On the Second Order Properties of Empirical Likelihood with Moment Restrictions

Song Xi Chen[a*] and Hengjian Cui[b]

[a*] *Department of Statistics, Iowa State University*
Ames, IA 50011-1210, USA
[b] *Department of Mathematics, Beijing Normal University*
Beijing, 100875, China

Summary This paper considers the second order properties of empirical likelihood for a parameter defined by moment restrictions, which is the framework operated upon by the Generalized Method of Moments. It is shown that the empirical likelihood defined for this general framework still admits the delicate second order property of Bartlett correction, which represents a substantial extension of all the established cases of Bartlett correction for the empirical likelihood. An empirical Bartlett correction is proposed, which is shown to work effectively in improving the coverage accuracy of confidence regions for the parameter.

Keywords: Bartlett correction, Coverage accuracy, Empirical likelihood, Generalized Method of Moments.

Running Title: Empirical Likelihood with Moment Restrictions.

1. INTRODUCTION

Generalized Method of Moments (GMM) introduced by Hansen (1982) is an important inferential framework in econometric studies. GMM is based on, upon given a model, some known functions $g(X, \theta)$ of a random observation $X \in R^d$ and an unknown parameter $\theta \in R^p$, where $g : R^{d+p} \to R^r$, such that $E\{g(X, \theta)\} = 0$ which constitutes moment restrictions on the relationship between X and θ. The power of GMM is in its allowing $r \geq p$, namely the number of moment restrictions (instruments) can be larger than the number of parameter, which leads to a full exploration of inference opportunities provided by the given model. There is a vast pool of literatures on GMM. Here we only cite the latest reviews of Andrews (2002), Brown and Newey (2002), Imbens (2002) and Hansen and West (2002).

Empirical likelihood (EL) introduced by Owen (1988) is a computer-intensive statistical method that facilitates a likelihood-type inference in a nonparametric or semiparametric setting. It is closely connected to the bootstrap as the EL effectively carries out the resampling implicitly. On certain aspects of inference, EL is more attractive than the bootstrap, for instance its ability of internal studentizing so as to avoid explicit variance estimation and producing confidence regions with natural shape and orientation; see Owen (2001) for an overview of EL. A key property of EL is that the log EL ratio is asymptotically chi-squared distributed, which resembles the Wilks' theorem in parametric likelihood. The Wilks' theorem was established in the original proposal of Owen (1988) for the means, in Hall and La Scala (1990) for smoothed function of means, Qin and Lawless (1994) for parameters defined by moment restrictions and Kitamura (1997) for weakly dependence observations.

There have been comprehensive studies of EL in the context of GMM in econometrics. Imbens (1997) shows that the maximum EL estimator of θ is a one-step variation of the two-stage GMM estimator in the over-identified case of $r > p$, and achieves the same asymptotic efficiency as the two-stage estimator. Testing is considered in Kitamura (2001) for moments restrictions, and Tripathi and Kitamura (2002) for conditional moment restrictions. Estimation and testing with conditional moment restrictions are studied in Donald, Imbens and Newey (2003) and Kitamura, Tripathi and Ahn (2002). They found that EL posses the attractive features of avoiding estimating optimal instruments and achieving asymptotic pivotalness. Tilted EL and other variations are studied in Kitamura and Stutzer (1997), Smith (1997) and Newey and Smith (2004). In particular, Newey and Smith (2004) find that the EL estimator is favorable in terms of the bias and the second order variance in comparison with the GMM estimator.

Another key property of the EL is Bartlett correction, which is a delicate second order property implying that a simple mean adjustment to the likelihood ratio can improve the approximation to the limiting chi-square distribution by one order of magnitude and hence can be used to enhance the coverage accuracy of likelihood-based confidence regions. In the context of testing hypotheses, the Bartlett correction reduces the errors between the nominal and actual significant levels of an EL test. Bartlett correction has been established for EL by DiCiccio, Hall and Romano (1991) for smoothed functions of means and Chen (1993, 1994) for linear regression. Baggerly (1998) shows that EL is the only member within the Cressie-Read power divergence family that is Bartlett correctable. Jing and Wood (1996) reveal that the exponentially tilted EL for the means is not Bartlett correction as the tilting alters the delicate second order mechanism of EL.

