value β, is defined by

$$\ell(\beta) = -2 \min_{\sum p_i z_i = 0} \sum \log(n p_i).$$

Using the Lagrange multiplier method, the optimal value for p_i may be shown to be given by,

$$p_i = \frac{1}{n} \frac{1}{1 + \lambda^T z_i} \qquad 1 \le i \le n.$$

This gives

$$\ell(\beta) = 2\sum \log(1 + \lambda^T z_i),$$

where λ is a $p \times 1$ vector satisfying

$$n^{-1}\sum \frac{z_i}{1 + \lambda^T z_i} = 0.$$

In terms of studentized variables $w_i = V_n^{-1/2} z_i$, for $1 \le i \le n$, we have

$$\ell(\beta) = 2\sum \log(1 + \lambda^T w_i), \tag{2.1}$$

where λ satisfies

$$n^{-1}\sum \frac{w_i}{1 + \lambda^T w_i} = 0. \tag{2.2}$$

Since analytic solution of equations (2.1) and (2.2) can rarely be attained, we have to derive an asymptotic expansion for $\ell(\beta)$. To this end, we assume the following regularity condition.

(2.3) There exist positive constants C_1 and C_2 such that uniformly in n, $C_1 < v_{pn} \le v_{1n} < C_2$; and $n^{-2}\sum_{j=1}^{n} E\|z_j\|^4 \to 0$, where $\|\ \|$ is the Euclidean norm.

Under condition (2.3), Owen (1991) shows that the λ appearing in (2.2) satisfies

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

We define

$$\bar{\alpha}^{j_1\cdots j_k} = n^{-1}\sum E(w_i^{j_1}\cdots w_i^{j_k}),$$
$$A^{j_1\cdots j_k} = n^{-1}\sum (w_i^{j_1}\cdots w_i^{j_k} - \bar{\alpha}^{j_1\cdots j_k}),$$

where w_i^j is the j-th component of w_i. In particular, $\bar{\alpha}^j = 0$, $\bar{\alpha}^{j\,k} = \delta^{j\,k}$, $\delta^{j\,k}$ is the Kronecker delta.

Notice that $\ell(\beta)$, given by (2.1) and (2.2), is similar to the empirical log likelihood ratio for means in the independent and identically distributed case. The only difference is that $\{w_i\}_{i=1}^n$ are independent but not identically distributed random variables due to the presence of the fixed design points. However, by modifying the expansion (3.6) in DiCiccio *et al.* (1988) we may obtain the following expansion for $\ell(\beta)$,

$$\begin{aligned} n^{-1}\ell(\beta) = {} & A^j A^j - A^{j\,k} A^j A^k + \frac{2}{3}\bar{\alpha}^{j\,k\,l} A^j A^k A^l + A^{j\,l} A^{k\,l} A^j A^k \\ & + \frac{2}{3} A^{j\,k\,l} A^j A^k A^l - 2\bar{\alpha}^{j\,k\,m} A^{l\,m} A^j A^k A^l \\ & + \bar{\alpha}^{j\,k\,n}\bar{\alpha}^{l\,m\,n} A^j A^k A^l A^m \\ & - \frac{1}{2}\bar{\alpha}^{j\,k\,l\,m} A^j A^k A^l A^m + O_p(n^{-5/2}). \end{aligned} \tag{2.4}$$

We use here the convention that terms with repeated indices are to be summed over. Based on expansion (2.4), we have

$$\ell(\beta) = (n^{1/2}R^T)(n^{1/2}R) + O_p(n^{-3/2}), \tag{2.5}$$

where $R = R_1 + R_2 + R_3$ is a p-dimensional vector and $R_l = O_p(n^{-l/2})$ for $l = 1, 2, 3$. Comparing terms in (2.4) with those in (2.5) yields,

$$R_1^j = A^j, \quad R_2^j = -\frac{1}{2}A^{j\ k}A^k + \frac{1}{3}\bar{\alpha}^{j\ k\ m}A^k A^m,$$
$$R_3^j = \frac{3}{8}A^{j\ m}A^{k\ m}A^k + \frac{1}{3}A^{j\ k\ m}A^k A^l - \frac{5}{12}\bar{\alpha}^{j\ k\ m}A^{l\ m}A^k A^l$$
$$- \frac{5}{12}\bar{\alpha}^{k\ l\ m}A^{j\ m}A^k A^l + \frac{4}{9}\bar{\alpha}^{j\ k\ n}\bar{\alpha}^{l\ m\ n}A^m A^k A^l - \frac{1}{4}\bar{\alpha}^{j\ k\ l\ m}A^m A^k A^l,$$

where R_l^j is the j-th component of R_l. Considering only the dominant term in expansion (2.4),

$$\begin{aligned}\ell(\beta) &= n^{-1}\left\{\sum(Y_i - x_i\beta)x_i\right\}V_n^{-1}\left\{\sum x_i^T(Y_i - x_i\beta)\right\} + O_p(n^{-1/2}) \\ &= (\beta_{\text{LS}} - \beta)^T\{\text{Var}(\beta_{\text{LS}})\}^{-1}(\beta_{\text{LS}} - \beta) + O_p(n^{-1/2}).\end{aligned} \tag{2.6}$$

Since $\beta_{\text{LS}} - \beta$ converges to $N\{0, \text{Var}(\beta_{\text{LS}})\}$ in distribution under condition (2.3),

$$(\beta_{\text{LS}} - \beta)^T \text{Var}^{-1}(\beta_{\text{LS}})(\beta_{\text{LS}} - \beta) \xrightarrow{d} \chi_p^2, \quad \text{as} \quad n \to \infty.$$

Thus we obtain

$$P\{\ell(\beta) < c\} = P(\chi_p^2 < c) + o(1) \quad \text{as} \quad n \to \infty, \tag{2.7}$$

which is the nonparametric version of Wilks' theorem, and first proved by Owen (1991).

