On Bartlett Correction of Empirical Likelihood in the Presence of Nuisance Parameters

By SONG XI CHEN

Department of Statistics and Applied Probability,

National University of Singapore, 117543, Singapore

stacsx@nus.edu.sg

AND HENGJIAN CUI

Department of Mathematics, Beijing Normal University, 100875, China

hjcui@bnu.eud.cn

Summary

Lazar and Mykland (1999) presented a case which indicated that an empirical likelihood defined by two estimating equations with a nuisance parameter was not Bartlett correctable. This was a surprising result considering the established cases of Bartlett correction for empirical likelihood. This paper shows that Bartlett correction of empirical likelihood in the presence of a nuisance parameter depends critically on the way the nuisance parameter is removed when formulating the likelihood for the parameter of interest. We show in the broad framework of estimating functions that the empirical likelihood is still Bartlett correctable if the nuisance parameter is profiled out given the value of the interested parameter.

Some key words: Bartlett correction; Empirical likelihood; Estimation equation; Nuisance parameter.

1. INTRODUCTION

Since its introduction by Owen (1988, 1990), the empirical likelihood has become an useful tool for conducting nonparametric or semiparametric inference. The empirical likelihood has been shown in a wide range of situations as outlined in Owen (2001) to admit limiting chi-square distributions, which is a nonparametric version of the Wilks' theorem in parametric likelihood. Another key property of the empirical likelihood which also resembles that of a parametric likelihood is the Bartlett correction. The Bartlett correction is a second order property which implies that a simple mean adjustment to the likelihood ratio leads to its distributional approximation to the limiting chi-square distribution improved by one order of magnitude. It is a delicate property reflecting certain internal higher order coherence of the likelihood ratio.

That the empirical likelihood is Bartlett correctable has been established for a range of situations, including Hall and La Scala (1990) for the mean parameter, DiCiccio, Hall and Romano (1991) for smooth functions of means, Chen and Hall (1993) for quantiles, Chen (1993, 1994) for linear regression and Cui and Yuan (2001) for quantiles in the presence of auxiliary information and others. Jing and Wood (1996) showed that the exponentially tilted empirical likelihood for the mean is not Bartlett correctable. This was not unexpected as the tilting breaks the delicate higher order mechanism of Bartlett correction. The most serious doubt on the Bartlett correctability of the empirical likelihood was casted in Lazar and Mykland (1999), who maintained that the empirical likelihood defined by two estimating equations in the presence of one nuisance parameter was not Bartlett correctable. The authors attributed the reason to some fundamental differences between estimating functions and smooth functions of means. However, after discussions with the authors in light of Chen (1994) and the general theory established in this paper, it is believed that the non-Bartlett correctability is due to plugging in a global maximum likelihood estimator of the nuisance parameter when constructing the empirical likelihood ratio for the parameter of interest.

In this paper, we consider the Bartlett property in a broader situation where there are r estimating equations and the dimension of the nuisance parameter is p ($p < r$), which is within the framework of the empirical likelihood for generalized estimating equations introduced in Qin and Lawless (1994). It is found that if the nuisance parameter is profiled out given the parameter of interest the empirical likelihood is still Bartlett correctable. This indicates that the Bartlett correctability of the empirical likelihood is crucially dependent on the method of the nuisance parameter removal when formulating the likelihood for the parameter of interest rather than any fundamental differences between estimating equations and the smooth function of means.

The paper is organized as follows. In Section 2, we discuss an example that highlights the effects of the nuisance parameter removal on the Bartlett correctability of the empirical likelihood. In Section 3, we consider a general class of empirical likelihood defined by a set of estimating equations in the presence of nuisance parameters of any dimensionality. The Bartlett correction of the empirical likelihood are presented in Section 4. A general discussion is given in Section 5.

2. An Example

In this section we consider a simple bivariate example with one parameter of interest and one nuisance parameter which is within the framework of Lazar and Mykland (1999). The purpose is to highlight that different ways of removing the nuisance parameter may lead to different conclusions on the Bartlett correctability.