In this paper we show that the EL with moment restrictions is Bartlett correctable. The finding represents a substantial extension of all the established cases of Bartlett correction, which all assume $r = p$ corresponding to the just-identified case in GMM. The establishment of the Bartlett correction for the just-identified case is a lot easier as the log maximum EL takes a constant value $-n \log(n)$ (n is the sample size). However, in the over-identified case the maximum EL is no longer a constant, rather it introduces many extra terms into the log EL ratio and makes the study of Bartlett correction far more challenging as can be seen from the analysis carried out in this paper. The establishment of Bartlett correction in this general case indicates that EL inherits the delicate second order mechanism of the parametric likelihood in a much wider situation. This together with the findings of Imbens

(1997), Kitamura (2001) and Newey and Smith (2004) and others suggests that the EL is an attractive inferential tool in the context of moment restrictions. The establishment of the Bartlett correction leads to a practical Bartlett correction, which is confirmed to work effectively for coverage restoration in our simulation studies reported in Section 4.

The paper is organized as follows. Section 2 provides an expansion for the log EL ratio for parameters defined by moment restrictions. Bartlett correction and coverage errors assessment of EL confidence regions are investigated in Section 3. Simulation results are reported in Section 4, followed by a general discussion in Section 5. All technical details are left in the appendix.

2. EL FOR GENERALIZED MOMENT RESTRICTIONS

Let $X_1, X_2, \cdots, X_n$ be $d-$dimensional independent and identically distributed random sample whose distribution depends on a $p-$dimensional parameter θ which takes values in a compact parameter space $\Theta \subseteq R^p$. The information about θ is summarized in the form of $r \geq p$ unbiased moment restrictions $g^j(x,\theta)$, $j = 1, 2, \cdots, r$, such that $E[g^j(X_1, \theta_0)] = 0$ for a unique θ_0, which is the true value of θ. Let

$$g(X,\theta) = (g^1(X,\theta), g^2(X,\theta), \cdots, g^r(X,\theta))^T \quad \text{and} \quad V = Var\{g(X_1,\theta_0)\}.$$

We assume the following regularity conditions:

(2.1) (i) V is a $r \times r$ positive definite matrix and the rank of $E[\partial g(X_1,\theta_0)/\partial\theta]$ is p;

(ii) For any j, $1 \leq j \leq p$, all the partial derivatives of $g^j(x,\theta)$ up to the third order with respect to θ are continuous in a neighborhood of θ_0 and are bounded by some integrable functions respectively in the neighborhood;

(iii) $\limsup_{|t|\to\infty} |E[\exp\{it^T g(X_1,\theta_0)\}]| < 1$ and $E\|g(X_1,\theta_0)\|^{15} < \infty$.

Conditions (i) and (ii) are standard requirements for establishing the Wilks' theorem and higher order Taylor expansions of the EL ratio. The first part of the condition (iii) is just the Cramér's condition on the characteristic function of $g(X,\theta_0)$. It and the requirement that $E\|g(X,\theta_0)\|^{15}<\infty$ are required for establishing the Edgeworth expansion.

To facilitate simpler expressions, we transform $g(X_i,\theta)$ to $w_i(\theta)=TV^{-1/2}g(X_i,\theta)$ where T is a $r\times r$ orthogonal matrix such that

$$TV^{-1/2}E\Big(\frac{\partial g(X_i,\theta_0)}{\partial\theta}\Big)U=\Big(\ \Lambda,0\ \Big)^{\tau}_{r\times p}. \tag{2.2}$$

Here $U=\Big(u^{kl}\Big)_{p\times p}$ is an orthogonal matrix and $\Lambda=diag(\lambda_1,\cdots,\lambda_p)$ is non-singular.

Let $p_1,p_2,\cdots,p_n$ be non-negative weights allocated to the observations. The EL for θ as proposed in Qin and Lawless (1994) is $L(\theta)=\prod_{i=1}^n p_i$ subject to $\sum_{i=1}^n p_i=1$ and $\sum_{i=1}^n p_iw_i(\theta)=0$. Let $\ell(\theta)=-2\log\{L(\theta)/n^n\}$. Standard derivations in EL show

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

where $\lambda=\lambda(\theta)$ is the solution of $n^{-1}\sum_{i=1}^n\frac{w_i(\theta)}{1+\lambda^Tw_i(\theta)}=0$. According to Qin and Lawless (1994), the maximum EL estimator $\hat\theta$ and its corresponding λ, denoted as $\hat\lambda$, are solutions of

$$Q_{1n}(\lambda,\theta)\ =\ n^{-1}\sum_{i=1}^n\frac{w_i(\theta)}{1+\lambda^Tw_i(\theta)}=0\quad\text{and} \tag{2.3}$$

$$Q_{2n}(\lambda,\theta)\ =\ n^{-1}\sum_{i=1}^n\frac{(\partial w_i(\theta)/\partial\theta)^T\lambda}{1+\lambda^Tw_i(\theta)}=0. \tag{2.4}$$

Then the log EL ratio is $r(\theta)=\ell(\theta)-\ell(\hat\theta)$.