From (2.6) we can see that $\ell(\beta)$ implicitly uses the true variance of β_{LS} to construct confidence regions for β. This is an advantage of empirical likelihood over other resampling techniques, such as jackknife and bootstrap, which depend on explicit estimates of $\text{Var}(\beta)$ and subsequently create problems about the quality of these estimates. This point was noted by Wu (1986). Empirical likelihood can avoid this problem, reflecting the feature "let the data themselves decide". And also note that the first term on the right of (2.6) is different from that given by Owen (1991), who uses an estimate of $\text{Var}(\beta_{\text{LS}})$. However the difference has no first order effect.

Using (2.7), a confidence region for β with nominal coverage level α can be constructed as follows. First find from χ_p^2 tables the value c_α such that

$$P(\chi_p^2 < c_\alpha) = \alpha.$$

Then $R_\alpha = \{\beta \mid \ell(\beta) < c_\alpha\}$ is the α confidence region for β, and (2.7) ensures that it has correct asymptotic coverage. Before discussing the coverage accuracy of R_α, let us define $j_1 = (p^2 + p)/2$, $j_2 = j_1/2 + p(p+1)(2p+1)/12$, and

$$\bar{U} = (A^1, \ldots, A^p, A^{1\ 1}, \ldots, A^{p\ p}, A^{1\ 1\ 1}, \ldots, A^{p\ p\ p}),$$

being the $p + j_1 + j_2$ dimensional vector consisting of all distinct first three order multivariate central moments of $w_i = V_n^{-1/2} z_i$'s. Note that there are j_1 and j_2 different second and third order multivariate central moments in $\bar{U}$. Let $T_n = n\,\mathrm{Cov}(\bar{U})$. We define square matrices H_1 and H_2 as follows,

$$H_1 = n^{-1} \sum (x_i^T \otimes x_i^T)(x_i \otimes x_i)\{\epsilon_i^4 - E^2(\epsilon_i^2)\},$$
$$H_2 = n^{-1} \sum (x_i^T \otimes x_i^T \otimes x_i^T)(x_i \otimes x_i \otimes x_i)\{\epsilon_i^6 - E^2(\epsilon_i^3)\}.$$

There are p^2 rows and columns for H_1 and p^3 for H_2. Notice that the ranks of H_1 and H_2 are not larger than j_1 and j_2, respectively. We denote $\eta_{j_1 n}$ and $\zeta_{j_2 n}$ as the j_1-th and j_2-th largest eigenvalues of H_1 and H_2 respectively. Moreover let S_i be the $(p-i) \times p$ matrix obtained by removing top i rows of $V_n^{-1/2}$, and $V_{nj}^{-1/2}$ the j-th row of $V_n^{-1/2}$. Clearly $S_o = V_n^{-1/2}$ and $S_{p-1} = V_{np}^{-1/2}$. We define $j_1 \times p^2$ and $j_2 \times p^3$ matrices B_1 and B_2 as follows,

$$B_1 = \begin{pmatrix} V_{n1}^{-1/2} \otimes S_o \\ \vdots \\ V_{np}^{-1/2} \otimes S_{p-1} \end{pmatrix} \quad \text{and} \quad B_2 = \begin{pmatrix} V_{n1}^{-1/2} \otimes V_{n1}^{-1/2} \otimes S_o \\ \vdots \\ V_{n1}^{-1/2} \otimes V_{np}^{-1/2} \otimes S_{p-1} \\ V_{n2}^{-1/2} \otimes V_{n2}^{-1/2} \otimes S_2 \\ \vdots \\ V_{n2}^{-1/2} \otimes V_{np}^{-1/2} \otimes S_{p-1} \\ \vdots \\ V_{np}^{-1/2} \otimes V_{np}^{-1/2} \otimes S_{p-1} \end{pmatrix}.$$

To derive an Edgeworth expansion for the distribution of $\ell(\beta)$, we need the following lemma, whose proof is deferred to the Appendix.

LEMMA 2.1. *Assume that*

(2.8) *both* $\eta_{j_1 n} - v_{pn}^{-1}$ *and* $\zeta_{j_2 n} - v_{pn}^{-1} - (\eta_{j_1 n} - v_{pn}^{-1})^{-1}$ *are positive and bounded away from zero; the ranks of* B_1 *and* B_2 *are* j_1 *and* j_2 *respectively.*

Then the smallest eigenvalue of T_n *is bounded away from zero.*

We use this lemma to establish the following theorem.

THEOREM 2.1. *Assume* (2.8) *and that*

(2.9) (i) *there exist positive constants* C_1, C_2 *such that uniformly in* n, $C_1 \le v_{pn} \le v_{1n} \le C_2$; (ii) $\|x_i\|$*'s for* $1 \le i \le n$ *are uniformly bounded;* (iii) $\sup_n n^{-1} \sum_{j=1}^{n} E|\epsilon_j|^{15} < \infty$; (iv) *for every positive* τ,

$$\lim_{n\to\infty} n^{-1}\sum_{j=1}^{n}\int_{|\epsilon_j|>\tau n^{1/2}} |\epsilon_j|^{15} = 0;$$ (v) *for every positive* b, *the characteristic function* g_n *of* ϵ_n *satisfies Cramér's condition* $\overline{\lim}_{n\to\infty} \sup_{|t|>b} |g_n(t)| < 1$.

Then $P\{\ell(\beta) < c_\alpha\} = \alpha - ac_\alpha g_p(c_\alpha)n^{-1} + O(n^{-3/2})$ *where* g_p *is the density of* χ_p^2 *distribution,* $P(\chi_p^2 < c_\alpha) = \alpha$, *and*

$$a = p^{-1}\left(\frac{1}{2}\bar{\alpha}^{j\ j\ m\ m} - \frac{1}{3}\bar{\alpha}^{j\ k\ m}\bar{\alpha}^{j\ k\ m}\right). \tag{2.10}$$

Theorem 2.1 states that the coverage error of the empirical likelihood confidence region R_α is of order n^{-1}. From (2.10) and the definitions of $\bar{\alpha}^{j\ j\ m\ m}$ and $\bar{\alpha}^{j\ k\ m}$, we have

$$a = p^{-1}\left[\frac{1}{2}n^{-1}\sum_{i=1}^{n} E(\epsilon_i^4)(x_iV_n^{-1}x_i^T)^2 - \frac{1}{3}n^{-2}\sum_{i,l}\{E(\epsilon_i^3)E(\epsilon_l^3)(x_iV_n^{-1}x_l)^3\}\right].$$

This reveals that the coverage error depends on a combination of the following five factors: (i) the moments of ϵ_i's, (ii) the nominal coverage level, (iii) the configuration of the fixed design points, (iv) the sample size n, and (v) dimension, p.