Let $X = (Z, Y)^T$ be a bivariate random vector with a distribution F, and $\{X_i = (Z_i, Y_i)\}_{i=1}^n$ be an independent and identically distributed sample from F. Let $\theta_0 = E(Z) = 0$ and $\psi = E(Y)$. Here θ is the parameter of interest and ψ is the nuisance parameter. There are two estimating equations: $g^1(X, \theta, \psi) = Z - \theta$ and $g^2(X, \theta, \psi) = Y - \psi$. We further assume that $Cov(X) = I_2$, the identity matrix in R^2, $E(Z^3) \neq 0$ and X admits conditions (2) given in the next section.

Using the same notations as Lazar and Mykland (1999), let

$$U_. = \left(\sum Z_i, \sum (Y_i - \psi), 0\right)^T, \qquad U_{..} = \begin{pmatrix} -\sum Z_i^2 & -\sum Z_i(Y_i - \psi) & 0 \\ -\sum Z_i(Y_i - \psi) & -\sum (Y_i - \psi)^2 & -n \\ 0 & -n & 0 \end{pmatrix},$$

$$\begin{aligned} \kappa_r &= n^{-1}E(U_r),\ \kappa_{rs} = n^{-1}E(U_{rs}),\ \kappa_{r,s} = n^{-1}Cov(U_r, U_s), \\ \kappa_{r,ts} &= n^{-1}Cov(U_r, U_{ts}) \quad \text{and} \quad \kappa_{rs,tu} = n^{-1}Cov(U_{rs}, U_{tu}). \end{aligned}$$

It is easy to check that $\kappa_r =: EU_r = 0$,

$$(\kappa_{rs}) = \begin{pmatrix} -1 & 0 & 0 \\ 0 & -1 & -1 \\ 0 & -1 & 0 \end{pmatrix}, \quad (\kappa_{r,s}) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (\kappa^{rs}) = (\kappa_{rs})^{-1} = \begin{pmatrix} -1 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 1 \end{pmatrix}$$

and $(\kappa^{r,s}) = (\kappa_{r,s})^{+}$, where $^{+}$ stands for the Moore-Penrose inverse.

Define $\beta^i_{rs} = \kappa^{i,\alpha}\kappa_{\alpha,rs} = \kappa_{i,rs}$, $\beta^i_{rst} = \kappa^{i,\alpha}\kappa_{\alpha,rst} = \kappa_{i,rst}$ for $i = 1, 2, 3$, and

$$U_r = V_r,\ U_{rs} = V_{rs} + \beta^i_{rs}V_i \quad \text{and} \quad U_{rst} = V_{rst} + \beta^i_{rs}V_{it}[3] + \beta^i_{rst}V_i.$$

Moreover let v denote the cumulants of V as in Lazar and Mykland (1999) and follow the notations of McCullagh (1987). It can be verified that $(v^{rs}) = (\kappa^{rs})$, $v^{11} = v^{23} = -v^{33} = -1$, $v^{13} = 0$ and $v_{111} = -E(Z^3)$, and $v_{113} = v_{133} = v_{13,13} = v_{1133} = 0$. Hence $b_0 = b_1 = c = 0$ in equation (5) of Lazar and Mykland (1999). Moreover,

$$\tilde{\omega}_4 = \tfrac{37}{18}n^{-1}v_{111}v_{111}v^{11}v^{11}v^{11} + O(n^{-2}) = -\tfrac{37}{18}n^{-1}\{E(Z^3)\}^2 + O(n^{-2}),\ \ \rho_4 = -\tfrac{37}{18}\{E(Z^3)\}^2 \neq 0.$$

This means that the fourth order cumulants of the sign root of the empirical likelihood for θ_0 are not at the order of n^{-4}, and thus the empirical likelihood would not be Bartlett correctable.