In the following we are to develop expansions to $\ell(\theta_0)$ and $\ell(\hat\theta)$ respectively. To expand $\ell(\theta_0)$, define

$$\begin{aligned}\alpha^{j_1\ldots j_k}&=E\{w_i^{j_1}(\theta_0)\ldots w_i^{j_k}(\theta_0)\}\quad\text{and}\\ A^{j_1\ldots j_k}&=n^{-1}\sum_{i=1}^n w_i^{j_1}(\theta_0)\ldots w_i^{j_k}(\theta_0)-\alpha^{j_1\ldots j_k}.\end{aligned}$$

Here we use a^j to denote the j-th component of a vector a. Then, it may be shown that

$$\begin{aligned}n^{-1}\ell(\theta_0)\ =\ &A^jA^j-A^{ji}A^jA^i+\tfrac{2}{3}\alpha^{jih}A^jA^iA^h+A^{ji}A^{hi}A^jA^h+\tfrac{2}{3}A^{jih}A^jA^iA^h\\ &-2\alpha^{jih}A^{gh}A^jA^iA^g+\alpha^{jgf}\alpha^{ihf}A^jA^iA^hA^g-\tfrac{1}{2}\alpha^{jihg}A^jA^iA^hA^g+O_p(n^{-5/2}).\end{aligned} \tag{2.5}$$

We use here a convention where if a superscript is repeated a summation over that superscript is understood. This expansion has the same form as DiCiccio, Hall and Romano (1991) for the mean parameter when $r = p$ and Chen (1993) for linear regression.

It is quite challenging to expand $\ell(\hat\theta)$ in the general case of $r > p$. Two new systems of notations are introduced to facilitate the expansion. Let $\eta = (\lambda, \theta)$, $Q(\eta) = (Q_{1n}^{\tau}(\eta), Q_{2n}^{\tau}(\eta))^{\tau}$, $S_{21} = U(\Lambda, 0)$ and $S_{12} = S_{21}^{\tau}$. Due to the early transformation in (2.2),

$$S =: E\Big\{\frac{\partial Q(0,\theta_0)}{\partial \eta}\Big\} = \begin{pmatrix} -I & S_{12} \\ S_{21} & 0 \end{pmatrix}.$$

Put $\Gamma(\eta) = S^{-1}Q(\eta)$. Now we can introduce the notations involving $\Gamma(\eta)$ and their derivatives

$$\beta^{j,j_1\ldots j_k} = E\Big(\frac{\partial^k \Gamma^j(0,\theta_0)}{\partial \eta_{j_1}\ldots\partial\eta_{j_k}}\Big) \quad \text{and} \quad B^{j,j_1\ldots j_k} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^k \Gamma^j(0,\theta_0)}{\partial \eta_{j_1}\ldots\partial\eta_{j_k}} - \beta^{j,j_1\ldots j_k}$$

and the notations involving $w_i(\theta)$ and their derivatives

$$\begin{aligned} \gamma^{j,j_1\ldots j_l;k,k_1\ldots k_m;\ldots;p,p_1\ldots p_n} &= E\Big(\frac{\partial^l w_i^j(\theta_0)}{\partial\theta^{j_1}\ldots\partial\theta^{j_l}}\frac{\partial^m w_i^k(\theta_0)}{\partial\theta^{k_1}\ldots\partial\theta^{k_m}}\cdots\frac{\partial^n w_i^p(\theta_0)}{\partial\theta^{p_1}\ldots\partial\theta^{p_n}}\Big) \quad \text{and} \\ C^{j,j_1\ldots j_l;k,k_1\ldots k_m;\ldots;p,p_1\ldots p_n} &= \frac{1}{n}\sum_{i=1}^{n}\frac{\partial^l w_i^j(\theta_0)}{\partial\theta^{j_1}\ldots\partial\theta^{j_l}}\frac{\partial^m w_i^k(\theta_0)}{\partial\theta^{k_1}\ldots\partial\theta^{k_m}}\cdots\frac{\partial^n w_i^p(\theta_0)}{\partial\theta^{p_1}\ldots\partial\theta^{p_n}} - \gamma^{j,j_1\ldots j_l;k,k_1\ldots k_m;\ldots;p,p_1\ldots p_n}. \end{aligned}$$