PROOF OF THEOREM 2.1. To prove the theorem we first derive an Edgeworth expansion for the distribution of $n^{1/2}R$. By the expansion $R = R_1+R_2+R_3$ and expressions for R_l, $l = 1, 2, 3$, we may prove that the cumulants $k_1, k_2, \ldots$ of $n^{1/2}R$ satisfy the following formulae:

$$k_1 = n^{-1/2}\mu + O(n^{-3/2}), \quad k_2 = I + n^{-1}\Delta + O(n^{-2}),$$
$$k_j = O(n^{-3/2}) \quad j \geq 3,$$

where I is the $p\times p$ identity matrix, $\mu = (\mu^1, \ldots, \mu^p)^T$, $\Delta = (\Delta_{i\ j})_{p\times p}$ and

$$\mu^j = -\frac{1}{6}\bar{\alpha}^{j\ k\ k}, \quad \Delta_{i\ j} = \frac{1}{2}\bar{\alpha}^{i\ j\ m\ m} - \frac{1}{3}\bar{\alpha}^{i\ k\ m}\bar{\alpha}^{j\ k\ m} - \frac{1}{36}\bar{\alpha}^{i\ j\ m}\bar{\alpha}^{m\ k\ k}.$$

Let $\mathcal{B}$ be a class of Borel sets satisfying

$$\sup_{B\in\mathcal{B}}\int_{(\partial B)^\epsilon} \phi(v)dv = O(\epsilon), \qquad \epsilon \downarrow 0, \tag{2.11}$$

where (∂B) and $(\partial B)^\epsilon$ are the boundary of B and ϵ-neighborhood of (∂B) respectively. A formal Edgeworth expansion for the distribution function of $n^{1/2}R$ is given as follows,

$$\sup_{B\in\mathcal{B}}\left|P(n^{1/2}R\in B) - \int_B \pi(v)\phi(v)dv\right| = O(n^{-3/2}), \tag{2.12}$$

where

$$\pi(v) = 1 + n^{-1/2}\mu^T v + \frac{1}{2}n^{-1}\{v^T(\mu\mu^T + \Delta)v - \mathrm{tr}(\mu\mu^T + \Delta)\},$$

$\phi(v)$ is the density function of standard p-dimensional normal distribution, and tr is the trace operation for square matrices.

Accepting that the Edgeworth expansion (2.12) may be justified, we shall develop an Edgeworth expansion for the distribution of $\ell(\beta)$. Put

$$H = (h_{i\ j})_{p\times p} = \mu\mu^T + \Delta.$$

From (2.5) and by the symmetry of $\phi(v)$ we have

$$\begin{aligned}
(2.13)\quad & P\{\ell(\beta) < c_\alpha\} \\
&= P\{(n^{1/2}R)^T(n^{1/2}R) < c_\alpha\} + O(n^{-3/2}) \\
&= \int_{\|v\|<c_\alpha^{1/2}} \pi(v)\phi(v)dv + O(n^{-3/2}) \\
&= P(\chi_p^2 < c_\alpha) \\
&\quad + \frac{1}{2}n^{-1}\int_{\|v\|<c_\alpha^{1/2}} \left\{\sum_{i=1}^{p} h_{i\ i}(v_i^2 - 1) + \sum_{i\neq j} h_{i\ j}v_iv_j\right\}\phi(v)dv \\
&\quad + O(n^{-3/2}) \\
&= \alpha - p^{-1}\sum_{i=1}^{p} h_{i\ i}c_\alpha g_p(c_\alpha)n^{-1} + O(n^{-3/2}).
\end{aligned}$$

After some simple algebra we may show that

$$p^{-1}\sum_{i=1}^{p} h_{i\ i} = p^{-1}\left(\frac{1}{2}\bar{\alpha}^{j\ j\ m\ m} - \frac{1}{3}\bar{\alpha}^{j\ k\ m}\bar{\alpha}^{j\ k\ m}\right).$$

Thus from (2.13) we obtain

$$P\{\ell(\beta) < c_\alpha\} = \alpha - ac_\alpha g_p(c_\alpha)n^{-1} + O(n^{-3/2}).$$

However, it remains to check that the formal expansion (2.12) is valid. Since

$$\bar{U} = (A^1, \ldots, A^p, A^{1\ 1}, \ldots, A^{p\ p}, A^{1\ 1\ 1}, \ldots, A^{p\ p\ p})^T = n^{-1}\sum U_i,$$

where

$$U_i = [x_i^T V_n^{-1/2}\epsilon_i, (x_i\otimes x_i)B_1^T\{\epsilon_i^2 - E(\epsilon_i^2)\}, (x_i\otimes x_i\otimes x_i)B_2^T\{\epsilon_i^3 - E(\epsilon_i^3)\}]^T,$$

we see that $\bar{U}$ is the mean of independent but not identically distributed random vectors due to the presence of the fixed design points. However an Edgeworth expansion for this case is established by Bhattacharya and Rao ((1976), Theorem 20.6). Using Lemma 2.1, it may be shown that conditions (2.8) and (2.9) imply the conditions of Theorem 20.6 of Bhattacharya and Rao (1976). Thus, we may establish the following Edgeworth expansion for the distribution of $\bar{U}$ under conditions (2.8) and (2.9),

$$\sup_{B\in\mathcal{B}}\left|P(\bar{U}\in B)-\int_B \xi_{n5}(u)du\right| = O(n^{-3/2}), \tag{2.14}$$

for every class $\mathcal{B}$ of Borel sets satisfying (2.11). In (2.14),

$$\xi_{n5}(u)=\sum_{r=0}^{3} P_r(-\phi : \{\chi_{\nu n}\})(u),$$

$\{\chi_{\nu n}\}$, $1\le\nu\le 5$, are the first five cumulants of $\bar{U}$, $P_r(-\phi : \{\chi_{\nu n}\})(u)$ is the density of the finite signed measure with characteristic function $\tilde{P}_r(it : \{\chi_{\nu n}\})\cdot\exp(-t^Tt/2)$, and $\tilde{P}_r$ is the Cramér-Edgeworth polynomial. From the expression for R, we see that there exists a smooth function f_n such that $n^{1/2}R = f_n(\bar{U})$. Hence from Theorem 3.2 and Remarks 3.3 and 3.4 of Skovgaard (1981), we may show in our case that the Edgeworth expansion (2.14) may be transformed by sufficiently smooth function f_n, to yield a valid Edgeworth expansion (2.12) under conditions (2.8) and (2.9). □