On the other hand, if ψ is profiled out given θ_0, the empirical likelihood ratio for $\theta_0 = 0$ is

$$\ell(\theta_0) = 2\sum_{i=1}^{n} \log\{1 + \tilde{\lambda}^1 g^1(Z_i, 0, \tilde{\psi}) + \tilde{\lambda}^2 g^2(Z_i, 0, \tilde{\psi})\} \tag{1}$$

where $\tilde{\lambda}^1, \tilde{\lambda}^2, \tilde{\psi}$ are the solutions of

$$\begin{cases} \displaystyle\sum_{i=1}^{n} \frac{Z_i}{1 + \lambda^1 Z_i + \lambda^2(Y_i - \psi)} = 0, \\ \displaystyle\sum_{i=1}^{n} \frac{Y_i - \psi}{1 + \lambda^1 Z_i + \lambda^2(Y_i - \psi)} = 0 \quad \text{and} \\ \displaystyle\sum_{i=1}^{n} \frac{-\lambda^2}{1 + \lambda^1 Z_i + \lambda^2(Y_i - \psi)} = 0. \end{cases}$$

From the third equation, $\tilde{\lambda}^2 = 0$, $\tilde{\lambda}^1$ satisfies

$$\sum_{i=1}^{n} \frac{Z_i}{1+\tilde{\lambda}^1 Z_i} = 0 \quad \text{and} \quad \tilde{\psi} = \frac{\sum_{i=1}^{n} Y_i/(1+\tilde{\lambda}^1 Z_i)}{\sum_{i=1}^{n} 1/(1+\tilde{\lambda}^1 Z_i)}.$$

Therefore, $\ell(\theta_0) = 2\sum_{i=1}^{n} \log(1+\tilde{\lambda}^1 Z_i)$. This is essentially the empirical likelihood for the mean of Z, which is the first known case of Bartlett correctable for the empirical likelihood given by Hall and La Scala (1990).

These seemingly conflicting results are due to different methods used to remove ψ in the construction of $\ell(\theta_0)$. Lazar and Mykland (1999) used the global maximum empirical likelihood estimator of ψ whereas in the second formulation given in (1) ψ is profiled out given the value of θ_0. The later approach is more commonly used and more natural for removing nuisance parameters. It is noted that Chen (1994) established Bartlett correction of empirical likelihood for a simple linear regression coefficient while treating the other coefficient as a nuisance parameter. There were two estimating equations too although the data are independent but not identically distributed.

Now the question is "Does this profiling out the nuisance parameter given the parameter of interest guarantee Bartlett correction for the empirical likelihood in general ?". The rest of the paper is devoted to address this question.

3. Empirical likelihood with nuisance parameters

Consider a random vector X with unknown distribution function F which depends on a r-dimensional parameter $(\theta, \psi) \in R^{r-p} \times R^p$. Here the interest is on the parameter θ while treating ψ as a p-dimensional nuisance parameter. We assume that the parameter (θ, ψ) is defined by r $(r > p)$ functionally unbiased estimating equations $g^j(x, \theta, \psi)$, $j = 1, 2, \cdots, r$ such that $E\{g^j(X_1, \theta_0, \psi_0)\} = 0$ where (θ_0, ψ_0) is the true parameter value. In particularly, we define

$$g(X, \theta, \psi) = \left(g^1(X, \theta, \psi), g^2(X, \theta, \psi), \cdots, g^r(X, \theta, \psi)\right)^T.$$

Assume that $\{X_1, X_2, \cdots, X_n\}$ is an independent and identically distributed sample drawn from F. Let $V = Cov\{g(X_i, \theta_0, \psi_0)\}$ and we assume the following regularity conditions:

(i) V is a $r \times r$ positive definite matrix and the rank of $E\{\partial g(X, \theta_0, \psi_0)/\partial\psi\}$ is p; (2)

(ii) For any j, $1 \le j \le p$, all the fourth order partial derivatives of $g^j(x, \theta_0, \psi)$ with respect to ψ are continuous in a neighborhood of θ_0 and are bounded by some integrable function $G(x)$ in the neighborhood;

(iii) $E\|g(X, \theta_0, \psi_0)\|^{15} < \infty$ and the characteristic function of $g(X, \theta_0, \psi_0)$ satisfies the Cramér condition: $\limsup_{|t|\to\infty} |E[\exp\{it^T g(X, \theta_0, \psi_0)\}]| < 1$.