Since $\hat\eta = (\hat\lambda, \hat\theta)$ is the solution of $\Gamma(\hat\eta) = 0$, by inverting this equation, derivations given in Appendix 2 show that for $j, k, l, m \in \{1, 2, \cdots, r+p\}$,

$$\begin{aligned} \hat\eta^j - \eta_0^j &= -B^j + B^{j,k}B^k - \tfrac{1}{2}\beta^{j,kl}B^kB^l - B^{j,k}B^{k,l}B^l + \tfrac{1}{2}\beta^{k,lm}B^{j,k}B^lB^m + \beta^{j,kl}B^{k,m}B^mB^l \\ &- \tfrac{1}{2}\beta^{j,kl}\beta^{k,mn}B^mB^nB^l - \tfrac{1}{2}B^{j,kl}B^kB^l + \tfrac{1}{6}\beta^{j,klm}B^kB^lB^m + O_p(n^{-2}). \end{aligned} \tag{2.6}$$

Note that (2.6) contains expansions for $\hat\lambda^j$ when $j \le r$ and for $\hat\theta^j$ when $j > r$ respectively. Note that

$$\ell(\hat\theta) = 2\sum_{i=1}^{n}\{\hat\lambda^T w_i(\hat\theta) - \tfrac{1}{2}[\hat\lambda^T w_i(\hat\theta)]^2 + \tfrac{1}{3}[\hat\lambda^T w_i(\hat\theta)]^3 - \tfrac{1}{4}[\hat\lambda^T w_i(\hat\theta)]^4\} + O_p(n^{-3/2}). \tag{2.7}$$

From now on, we fix the ranges of the superscripts $a, b, c, d \in \{1, 2, \cdots, r-p\}$, $f, g, h, i, j \in \{1, 2, \cdots, r\}$, $k, l, m, n, o \in \{1, 2, \cdots, p\}$ and $q, s, t, u \in \{1, 2, \cdots, r+p\}$. It is shown in

Appendix 2 by substituting (2.6) into (2.7) that

$$
\begin{aligned}
n^{-1}l(\hat{\theta}) \;=\;& -2B^jA^j - B^jB^j + 2C^{i,k}B^iB^{r+k,q}B^q + \tfrac{1}{2}\beta^{j,uq}\beta^{r+k,st}\gamma^{j,k}B^uB^qB^sB^t \\
-\;& \beta^{j,uq}B^uB^qB^{r+k,s}B^s\gamma^{j,k} - \beta^{r+k,uq}B^uB^qC^{i,k}B^i - B^jB^iA^{ji} - \tfrac{2}{3}\alpha^{jih}B^jB^iB^h \\
+\;& 2C^{j,k}\{B^jB^{r+k} - B^{j,p}B^pB^{r+k}[2,j,r+k] + \tfrac{1}{2}\beta^{j,uq}B^uB^qB^{r+k}[2,j,r+k]\} \\
+\;& \gamma^{j,kl}\{-B^jB^{r+k}B^{r+l} + B^jB^{r+k}B^{r+l,q}B^q[3,j,r+k,r+l] \\
-\;& \tfrac{1}{2}\beta^{j,uq}B^{r+k}B^{r+l}B^uB^q[3,j,r+k,r+l]\} - C^{j,kl}B^jB^{r+k}B^{r+l} - \tfrac{2}{3}A^{jih}B^jB^iB^h \\
-\;& B^{j,u}B^uB^{j,q}B^q - \tfrac{1}{4}\beta^{j,uq}\beta^{j,st}B^uB^qB^sB^t + \beta^{j,uq}B^uB^qB^{j,s}B^s + 2\gamma^{j;i;h,k}B^jB^iB^hB^{r+k} \\
+\;& B^jB^{i,q}B^qA^{ji}[2,j,i] - \tfrac{1}{2}\beta^{j,uq}B^uB^qB^iA^{ji}[2,j,i] + \tfrac{1}{3}\gamma^{j;k,lm}B^jB^{r+k}B^{r+l}B^{r+m} \\
+\;& 2\gamma^{j;i,l}\{B^jB^iB^{r+l} - B^jB^iB^{r+l,q}B^q + \tfrac{1}{2}\beta^{r+l,uq}B^jB^iB^uB^q - B^{r+l}B^iB^{j,q}B^q[2,j,i] \\
+\;& \tfrac{1}{2}\beta^{j,uq}B^uB^qB^iB^{r+l}[2,j,i]\} + 2B^jB^iB^{r+l}C^{j;i,l} - (\gamma^{j;i,lk} + \gamma^{j,l;i,k})B^jB^iB^{r+l}B^{r+k} \\
+\;& 2\alpha^{jih}B^jB^iB^{h,q}B^q - \alpha^{jih}\beta^{j,uq}B^uB^qB^iB^h - \tfrac{1}{2}\alpha^{jihg}B^jB^iB^hB^g + O_p(n^{-5/2})
\end{aligned}
\tag{2.8}
$$