3. Bartlett correction

In Section 2 we showed that the coverage errors of empirical likelihood confidence regions for β are of order n^{-1}. It is well known that part of the coverage error is due to the fact that the mean of $\ell(\beta)$ does not agree with the mean of χ^2_p, that is $E\{\ell(\beta)\}\ne p$. This disagreement can be eliminated by rescaling $\ell(\beta)$ so that it has correct mean. We demonstrate in this section that the empirical likelihood confidence region for β is Bartlett correctable. Thus, a simple empirical correction for scale can reduce the size of coverage error from order n^{-1} to order n^{-2}. For practical implementation of Bartlett correction, we give an empirical Bartlett correction.

From expansion (2.4), we may obtain an expansion for $E\{\ell(\beta)\}$ as follows,

$$E\{\ell(\beta)\} = p(1+an^{-1}) + O(n^{-2}), \tag{3.1}$$

where a is given by (2.10). The Bartlett correctability of empirical likelihood confidence regions for β is discussed in following theorem.

THEOREM 3.1. *Assume conditions* (2.8) *and* (2.9). *For any* $c_\alpha > 0$,

$$P\{\ell(\beta) < c_\alpha(1+\zeta n^{-1})\} = \alpha + O(n^{-2}),$$

where $P(\chi^2_p < c_\alpha)=\alpha$ *and* ζ *is either* a *or an* $n^{1/2}$*-consistent estimate of* a.

PROOF. We establish only the case $\zeta = a$; the case ζ is an $n^{1/2}$-consistent estimate of a can be treated in a similar way. According to Theorem 2.1, under condition (2.8) and (2.9),

$$P\{\ell(\beta) < c_\alpha(1+an^{-1})\} = P\{\chi_p^2 < c_\alpha(1+an^{-1})\} - ac_\alpha g_p\{c_\alpha(1+an^{-1})\}n^{-1} + O(n^{-3/2}). \tag{3.2}$$

Note that $g_p(v)$ is the density of χ_p^2 distribution,

$$P\{\chi_p^2 < c_\alpha(1+an^{-1})\} = P(\chi_p^2 < c_\alpha) + acg_p(c_\alpha)n^{-1} + O(n^{-2}), \tag{3.3}$$

and that

$$g_p\{c_\alpha(1+an^{-1})\} = g_p(c_\alpha) + O(n^{-1}). \tag{3.4}$$

Substituting (3.3) and (3.4) into (3.2), gives

$$P\{\ell(\beta) < c_\alpha(1+an^{-1})\} = P(\chi_p^2 < c_\alpha) + O(n^{-3/2}). \tag{3.5}$$

Moreover, by an arguments based on the oddness and evenness of polynomials in the Edgeworth expansion (see for example Barndorff-Nielsen and Hall (1988)), the $O(n^{-3/2})$ term in (3.5) is actually $O(n^{-2})$, when either a or an $n^{1/2}$-consistent estimate of a is used. Thus the theorem is proved. □

From (2.10), we know that the Bartlett correction is given by

$$a = p^{-1}\left(\frac{1}{2}\bar{\alpha}^{j\ j\ m\ m} - \frac{1}{3}\bar{\alpha}^{j\ k\ m}\bar{\alpha}^{j\ k\ m}\right),$$

where

$$\bar{\alpha}^{j\ k\ m} = n^{-1}\sum_{i=1}^{n} E(\epsilon_i^3)V_{n\ j}^{-1/2}x_i^T V_{n\ k}^{-1/2}x_i^T V_{n\ m}^{-1/2}x_i^T,$$

$$\bar{\alpha}^{j\ j\ m\ m} = n^{-1}\sum_{i=1}^{n} E(\epsilon_i^4)(x_i V_n^{-1}x_i^T)^2,$$

and $V_{n\ j}^{-1/2}$ is the j-th row of $V_n^{-1/2}$. But in practice, the Bartlett correction is unknown because V_n and the moments of ϵ_i's are unknown. In order to give an $n^{1/2}$-consistent estimate of a, we define

$$\hat{V}_n = n^{-1}\sum x_i^T x_i \hat{\epsilon}_i^2,$$

which is an estimate of covariance matrix V_n. Accordingly, we let $\hat{V}_n^{-1}$ be the inverse matrix of $\hat{V}_n$ and $\hat{V}_n^{-1/2}$ be the square root matrix of $\hat{V}_n^{-1}$. Now an estimate of a, $\hat{a}$ say, may be proposed as follows,

$$\hat{a} = p^{-1}\left(\frac{1}{2}\hat{\bar{\alpha}}^{j\ j\ m\ m} - \frac{1}{3}\hat{\bar{\alpha}}^{j\ k\ m}\bar{\alpha}^{j\ k\ m}\right), \tag{3.6}$$

where

$$\hat{\bar{\alpha}}^{j\ k\ m} = n^{-1}\sum_{i=1}^{n} \hat{\epsilon}_i^3 \hat{V}_{n\ j}^{-1/2} x_i^T \hat{V}_{n\ k}^{-1/2} x_i^T \hat{V}_{n\ m}^{-1/2} x_i^T,$$

$$\hat{\bar{\alpha}}^{j\ j\ m\ m} = n^{-1}\sum_{i=1}^{n} \hat{\epsilon}_i^4 (x_i \hat{V}_n^{-1} x_i^T)^2.$$

We can see that $\hat{\bar{\alpha}}^{j\ k\ m}$ and $\hat{\bar{\alpha}}^{j\ j\ m\ m}$ are established by replacing ϵ_i, $V_{n\ j}^{-1/2}$ in $\bar{\alpha}^{j\ k\ m}$ and $\bar{\alpha}^{j\ j\ m\ m}$ with their corresponding estimates $\hat{\epsilon}_i$ and $\hat{V}_{n\ j}^{-1/2}$, where $\hat{V}_{n\ j}^{-1/2}$ denotes the j-th row of $\hat{V}_n^{-1/2}$.