To simplify derivations, let us first rotate the original estimating functions by defining

$$w_i(\theta, \psi) =: TV^{-1/2} g(X_i, \theta, \psi),$$

where T is a $r \times r$ orthogonal matrix such that

$$TV^{-1/2} E\Big\{\frac{\partial g(X, \theta_0, \psi_0)}{\partial \psi}\Big\} U = \begin{pmatrix}\Lambda & 0\end{pmatrix}^T_{r\times p}$$

$U = (u^{kl})_{p\times p}$ is an orthogonal matrix and $\Lambda = diag(\lambda_1, \cdots, \lambda_p)$ is a non-singular $p \times p$ diagonal matrix. Furthermore, let us define $\Omega = \left(\omega^{kl}\right)_{p\times p} = U\Lambda^{-1}$.

Let $p_1, \cdots, p_n$ be non-negative weights allocated to the observations. The empirical likelihood for the parameter (θ, ψ) is

$$L(\theta, \psi) = \prod_{i=1}^{n} p_i$$

subject to $\sum p_i = 1$ and the constraints $\sum p_i w_i(\theta, \psi) = 0$. Let $\ell(\theta, \psi) = -2\log\{n^n L(\theta, \psi)\}$ be the log empirical likelihood ratio. Standard derivations in the empirical likelihood show that

$$\ell(\theta, \psi) = 2\sum_{i=1}^{n} \log\{1 + \lambda^T(\theta, \psi) w_i(\theta, \psi)\},$$

where $\lambda(\theta, \psi)$ satisfies:

$$n^{-1}\sum_{i=1}^{n} \frac{w_i(\theta, \psi)}{1 + \lambda^T(\theta, \psi) w_i(\theta, \psi)} = 0. \tag{3}$$

To obtain the empirical likelihood ratio at θ_0, we profile out the nuisance parameter ψ. To simplify notation, let us write $w_i(\psi) = w_i(\theta_0, \psi)$ and let $\tilde{\psi} =: \tilde{\psi}(\theta_0)$ be the minima of $\ell(\theta_0, \psi)$ given $\theta = \theta_0$ and $\tilde{\lambda} = \lambda(\theta_0, \tilde{\psi})$ be the solution of (3) at $(\theta_0, \tilde{\psi})$. Let $(\hat{\theta}, \hat{\psi})$ be the maximum empirical likelihood estimate of parameter (θ, ψ). Since the number of estimating functions equal to the dimension of parameter (θ, ψ), then $\ell(\hat{\theta}, \hat{\psi}) = 0$. This means that the log empirical likelihood ratio for θ_0 is just

$$r(\theta_0) =: \ell(\theta_0, \tilde{\psi}(\theta)) = 2\sum_{i=1}^{n} \log\{1 + \tilde{\lambda}^T w_i(\tilde{\psi})\}. \tag{4}$$

In order to develop an expansion for $r(\theta_0)$, we need to derive expansions for $\tilde{\lambda}$ and $\tilde{\psi}$ first. We notice from Qin and Lawless (1994) that $(\tilde{\lambda}, \tilde{\psi})$ are the solutions of

$$Q_{1n}(\lambda, \psi) = n^{-1}\sum_{i=1}^{n} \frac{w_i(\psi)}{1 + \lambda^T w_i(\psi)} = 0 \tag{5}$$

$$Q_{2n}(\lambda, \psi) = n^{-1}\sum_{i=1}^{n} \frac{(\partial w_i(\psi)/\partial\psi)^T \lambda}{1 + \lambda^T w_i(\psi)} = 0. \tag{6}$$