where $[2,j,i]$ indicates there are two terms by exchanging the super-scripts i and j, and the same is understood for $[3,i,j,k]$. Expansion (2.8) for $\ell(\hat{\theta})$ is much more complicated than the just-identified case of $r = p$. In that case, all the $B^j = 0$ from a result established in (A.1), which means $\ell(\hat{\theta}) = 0$ and $r(\theta_0) = \ell(\theta_0)$. This is the situations of all the existing studies on Bartlett corretability of the EL. When $r > p$, the expansion of $\ell(\hat{\theta})$ contains more terms than that of $\ell(\theta_0)$, which increases substantially the difficulty of the second order study.

Combining (2.5) and (2.8), and carrying out further simplifications,

$$
\begin{aligned}
n^{-1}r(\theta_0) \;=\;& A^lA^l - A^{kl}A^kA^l - 2A^{l\ p+a}A^{p+a}A^l + \tfrac{2}{3}\alpha^{klm}A^kA^mA^l + 2\omega^{kl}C^{p+a,k}A^{p+a}A^l \\
+\;& \left(2\alpha^{kl\ p+a} - \gamma^{p+a,mn}w_{mk}w_{nl}\right)A^{p+a}A^kA^l + A^{ji}(A^{hi}A^jA^h - B^{i,q}B^qB^j[2,i,j]) \\
+\;& 2\left(\alpha^{l\ p+a\ p+b} - \gamma^{p+a;p+b,k}\omega^{kl}\right)A^{p+a}A^{p+b}A^l + B^{j,u}B^{j,q}B^uB^q + 2C^{j,k}B^{j,q}B^{r+k}B^q \\
-\;& \gamma^{j,kl}B^{r+k}B^{r+l}B^{j,q}B^q - 2\gamma^{j,kl}B^jB^{r+l}B^{r+k,q}B^q - 2\alpha^{jih}B^jB^iB^{h,q}B^q \\
+\;& 2\gamma^{j;i,l}(B^jB^iB^{r+l,q}B^q + B^{r+l}B^iB^{j,q}B^q[2,j,i]) + (\gamma^{j;i,lk} + \gamma^{j,l;i,k})B^jB^iB^{r+l}B^{r+k} \\
+\;& (\tfrac{1}{4}\beta^{j,uq}\beta^{j,st} - \tfrac{1}{2}\beta^{j,uq}\beta^{r+k,st}\gamma^{j,k})B^uB^qB^sB^t + (\alpha^{jih}\beta^{h,uq} - \gamma^{j;i,l}\beta^{r+l,uq})B^jB^iB^uB^q
\end{aligned}
$$

$$\begin{aligned}
&+ \ (\gamma^{j,kl}\beta^{r+l,uq} - \gamma^{i;j,k}\beta^{i,uq} - \gamma^{j;i,k}\beta^{i,uq})B^uB^qB^jB^{r+k} + \tfrac{1}{2}\gamma^{j,kl}\beta^{j,uq}B^uB^qB^{r+l}B^{r+k} \\
&- \ \tfrac{1}{3}\gamma^{j,klm}B^jB^{r+k}B^{r+l}B^{r+m} - 2\gamma^{j;i;h,k}B^jB^iB^hB^{r+k} + \tfrac{1}{2}\alpha^{jihg}B^jB^iB^hB^g \\
&+ \ C^{j,kl}B^jB^{r+k}B^{r+l} - 2C^{j;i,l}B^jB^iB^{r+l} + \tfrac{2}{3}A^{jih}(B^jB^iB^h + A^jA^iA^h) \\
&- \ 2\alpha^{jih}A^{gh}A^jA^iA^g + \alpha^{jgf}\alpha^{ihf}A^jA^iA^hA^g - \tfrac{1}{2}\alpha^{jihg}A^jA^iA^hA^g + O_p(n^{-5/2}).
\end{aligned} \tag{2.9}$$