We want to prove that $\hat{a}$ is a $n^{1/2}$-consistent estimate of a. To this end we assume that

(3.7) there exist positive constants C_1, C_2 such that uniformly in n, $C_1 \le v_{pn} \le v_{1n} \le C_2$; and there exist constants $q_1, q_2 > 0$ such that $q_1 \le \inf \|x_i\| \le \sup \|x_i\| \le q_2$; and $\sup_n n^{-1} \sum E(\epsilon_i^8) \le +\infty$.

THEOREM 3.2. *Assume condition* (3.7). *Then,*

$$\hat{a} = a + O_p(n^{-1/2}).$$

Clearly Theorem 3.2 is a direct consequence of the following Lemma 3.1.

LEMMA 3.1. *Assume condition* (3.7). *Then,*

$$\hat{\bar{\alpha}}^{j\ j\ m\ m} = \bar{\alpha}^{j\ j\ m\ m} + O_p(n^{-1/2}) \quad \textit{and}$$
$$\hat{\bar{\alpha}}^{j\ k\ m} = \bar{\alpha}^{j\ k\ m} + O_p(n^{-1/2}).$$

PROOF. Since condition (3.7) means that the eigenvalues of V_n and design points x_i are all uniformly bounded respect to n, there must exist positive constants q_3 and q_4 such that for any $1 \le i, j \le n$

$$|V_{n\ j}^{-1/2} x_i^T| \le q_3, \qquad \|(n^{-1}X^T X)^{-1} x_i^T\| \le q_4. \tag{3.8}$$

Using (3.8) and Chebyshev's inequality gives us that

$$\hat{V}_n = V_n + O_p(n^{-1/2}),$$

which implies that

$$\hat{V}_n^{-1} = V_n^{-1} + O_p(n^{-1/2}) \quad \text{and} \quad \hat{V}_{n\ j}^{-1} = V_{n\ j}^{-1} + O_p(n^{-1/2}), \quad 1 \le j \le p. \tag{3.9}$$

Put

$$\bar{\alpha}_0^{j\ k\ m} = n^{-1}\sum_{i=1}^{n} \epsilon_i^3 V_{n\ j}^{-1/2} x_i^T V_{n\ k}^{-1/2} x_i^T V_{n\ m}^{-1/2} x_i^T \quad \text{and}$$

$$\bar{\alpha}_0^{j\ j\ m\ m} = n^{-1}\sum_{i=1}^{n} \epsilon_i^4 (x_i V_n^{-1} x_i^T)^2.$$

For any $M > 0$, using (3.8) and Chebyshev's inequality again, we have

$$\begin{aligned}
&P\{n^{1/2}(\bar{\alpha}_0^{j\ j\ m\ m} - \bar{\alpha}^{j\ j\ m\ m}) > M\} \\
&= P\left[n^{-1/2}\sum_{i=1}^{n}\{\epsilon_i^4 - E(\epsilon_i^4)\} V_{n\ j}^{-1/2} x_i^T V_{n\ k}^{-1/2} x_i^T V_{n\ m}^{-1/2} x_i^T > M\right] \\
&\le M^{-2} n^{-1} E\left[\sum_{i=1}^{n}\{\epsilon_i^4 - E(\epsilon_i^4)\} V_{n\ j}^{-1/2} x_i^T V_{n\ k}^{-1/2} x_i^T V_{n\ m}^{-1/2} x_i^T\right]^2 \\
&\le M^{-2} q_3^6 n^{-1} \sum E\{\epsilon_i^3 - E(\epsilon_i^3)\}^2.
\end{aligned}$$

From (3.7), we know that $n^{-1}\sum E\{\epsilon_i^3 - E(\epsilon_i^3)\}^2 < \infty$. Therefore for any $\epsilon > 0$, there exists a $M_\epsilon > 0$ such that for any $M > M_\epsilon$,

$$P\{n^{1/2}(\bar{\alpha}_0^{j\ j\ m\ m} - \bar{\alpha}^{j\ j\ m\ m}) > M\} < \epsilon$$

uniformly in n. Thus we obtain

$$\bar{\alpha}_0^{j\ j\ m\ m} = \bar{\alpha}^{j\ j\ m\ m} + O_p(n^{-1/2}). \tag{3.10}$$

In a similar way we can prove that

$$\bar{\alpha}_0^{j\ k\ m} = \bar{\alpha}^{j\ k\ m} + O_p(n^{-1/2}). \tag{3.11}$$

Now to prove Lemma 3.1 it is sufficient to show that

$$\hat{\bar{\alpha}}^{j\ k\ m} = \bar{\alpha}_0^{j\ k\ m} + O_p(n^{-1/2}), \tag{3.12}$$

$$\hat{\bar{\alpha}}^{j\ j\ m\ m} = \bar{\alpha}_0^{j\ j\ m\ m} + O_p(n^{-1/2}). \tag{3.13}$$

We only give the proof of (3.13) here, since (3.12) may be derived using the same method. By Taylor expansion, (3.9) and Schwarz inequality, we have

$$\begin{aligned}
&(x_i \hat{V}_n^{-1} x_i^T)^2 - (x_i V_n^{-1} x_i^T)^2 \\
&\quad = 2(x_i V_n^{-1} x_i^T)\{x_i(\hat{V}_n^{-1} - V_n^{-1})x_i^T\} + o_p(n^{-1/2}),
\end{aligned} \tag{3.14}$$

$$|x_i V_n^{-1} x_i^T| = |x_i V_n^{-1/2} V_n^{-1/2} x_i^T| \le \|x_i V_n^{-1/2}\| \|V_n^{-1/2} x_i^T\| \le p q_3^2. \tag{3.15}$$

Following (3.9) and (3.15), we may show that there exist positive constants C_3 and C_4 such that for $1 \le i \le n$,

$$|x_i \hat{V}_n^{-1} x_i^T| \le C_3 + |\Delta_1|, \tag{3.16}$$

$$|(x_i \hat{V}_n^{-1} x_i^T)^2 - (x_i V_n^{-1} x_i^T)^2| \le C_4 \|\hat{V}_n^{-1} - V_n^{-1}\| + |\Delta_2|, \tag{3.17}$$

where $|\Delta_i| = O_p(n^{-1/2})$ for $i = 1, 2$ and $\|A\| = \max_{i\,j} |a_{i\,j}|$ for any matrix $A = (a_{i\,j})$.