Let $\eta = (\lambda^T, \psi^T)^T$, $\eta_0 = (0, \psi_0)$,

$$Q(\eta) = \begin{pmatrix} Q_{1n}(\eta) \\ Q_{2n}(\eta) \end{pmatrix} \quad \text{and} \quad S = E\frac{\partial Q(0, \psi_0)}{\partial \eta} = \begin{pmatrix} -I & S_{12} \\ S_{21} & 0 \end{pmatrix},$$

where $S_{21} = U(\Lambda, 0)$ and $S_{12} = S_{21}^T$. To facilitate easy expressions, we standardize Q to $\Gamma(\eta) = S^{-1}Q(\eta)$. Let $w_i^j(\psi)$ and $\Gamma^j(\eta)$ denote respectively the j-th component of $w_i(\psi)$ and $\Gamma(\eta)$. The following $\alpha - A$ system of notations was first used by DiCiccio, Hall and Romano (1988):

$$\alpha^{j_1 \dots j_k} = E\{w^{j_1}(\psi_0) \dots w^{j_k}(\psi_0)\}$$

$$A^{j_1 \dots j_k} = n^{-1}\sum_{i=1}^{n} w^{j_1}(\psi_0) \dots w^{j_k}(\psi_0) - \alpha^{j_1 \dots j_k}.$$

We also need to define

$$\beta^{j, j_1 \dots j_k} = E\Big\{\frac{\partial^k \Gamma^j(0, \psi_0)}{\partial \eta_{j_1} \dots \partial \eta_{j_k}}\Big\}, \quad B^{j, j_1 \dots j_k} = \frac{\partial^k \Gamma^j(0, \psi_0)}{\partial \eta_{j_1} \dots \partial \eta_{j_k}} - \beta^{j, j_1 \dots j_k}$$

and

$$\gamma^{j,j_1...j_l;k,k_1...k_m;...;p,p_1...p_n} = E\Big\{\frac{\partial^l w_i^j(\psi_0)}{\partial\psi^{j_1}...\partial\psi^{j_l}}\frac{\partial^m w_i^k(\psi_0)}{\partial\psi^{k_1}...\partial\psi^{k_m}}\cdots\frac{\partial^n w_i^p(\psi_0)}{\partial\psi^{p_1}...\partial\psi^{p_n}}\Big\}$$

$$C^{j,j_1...j_l;k,k_1...k_m;...;p,p_1...p_n} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^l w_i^j(\psi_0)}{\partial\psi^{j_1}...\partial\psi^{j_l}}\frac{\partial^m w_i^k(\psi_0)}{\partial\psi^{k_1}...\partial\psi^{k_m}}\cdots\frac{\partial^n w_i^p(\psi_0)}{\partial\psi^{p_1}...\partial\psi^{p_n}} - \gamma^{j,j_1...j_l;k,k_1...k_m;...;p,p_1...p_n}.$$

4. MAIN RESULTS

We only provide a brief sketch about the steps taken to reach the Bartlett correction for this general case. A full account on the technical details are available in Chen and Cui (2002).

We need to first develop an expansion to the empirical likelihood ratio $\ell(\theta_0)$. Derivations given in Chen and Cui (2002) show that the Langrage multiplier and the nuisance parameter at θ_0 admit the following expansions:

$$\begin{aligned}\tilde{\lambda}^j &= -B^j + B^{j,q}B^q - \tfrac{1}{2}\beta^{j,uq}B^uB^q - B^{j,u}B^{u,q}B^q + \tfrac{1}{2}\beta^{u,qs}B^{j,u}B^qB^s + \beta^{j,uq}B^{u,s}B^sB^q \\ &- \tfrac{1}{2}\beta^{j,uq}\beta^{u,st}B^sB^tB^q - \tfrac{1}{2}B^{j,uq}B^uB^q + \tfrac{1}{6}\beta^{j,uqs}B^uB^qB^s + O_p(n^{-2}) \qquad (7)\end{aligned}$$