This expansion leads to the following signed root decomposition

$$n^{-1}r(\theta_0) = R^jR^j + O_p(n^{-5/2})$$

where $R = R_1 + R_2 + R_3$ and $R_i = O_p(n^{-i/2})$ for $i = 1, 2$ and 3. Clearly, the terms appeared in the first two lines of (2.9) fully determine R_1 and R_2, namely

$$R_1^l = A^l, \tag{2.10}$$

$$\begin{aligned}
R_2^l \ &= \ -\tfrac{1}{2}A^{kl}A^k - A^{l\ p+a}A^{p+a} + \tfrac{1}{3}\alpha^{klm}A^kA^m + \omega^{kl}C^{p+a,k}A^{p+a} \\
&+ \ [\alpha^{kl\ p+a} - \tfrac{1}{2}\gamma^{p+a,mn}w_{mk}\omega^{nl}]A^{p+a}A^k + [\alpha^{l\ p+a,p+b} - \gamma^{p+a;p+b,k}\omega^{kl}]A^{p+a}A^{p+b}
\end{aligned} \tag{2.11}$$

and $R_1^j = R_2^j = 0$ for $j \in \{p+1, \cdots, r\}$. An expression for R_3^l is given in Appendix 2.

From (2.9), $r(\theta_0) = nA^lA^l + o_p(1)$ which means that $r(\theta_0) \xrightarrow{d} \chi_p^2$ and leads to an EL confidence region for θ with nominal confidence level $1-\alpha$: $I_\alpha = \{\theta | r(\theta) \le c_\alpha\}$ where c_α is the upper α-quantile of χ_p^2 distribution.

3. THE SECOND ORDER PROPERTIES

The coverage accuracy of the EL confidence region I_α is evaluated in the following theorem.

Theorem 1. Under conditions (2.1),

$$P\{r(\theta_0) < c_\alpha\} = \alpha - n^{-1}p^{-1}B_cc_\alpha f_p(c_\alpha) + O(n^{-2})$$

where $f_p(\cdot)$ is the density of χ_p^2 distribution, $B_c = p^{-1}(\sum_{l=1}^p \Delta^{ll} + \frac{1}{36}\alpha^{lkk}\alpha^{lmm})$ and Δ^{ll} is defined in (A.15).

Theorem 1 indicates that the coverage error of the EL confidence region I_α is $O(n^{-1})$, which is the same order as a standard two sided confidence region based on the asymptotic normality of $\hat{\theta}$. The attractions of the EL confidence region are (i) there is no need to carry out any secondary estimation procedure in formulating the confidence region whereas the covariance matrix has to be estimated for the confidence region based on the asymptotic normality; and (ii) the shape and the orientation of the region are naturally determined by the likelihood surface, free of any subjective intervention.

Theorem 2. Under conditions (2.1),

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

The theorem shows that the Bartlett correction is maintained by the EL for the situation of general moment restrictions, despite that r may be larger than p and $\ell(\hat{\theta})$ has a rather complex expression. This indicates that the EL is resilient in sharing this delicate second order property with a parametric likelihood and the existence of certain internal mechanism in the EL that resembles that of the parametric likelihood.

It can be seen from Appendix A.4. that the Bartlett factor B_c has a rather involved expression for a general over-identified case of $r > p$ due to the lengthy expressions of Δ^{ll}. However, it admits simpler expression in two special situations. One is in the situation of just-identified moment restrictions with $r = p$. It may be easily checked from (A.15) that

$$B_c = p^{-1}\left(\tfrac{1}{2}\alpha^{llkk} - \tfrac{1}{3}\alpha^{lkm}\alpha^{lkm}\right),$$

which is the Bartlett factor obtained in DiCiccio, Hall and Romano (1991) for smooth function of means and Chen (1993) for linear regression. The other situation is when $r > p$, but (i) $Cov\{g^j(X, \theta_0), g^{p+a}(X, \theta_0)\} = 0$ for any $j \le p$ and $a \le r - p$ and (ii) $g^j(x, \theta) = g^j(x)$ does not depend on θ for $j = p + 1, \cdots, r$. The assumption (i) means that the first p estimating equations are uncorrelated with the last $r - p$ estimating equations at θ_0 and (ii) means that the last $r - p$ estimating equations are free of parameters. In this case,

$$B_c = p^{-1}\left(\tfrac{1}{2}\alpha^{llkk} + \alpha^{ll\ p+a\ p+a} - \tfrac{1}{3}\alpha^{lkm}\alpha^{lkm} - \alpha^{llk}\alpha^{kp+a\ p+a} - \alpha^{lf\ p+a}\alpha^{lf\ p+a}\right).$$