From the fact that $\hat\epsilon_i = (\beta - \beta_{\rm LS})x_i + \epsilon_i$, and using the Binomial theorem, we may show that for each integer k there exists a constant D_k such that

$$|\hat\epsilon_i^k - \epsilon_i^k| \le D_k |x_i(\beta_{\rm LS} - \beta)| \{|\epsilon_i|^{k-1} + |x_i(\beta_{\rm LS} - \beta)|^{k-1}\}. \tag{3.18}$$

Now from (3.16)–(3.18),

$$\begin{aligned}
|\hat\alpha^{j\,j\,m\,m} - \bar\alpha_0^{j\,j\,m\,m}| \quad & \\
= n^{-1} \Big| \sum [(\hat\epsilon_i^4 - \epsilon_i^4)(x_i \hat V_n^{-1} x_i^T)^2 & \\
+ \epsilon_i^4 \{(x_i \hat V_n^{-1} x_i^T)^2 - (x_i V_n^{-1} x_i^T)^2\}] \Big| & \\
\le D_4 \|\beta_{\rm LS} - \beta\| n^{-1} & \\
\cdot \sum \|x_i\| \{|\epsilon_i|^3 + \|x_i\|^3 \|\beta_{\rm LS} - \beta\|^3\} (x_i \hat V_n^{-1} x_i^T)^2 & \\
+ (C_4 \|\hat V_n^{-1} - V_n^{-1}\| + |\Delta_2|) n^{-1} \sum \epsilon_i^4 & \\
\le \Big(q_2 D_4 \|\beta_{\rm LS} - \beta\| n^{-1} \sum |\epsilon_i|^3 & \\
+ q_2^4 C_3 D_4 \|\beta_{\rm LS} - \beta\|^4 \Big) (C_3 + |\Delta_1|) & \\
+ (C_4 \|\hat V_n^{-1} - V_n^{-1}\| + |\Delta_2|) n^{-1} \sum \epsilon_i^4. &
\end{aligned} \tag{3.19}$$

Since (3.8) implies $n^{-1}\sum |\epsilon_i|^3 = O_p(1)$, and since $n^{-1}\sum \epsilon_i^4 = O_p(1)$, (3.14) can be proved from (3.9), (3.19) and the fact that $\beta_{\rm LS} = \beta + O_p(n^{-1/2})$. □

After some simplification we may show that $\hat a$ has the following explicit form

$$\hat a = p^{-1} \left[\frac{1}{2} n^{-1} \sum_{i=1}^{n} \hat\epsilon_i^4 (x_i \hat V_n^{-1} x_i^T)^2 - \frac{1}{3} n^{-2} \sum_{i,l} \{\hat\epsilon_i^3 \hat\epsilon_l^3 (x_i \hat V_n^{-1} x_l)^3\} \right].$$

In some special cases, $\hat a$ has simpler forms.

1) If $\epsilon_1, \ldots, \epsilon_n$ are i.i.d, which implies that model (1.1) is a homoscedastic regression model, then

$$\hat a = p^{-1} n \left[\frac{1}{2} \hat\mu_{4\,\epsilon} \hat\sigma^{-4} \sum \{x_i (X^T X)^{-1} x_i^T\}^2 - \frac{1}{3} \hat\mu_{3\,\epsilon}^2 \hat\sigma^{-6} \sum_{i\,l} \{x_i (X^T X)^{-1} x_l^T\} \right],$$

where $\hat\mu_{k\,\epsilon} = n^{-1}\sum_{i=1}^n \hat\epsilon_i^k$ for $k = 3, 4$, $\hat\sigma^2 = n^{-1}\sum_{i=1}^n \hat\epsilon_i^2$.

2) If $\epsilon_1, \ldots, \epsilon_n$ are i.i.d and have a symmetric distribution, then the model implies $E(\epsilon_i^3) = 0$, and we may take

$$\hat a = p^{-1} n \frac{1}{2} \hat\mu_{4\,\epsilon} \hat\sigma^{-4} \sum \{x_i (X^T X)^{-1} x_i^T\}^2.$$

4. Simulation results

In this section we use Monte Carlo simulation to examine the coverages of the empirical likelihood confidence regions proposed in previous sections. Under consideration is the following simple linear regression model:

$$Y_i = 1 + x_i^o + \epsilon_i, \qquad i = 1, \ldots, n.$$

Table 1. The data set x_i^0 for $1 \le i \le 150$.

i	x_i^0	i	x_i^0	i	x_i^0	i	x_i^0	i	x_i^0
1	1.00	31	8.90	61	14.89	91	23.80	121	37.20
2	1.40	32	9.30	62	15.01	92	24.10	122	37.60
3	1.50	33	9.70	63	15.67	93	24.20	123	37.80
4	1.70	34	9.90	64	15.71	94	24.70	124	38.30
5	2.00	35	10.00	65	15.85	95	24.98	125	38.70
6	2.30	36	10.30	66	15.97	96	25.30	126	38.90
7	2.50	37	10.40	67	16.29	97	26.00	127	39.40
8	2.67	38	10.55	68	16.38	98	27.00	128	39.80
9	3.00	39	10.70	69	16.71	99	29.00	129	40.00
10	3.30	40	11.00	70	17.00	100	29.50	130	40.50
11	3.46	41	11.23	71	17.20	101	29.90	131	40.90
12	3.50	42	11.47	72	17.35	102	30.10	132	41.10
13	4.00	43	11.66	73	17.62	103	30.60	133	41.60
14	4.40	44	11.89	74	18.00	104	31.00	134	42.00
15	4.50	45	12.09	75	18.50	105	31.20	135	42.20
16	4.90	46	12.21	76	18.50	106	31.70	136	42.70
17	5.00	47	12.43	77	19.00	107	32.10	137	43.10
18	5.20	48	12.64	78	19.33	108	32.30	138	43.30
19	5.50	49	12.91	79	19.42	109	32.80	139	43.80
20	6.00	50	13.00	80	19.78	110	33.20	140	44.20
21	6.30	51	13.23	81	19.98	111	33.40	141	44.40
22	6.70	52	13.44	82	20.02	112	33.90	142	44.90
23	6.85	53	13.51	83	20.51	113	34.30	143	45.30
24	7.00	54	13.66	84	21.00	114	34.50	144	45.50
25	7.15	55	13.79	85	21.31	115	35.00	145	46.00
26	7.30	56	13.81	86	21.79	116	35.40	146	46.40
27	7.70	57	13.81	87	22.69	117	35.60	147	46.60
28	8.00	58	14.04	88	22.81	118	36.10	148	47.10
29	8.20	59	14.19	89	23.00	119	36.50	149	47.50
30	8.50	60	14.34	90	23.40	120	36.70	150	47.70