and

$$\begin{aligned}\tilde{\psi}^k &= -B^{r+k} + B^{r+k,q}B^q - \tfrac{1}{2}\beta^{r+k,uq}B^uB^q - B^{r+k,u}B^{u,q}B^q + \tfrac{1}{2}\beta^{u,qs}B^{r+k,u}B^qB^s \\ &+ \beta^{r+k,uq}B^{u,s}B^sB^q - \tfrac{1}{2}\beta^{r+k,uq}\beta^{u,st}B^sB^tB^q - \tfrac{1}{2}B^{r+k,uq}B^uB^q + \tfrac{1}{6}\beta^{r+k,uqs}B^uB^qB^s \\ &+ O_p(n^{-2}) \qquad (8)\end{aligned}$$

for $j \in \{1, \cdots, r\}$, $k \in \{1, \cdots, p\}$ and $q, s, t, u \in \{1, \cdots, r+p\}$. Here and in the rest of the paper, we use the tensor notation where if a superscript is repeated a summation over that superscript is understood; see McCullagh (1987) for a systematic description.

Substitute the above expansions to (4),

$$n^{-1}\ell(\theta_0) = A^{p+a}A^{p+a} - A^{p+a\ p+b}A^{p+a}A^{p+b} - 2\omega^{kl}C^{p+a,k}A^{p+a}A^l$$

$$
\begin{aligned}
&+ \quad 2\gamma^{p+a;p+b,k}\omega^{kl}A^{p+a}A^{p+b}A^{l} + \tfrac{2}{3}\alpha^{p+a\ p+b\ p+c}A^{p+a}A^{p+b}A^{p+c} \\
&+ \quad \gamma^{p+a,kl}\omega^{km}\omega^{ln}A^{p+a}A^{m}A^{n} + A^{ji}B^{i,q}B^{q}B^{j}[2,i,j] - B^{j,u}B^{j,q}B^{u}B^{q} \\
&- \quad 2C^{j,k}B^{j,q}B^{r+k}B^{q} + \gamma^{j,kl}B^{r+k}B^{r+l}B^{j,q}B^{q} + 2\gamma^{j,kl}B^{j}B^{r+l}B^{r+k,q}B^{q} \\
&- \quad 2\gamma^{j;i,l}(B^{j}B^{i}B^{r+l,q}B^{q} + B^{r+l}B^{i}B^{j,q}B^{q}[2,j,i]) + 2\alpha^{jih}B^{j}B^{i}B^{h,q}B^{q} \\
&+ \quad (\tfrac{1}{2}\beta^{j,uq}\beta^{r+k,st}\gamma^{j,k} - \tfrac{1}{4}\beta^{j,uq}\beta^{j,st})B^{u}B^{q}B^{s}B^{t} - \tfrac{1}{2}\gamma^{j,kl}\beta^{j,uq}B^{u}B^{q}B^{r+l}B^{r+k} \\
&+ \quad (\gamma^{i;j,k}\beta^{i,uq} + \gamma^{j;i,k}\beta^{i,uq} - \gamma^{j,kl}\beta^{r+l,pq})B^{u}B^{q}B^{j}B^{r+k} + 2\gamma^{j;i;h,k}B^{j}B^{i}B^{h}B^{r+k} \\
&- \quad (\gamma^{j;i,lk} + \gamma^{j,l;i,k})B^{j}B^{i}B^{r+l}B^{r+k} + \tfrac{1}{3}\gamma^{j,klm}B^{j}B^{r+k}B^{r+l}B^{r+m} \\
&- \quad \tfrac{1}{2}\alpha^{jihg}B^{j}B^{i}B^{h}B^{g} + (\gamma^{j;i,l}\beta^{r+l,uq} - \alpha^{jih}\beta^{h,uq})B^{j}B^{i}B^{u}B^{q} \\
&- \quad C^{j,kl}B^{j}B^{r+k}B^{r+l} + 2C^{j;i,l}B^{j}B^{i}B^{r+l} - \tfrac{2}{3}A^{jih}B^{j}B^{i}B^{h} + O_p(n^{-5/2}) \qquad (9)
\end{aligned}
$$