To practically implement the Bartlett correction in a general situation, $\widetilde{B}_c =: 1 + B_c n^{-1}$ has to be estimated. It is noted that the direct plug-in estimator of $\widetilde{B}_c$ can be obtained by substituting all the populations moments involved by their corresponding sample moments. However, considering the rather lengthy forms of $\widetilde{B}_c$, we propose using the bootstrap to estimate $\widetilde{B}_c$, which is based on the fact that

$$E\{r(\theta_0)\} = p(1 + B_c n^{-1}) + O(n^{-2}) = p\widetilde{B}_c + O(n^{-2}) \tag{3.1}$$

by combining the expressions of $E(R_i^l R_j^l)$ given in Appendix 3. The bootstrap procedure is

Step 1: generate a bootstrap resample $\{X_i^*\}_{i=1}^n$ by sampling with replacement from the original sample $\{X_i\}_{i=1}^n$ and compute $r^*(\hat{\theta}) = \ell^*(\hat{\theta}) - \ell^*(\hat{\theta}^*)$, where ℓ^* and $\hat{\theta}^*$ are respectively the Log EL ratio and the maximum EL estimate based on the resample;

Step 2: for a large integer N, repeat Step 1 N times and obtain $r^{*1}(\hat{\theta}), \cdots, r^{*N}(\hat{\theta})$.

As $N^{-1}\sum_{b=1}^N r^{*b}(\hat{\theta})$ estimates $E\{r(\theta_0)\}$, a bootstrap estimate of $\widetilde{B}_c$ is

$$\hat{\widetilde{B}}_c = (Np)^{-1}\sum_{b=1}^N r^{*b}(\hat{\theta}).$$

Let $\mathcal{X}_\mathbf{n} = \{X_1, \cdots, X_n\}$ be the original sample. It may be shown by standard bootstrap arguments, for instance those given in Hall (1992), that

$$E\{\hat{\widetilde{B}}_c | \mathcal{X}_\mathbf{n}\} = (1 + B_c n^{-1})\{1 + O_p(n^{-1/2})\} \tag{3.2}$$

which means that the bootstrap estimate of $\widetilde{B}_c$ is $\sqrt{n}$-consistent. Now a practical Bartlett corrected confidence region is $I_{\alpha,bc} = \{\theta | r(\theta) \le c_\alpha \hat{\widetilde{B}}_c\}$. It can be shown from Theorem 2 and (3.2) that the coverage error of $I_{\alpha,bc}$ is $O(n^{-3/2})$, which improves that of I_α.

The above use of the bootstrap to estimate the Bartlett factor B_c or $\widetilde{B}_c$ naturally leads ones to think of using the bootstrap to calibrate directly on the distribution of the EL ratio $r(\theta_0)$. Let us rank $\{r^{*i}(\hat{\theta})\}_{i=1}^N$ such that $r^{*1}(\hat{\theta}) \le r^{*2}(\hat{\theta}) \le \cdots \le r^{*N}(\hat{\theta})$. Then a direct bootstrap confidence interval with a nominal level $1 - \alpha$ is $I_{\alpha,bt} = (r^{*[\alpha N/2]+1}(\hat{\theta}), r^{*[(1-\alpha)N/2]+1}(\hat{\theta}))$ where $[\cdot]$ is the integer truncation operator.

The cumulants and expansions which are quite expensively derived for the purpose of establishing Bartlett correction are needed in assessing the coverage accuracy of the direct Bootstrap confidence interval $I_{\alpha,bt}$. It may be shown that under conditions (2.1),[1]

$$P\left(\theta_0 \in I_{\alpha,bt}\right) = \alpha + O(n^{-3/2}) \tag{3.3}$$

which indicates that the coverage error of the bootstrap confidence interval $I_{\alpha,bt}$ and the Bartlett corrected interval $I_{\alpha,bc}$ is at the same order. This is indeed confirmed by our simulation studies reported in the next section, although we observe that the performance of the Bartlett corrected interval is more robust.

4. SIMULATION RESULTS

We report in this section results of two simulation studies designed to confirm the theoretical finding of Bartlett correction of the EL by implementing the proposed empirical Bartlett correction. For comparison purposes, the bootstrap confidence intervals $I_{\alpha,bt}$ is also evaluated.

In the first simulation study, $X_1, \cdots, X_n$ are independent and identically $N(\theta, \theta^2+1)$ distributed, as considered in an example of Qin and Lawless (1994). The relationship between the mean and variance leads to moment restrictions: $g_1(X_1,\theta) = X_1 - \theta$ and $g_2(X_1,\theta) = X_1^2 - 2\theta - 1$. This is an over-identified case as there are two moment restrictions and one parameter of interest, i.e. $r = 2$ and $p = 1$. Like Qin and Lawless, the value of θ is chosen to be 0 and 1 respectively. The sample size used in the simulation study is $n = 20, 30, 40$ and 50 respectively.