Table 2. Estimated true coverages, from 20,000 simulations, of α-level empirical likelihood confidence regions for β. Rows headed "predic.", "uncorr.", "a" and "$\hat{a}$" give the predicted uncorrected and Bartlett-corrected coverages respectively. The figures in parentheses are 10^2 times the standard errors associated with the coverage probabilities.

(a) Normal error patterns

	ϵ_i	$N(0,1)$		$(x_i^0/2)^{1/2}N(0,1)$	
n	α	0.90	0.95	0.90	0.95
30	predic.	0.872	0.931	0.868	0.930
	uncorr.	0.839 (0.26)	0.904 (0.21)	0.833 (0.26)	0.897 (0.21)
	a	0.870 (0.24)	0.924 (0.19)	0.867 (0.24)	0.921 (0.19)
	$\hat{a}$	0.867 (0.24)	0.922 (0.19)	0.858 (0.25)	0.915 (0.20)
50	predic.	0.884	0.939	0.884	0.939
	uncorr.	0.872 (0.24)	0.928 (0.18)	0.869 (0.24)	0.927 (0.18)
	a	0.888 (0.22)	0.939 (0.17)	0.886 (0.22)	0.940 (0.17)
	$\hat{a}$	0.887 (0.22)	0.939 (0.17)	0.883 (0.23)	0.938 (0.17)
100	predic.	0.891	0.944	0.889	0.943
	uncorr.	0.890 (0.22)	0.942 (0.17)	0.888 (0.22)	0.941 (0.17)
	a	0.899 (0.21)	0.948 (0.16)	0.899 (0.21)	0.948 (0.16)
	$\hat{a}$	0.899 (0.21)	0.948 (0.16)	0.897 (0.21)	0.947 (0.16)
150	predic.	0.894	0.946	0.894	0.946
	uncorr.	0.894 (0.22)	0.946 (0.16)	0.893 (0.22)	0.948 (0.16)
	a	0.900 (0.21)	0.949 (0.15)	0.898 (0.21)	0.951 (0.15)
	$\hat{a}$	0.900 (0.21)	0.949 (0.15)	0.898 (0.21)	0.951 (0.15)

The data set x_i^0 for $1 \leq i \leq 150$ is displayed in Table 1. For sample size $n < 150$, we use the first n x_i^0 as the fixed design points. Four error patterns were considered. They are two homoscedastic error patterns $\epsilon_i = N(0,1)$ and $\epsilon_i = \mathcal{E}(1.00) - 1.00$, and two heteroscedastic error patterns $\epsilon_i = (1/2x_i^0)^{1/2}N(0,1)$ and $\epsilon_i = (1/2x_i^0)^{1/2}\{\mathcal{E}(1.00) - 1.00\}$, where $N(0,1)$ and $\mathcal{E}(1.00)$ are random variables with standard normal distribution and exponential distribution with unit mean, respectively. For each of these four error patterns we chose sample size $n = 30, 50, 100, 150$, and nominal coverage levels $\alpha = 0.90, 0.95$. The normal and exponential random variables were generated by the routines of Press *et al.* (1989).

We give in Table 2 the coverages of the uncorrected confidence regions and two corrected confidence regions based on 20,000 simulations. One of the corrected confidence regions uses the theoretical Bartlett correction a, another uses the empirical Bartlett correction $\hat{a}$. Since we know the error pattern, sample size and nominal coverage level α, we can calculate the theoretical coverages up to second order by using Edgeworth expansion in Theorem 2.1. Because the theoretical coverages can be computed without simulation, we call these "predicted coverages". We compare the "predicted coverages" with the uncorrected coverages in order to see if the theoretical results are consistent with the empirical outputs. Also,

Table 2. (continued).

(b) Exponential error patterns

	ϵ_i	$\mathcal{E}(1.00) - 1.00$		$(x_i^0/2)^{1/2}\{\mathcal{E}(1.00) - 1.00\}$	
n	α	0.90	0.95	0.90	0.95
30	predic.	0.835	0.908	0.829	0.904
	uncorr.	0.800 (0.28)	0.864 (0.24)	0.788 (0.29)	0.854 (0.25)
	a	0.863 (0.24)	0.914 (0.20)	0.847 (0.25)	0.906 (0.21)
	$\hat{a}$	0.838 (0.26)	0.895 (0.22)	0.812 (0.28)	0.874 (0.23)
50	predic.	0.863	0.926	0.863	0.926
	uncorr.	0.837 (0.26)	0.900 (0.21)	0.836 (0.26)	0.898 (0.21)
	a	0.872 (0.24)	0.927 (0.18)	0.872 (0.24)	0.924 (0.18)
	$\hat{a}$	0.860 (0.25)	0.919 (0.19)	0.853 (0.25)	0.910 (0.20)
100	predic.	0.880	0.937	0.876	0.934
	uncorr.	0.871 (0.24)	0.926 (0.18)	0.869 (0.24)	0.924 (0.18)
	a	0.893 (0.22)	0.942 (0.17)	0.892 (0.22)	0.942 (0.17)
	$\hat{a}$	0.888 (0.22)	0.938 (0.17)	0.880 (0.22)	0.932 (0.17)
150	predic.	0.888	0.942	0.886	0.941
	uncorr.	0.884 (0.23)	0.939 (0.17)	0.884 (0.23)	0.934 (0.17)
	a	0.896 (0.22)	0.947 (0.16)	0.897 (0.22)	0.945 (0.16)
	$\hat{a}$	0.895 (0.22)	0.946 (0.16)	0.895 (0.22)	0.944 (0.16)

standard errors are given for each simulated coverage and these serve as one of the criteria for comparing accuracies among different kinds of simulated coverages. The following conclusions may be drawn from the results shown in Table 2:

1) The simulated uncorrected coverages converge to the "predicted coverages" as n increases. This empirically justifies the Edgeworth expansion developed in Theorem 2.1.