where $h, i, j \in \{1, \cdots, r\}$, $k, l, m \in \{1, \cdots, p\}$, $a, b, c \in \{1, \cdots, r-p\}$ and $q, s, t, u \in \{1, \cdots, r+p\}$. When $p = 0$, namely there is no nuisance parameter, the above expansion takes the standard form as given in DiCiccio, Hall and Romano (1991) by noting that $S = -I$, $B^j = -A^j$, $B^{i,j} = A^{ij}$, $\beta^{j,j_1j_2} = -2\alpha^{jj_1j_2}$, and all the γs and Cs are zero.

Let $R = R_1 + R_2 + R_3$ be a signed root decomposition of $n^{-1}\ell(\theta_0)$ such that

$$n^{-1}\ell(\theta_0) = R^q R^q + O(n^{-5/2})$$

where $R_j = O_p(n^{-j/2})$ for $j = 1, 2$ and 3. Clearly, R_1 and R_2 can be determined from the terms of $O_p(n^{-1})$ and $O_p(n^{-3/2})$ respectively in (9). Specifically, for $a, b, c, d, e \in \{1, \cdots, r-p\}$ and $l, k, m, n, o, v, m', n' \in \{1, \cdots, p\}$

$$
\begin{aligned}
R_1^q &= R_2^q = 0 \quad \text{for } q \le p, \quad R_1^{p+a} = A^{p+a} \quad \text{and} \\
R_2^{p+a} &= -\tfrac{1}{2}A^{p+a\ p+b}A^{p+b} - \omega^{kl}C^{p+a,k}A^{l} + \gamma^{p+a;p+b,k}\omega^{kl}A^{p+b}A^{l} \\
&+ \tfrac{1}{3}\alpha^{p+a\ p+b\ p+c}A^{p+b}A^{p+c} + \tfrac{1}{2}\gamma^{p+a,kl}\omega^{km}\omega^{ln}A^{m}A^{n}.
\end{aligned}
$$

The expression for R_3 is obtained after removing terms induced by $R_1^{p+a}R_1^{p+a}$ and $R_2^{p+a}R_2^{p+a}$ from (9); see Chen and Cui (2002) for details.

The key in checking on the Bartlett correctability of the empirical likelihood is to examine if the third and the fourth order joint cumulants of R are at the orders of n^{-3} and n^{-4} respectively. This is the path taken by DiCiccio, Hall and Romano (1991), Chen (1994), Jing and Wood (1996) and Lazar and Mykland (1999). A formal establishment of the Bartlett correction can be then made by developing Edgeworth expansions for the empirical likelihood ratio under condition (2). After a rather length expedition as documented in Chen and Cui (2002), it is found that the joint third and fourth order cumulants of R are indeed at the orders of n^{-3} and n^{-4} respectively. Hence, the empirical likelihood is still Bartlett correctable in this general case of estimating equations with nuisance parameters.