2) Standard errors and absolute coverage errors both show that the Bartlett corrected confidence regions have more accurate coverage than corresponding uncorrected ones.

3) The empirically corrected confidence regions perform similarly to their theoretically corrected counterparts, except for the cases of skewed error patterns with sample sizes $n = 30$ and 50. It seems that we need a larger sample size to ensure $\hat{a}$ as a good estimate of a when the error ϵ_i's are skewed.

Comparing Table 2(a) with Table 2(b), we observe that skewness in the error patterns reduces the overall coverages. However this has little surprise for us since it has been foreseen by their corresponding "predicted coverages". In the examples considered, we see some reduction in coverages caused by heterscedasticity when n is small. Nevertheless there is no clear evidence to say generally that heterscedasticity reduces coverage accuracy when sample size is large. Our theory shows that real coverage depends on the configuration of the fixed design points and the moments of the residuals when sample size, nominal coverage level and dimensionality are all fixed.

Acknowledgements

The author thanks Professor Peter Hall for his helpful comments and encouragements. The author would also like to thank two referees and the Managing Editor for suggestions leading to improve the proof of Theorem 2.1, the presentation and simulation of the paper.

Appendix

Proof of Lemma 2.1. Since $T_n = \text{Cov}(\bar{U}) = n^{-1}\sum \text{Cov}(U_i)$ and

$$U_i = [x_i^T V_n^{-1/2}\epsilon_i, (x_i\otimes x_i)B_1^T\{\epsilon_i^2 - E(\epsilon_i^2)\}, (x_i\otimes x_i\otimes x_i)B_2^T\{\epsilon_i^3 - E(\epsilon_i^3)\}]^T,$$

we have

$$T_n = \begin{pmatrix} I_p & \Gamma_{12} & \Gamma_{13} \\ \Gamma_{12}^T & \Gamma_{22} & \Gamma_{23} \\ \Gamma_{13}^T & \Gamma_{23}^T & \Gamma_{33} \end{pmatrix}$$

where

$$\Gamma_{12} = V_n^{-1/2}\left\{n^{-1}\sum x_i^T(x_i\otimes x_i)E(\epsilon_i^3)\right\}B_1^T,$$
$$\Gamma_{13} = V_n^{-1/2}\left\{n^{-1}\sum x_i^T(x_i\otimes x_i\otimes x_i)E(\epsilon_i^4)\right\}B_2^T,$$
$$\Gamma_{22} = B_1\left[n^{-1}\sum(x_i^T\otimes x_i^T)(x_i\otimes x_i)\{E(\epsilon_i^4) - E^2(\epsilon_i^2)\}\right]B_1^T,$$
$$\Gamma_{33} = B_2\left[n^{-1}\sum(x_i^T\otimes x_i^T\otimes x_i^T)(x_i\otimes x_i\otimes x_i)\{E(\epsilon_i^6) - E^2(\epsilon_i^3)\}\right]B_2^T.$$

There exists a nonsigular matrix S such that

$$S^T T_n S = \begin{pmatrix} I_p & 0 & 0 \\ 0 & \Gamma_{22} - \Gamma_{12}^T\Gamma_{12} & 0 \\ 0 & 0 & \Gamma_{33} - \Gamma_{13}^T\Gamma_{13} - Q \end{pmatrix}$$

where

$$Q = (\Gamma_{23}^T - \Gamma_{13}^T\Gamma_{12})(\Gamma_{22} - \Gamma_{12}^T\Gamma_{12})^{-1}(\Gamma_{23} - \Gamma_{12}^T\Gamma_{13}).$$

Now Lemma 2.1 can be proved by noting that the smallest eigenvalue of $\Gamma_{22} - \Gamma_{12}^T\Gamma_{12}$ is larger than $\eta_{j_1 n} - v_{pn}^{-1}$ and the smallest eigenvalue of $\Gamma_{33} - \Gamma_{13}^T\Gamma_{13} - Q$ is larger than $\zeta_{j_2 n} - v_{pn}^{-1} - (\eta_{j_1 n} - v_{pn}^{-1})^{-1}$.

References

Barndorff-Nielsen, O. E. and Hall, P. (1988). On the level-error after Bartlett adjustment of the likelihood ratio statistics, *Biometrika*, **75**, 374–378.

Bhattacharya, R. N. and Rao, R. R. (1976). *Normal Approximation and Asymptotic Expansions*, Wiley, New York.

Chen, S. X. and Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles, *Ann. Statist.* (to appear).

Diciccio, T. J., Hall, P. and Romano, J. P. (1988). Bartlett adjustment for empirical likelihood, Tech. Report, No. 298, Dept. of Statistics, Stanford University.

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

Owen, A. (1988). Empirical likelihood ratio confidence intervals for a 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 model, *Ann. Statist.*, **19**, 1725–1747.

Press, W. H., Flannery, B. F., Teukolsky, S. A. and Vetterling, W. T. (1989). *Numerical Recipes: the Art of Scientific Computing*, Cambridge University Press, Cambridge.

Skovgaard, I. B. M. (1981). Transformation of an Edgeworth expansion by a sequence of smooth functions, *Scand. J. Statist.*, **8**, 207–217.

Wu, C. F. J. (1986). Jackknife, Bootstrap and other resampling methods in regression analysis (with discussion), *Ann. Statist.*, **14**, 1261–1350.