6. DISCUSSION

The established Bartlett correction provides the theoretical justification for an empirical Bartlett correction to the confidence regions. Let c_α be the upper α quantile of the χ^2_{r-p} distribution. Based on the Wilks' theorem, a $1-\alpha$ level confidence region for θ is $I_\alpha = \{\theta | \ell(\theta) \le c_\alpha\}$. It may be shown by developing Edgeworth expansions using the derived cumulants given in Chen and Cui (2002) that the coverage error of I_α is of $O(n^{-1})$. As $\ell(\theta_0)$ is Bartlett correctable, it may be shown under condition (2) that

$$P\{l(\theta_0) < c_\alpha(1 + n^{-1}B_c)\} = 1 - \alpha + O(n^{-2}),$$

where B_c is the Bartlett factor whose expression is given in Chen and Cui (2002). The rather complicated form of B_c means that the direct plug-in method would not work for its estimation. We propose a method based on the following bootstrap procedure:

Step 1: generate bootstrap resamples of size n by sampling with replacement from the original sample $\{X_i\}_{i=1}^n$; compute the empirical likelihood ratio $\ell^\star(\hat\theta)$ based on the resamples, where $\hat\theta$ is the global maximum empirical likelihood estimator of θ based on the original sample.

Step 2: repeat Step 1 B times to obtain $\ell^{\star 1}(\hat\theta), \cdots, \ell^{\star B}(\hat\theta)$ and $B^{-1}\sum_{b=1}^B \ell^{b\star}(\hat\theta)$, which is the bootstrap estimate of $E\{\ell(\hat\beta)\}$.

The bootstrap estimator to $1+n^{-1}B_c$ is $\tau = (r-p)^{-1}B^{-1}\sum_{b=1}^{B}\ell^{b\star}(\hat{\beta})$, which can be used to construct $I_{\alpha,bc} = \{\theta|\ell(\theta) \le \tau c_\alpha\}$, the Bartlett corrected confidence region. It can be shown based on the same Edgeworth expansion mentioned earlier that the coverage error of this Bartlett corrected region is $O(n^{-3/2})$, which is one order of magnitude smaller than that of I_α.

Acknowledgement

The authors would like to thank Professor Nicole Lazar for informative discussions, and an anonymous reviewer and the Editor for comments and suggestions which improve the presentation of the paper. The project is supported by an Academic Research Grant of the National University of Singapore (R-155-000-018-112), the NSFC (10071009) and RFDP (20020027010) of China.

References

Chen S. X. (1993). On the coverage accuracy of empirical likelihood 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.

Chen, S. X. & Cui, H. J. (2002). On Bartlett correction of empirical likelihood in the presence of nuisance parameters. Technical report, Department of Statistics and Applied Probability, National University of Singapore.

Chen, S. X. & Hall, P. (1993). Smoothed empirical likelihood confidence intervals for quantiles. *Ann. Statist.* **21**, 1166-1181.

Cui H. J. & Yuan X. J. (2001). Smoothed empirical likelihood confidence interval for quantile in the partial symmetric auxiliary information. *J. Sys. Sci. and Math. Scis.* 21(2), 172-181.

DiCiccio, T. J., Hall, P. & Romano, J. P. (1988). Empirical likelihood is Bartlett correctable. Technical Report, Department of Statistics, Stanford University.

DiCiccio, T. J., Hall, P. & Romano, J. P. (1991). Empirical likelihood is Bartlett correctable. *Ann. Statist.* **19**, 1053-1061.

DiCiccio, T. J. & Romano, J. P. (1989). On adjustments based on the signed root of the empirical likelihood ratio statistic. *Biometrika* **76**, 447-56.

Jing, B. Y. & Wood, A. T. A. (1996). Exponential empirical likelihood is not Bartlett correctable. *Ann. Statist.* **24** 365-369.

Hall, P. and La Scala, B. (1990). Methodology and algorithms of empirical likelihood. *Internat. Statist. Rev.* **58**, 109–127.

Lazar, N. A. & Mykland, P. A. (1999). Empirical likelihood in the presence of nuisance parameters. *Biometrika* **86**, 203-211.

McCullagh, P. (1987). *Tensor Methods in Statistics.* London: Chapman and Hall.

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

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

Owen, A. B. (2001). *Empirical Likelihood.* Boca Raton: Chapman and Hall/CRC.

Qin J. & Lawless, J. (1994). Empirical likelihood and general estimation equations. *Ann. Statist.* **22**, 300-325.