In the second simulation study, we consider the following autoregressive panel data model, which is an example considered in Brown and Newey (2002)

$$X_{it} = \theta X_{it-1} + \alpha_i + \epsilon_{it}, \quad X_{i0} = \frac{\alpha_i}{1-\rho} + e_i, \tag{4.1}$$

---

[1] We would not provide the proof here due to a limited space. It can be carried out by taking a similar route as in Hall (1992) and utilizing the Edgeworth expansion established in the proof of Theorem 1.

for $t = 1, \cdots, 4$ and $i = 1, \cdots, n$, where $|\theta| < 1$, $\{\epsilon_{it}\}_{t=1}^4$ and α_i are mutually independent standard normal random variables, $v_i \sim N(0, (1-\theta^2)^{-1})$ and independent of $\{\epsilon_{it}\}_{t=1}^4$ and α_i. Let $X_i = (X_{i1}, \ldots, X_{i4})$. The moment restrictions after taking time differencing are $g_1(X_i, \theta) = X_{i1}(\Delta X_{i3} - \theta \Delta X_{i2})$, $g_2(X_i, \theta) = X_{i1}(\Delta X_{i4} - \theta \Delta X_{i3})$ and $g_3(X_i, \theta) = X_{i2}(\Delta X_{i3} - \theta \Delta X_{i2})$ where $\Delta X_{it} = X_{it} - X_{it-1}$. It is easy to check from model (4.1) that $E\{g_j(X_i, \theta)\} = 0$. Hence, there are three constraints and one parameter, i.e. $r = 3$ and $p = 1$, another over-identified case. The parameter θ, which is the autoregressive coefficient is assigned values of 0.5 and 0.9 to obtain different levels of correlations. The sample size is chosen at $n = 50$ and 100 respectively.

In both simulation studies, the empirical coverage and length of the EL, Bartlett corrected EL and the direct bootstrap calibrated intervals are evaluated with nominal coverage levels of 90% and 95% respectively. The bootstrap resample size N used in the Bartlett correction is 250 and the number of simulation is 1000.

Tables 1 and 2 contain the empirical coverage and the averaged length of the three types of confidence intervals, which can be summarized as follows. First of all, the need for carrying out the second order correction to the EL confidence interval I_α is quite obvious as the original EL interval has quite severe under coverage for all the cases considered even for a sample size of 100 for the panel data model. The under coverage is particularly severe when the sample size is small for the normal mean model $N(1, 2)$ and for the panel data model, These are the situations where the Bartlett correction is needed. It is observed that in all the cases considered the Bartlett correction improves significantly the coverage of I_α. The restoration of coverage by the Bartlett correction is very impressive. We also observed that, as anticipated in (3.3), the direct bootstrap confidence interval has similar performance with the Bartlett corrected intervals in most of the cases. However, in the normal mean models with 90% nominal coverage level, the coverage of the direct bootstrap interval $I_{\alpha,bt}$ is not as good as the Bartlett corrected interval $I_{\alpha,bc}$. The robust performance of the Bartlett corrected interval is due to the fact that the estimation of the Bartlett factor

$\tilde{B}_c$, which involves simple bootstrap averaging, is more robust than the bootstrap estimation of the extreme quantiles of the distribution of $r(\hat{\theta})$. We also observed in passing that as the sample size increases the EL interval I_α improves both in its coverage and length whereas the improvement on the Bartlett intervals and the bootstrap interval is more in reducing the length of the intervals.

5. CONCLUSIONS

The main finding of the paper is that the EL with general moment restrictions are Bartlett correctable. This is a substantial extension of the previously established cases of Bartlett correction of EL, including the case of smoothed functions of means by DiCiccio, Hall and Romano (1991) more than one decade ago. It shows that the dedicate Bartlett property of the EL is still preserved even in the case of over-identification. Although the Bartlett factor admits a very involved expression with over-identified moment restrictions, proving that the EL is Bartlett correctable in the general case provides the theoretical foundation to the proposed easily implementable empirical Bartlett correction.

The use of the bootstrap to carry out the Bartlett correction empirically is due to a rather involved expression for the Bartlett factor. Although it may be expected that the direct bootstrap calibration would give the same effect as the Bartlett correction, the justification of the direct bootstrap method inevitably needs those cumulants and the Edgeworth expansions established in this paper.

The results established in Theorems 1 and 2 can be extended to independent but not identically distributed samples, for instance those arisen in a regression study. We need to modify α, β and γ as follows:

$$\alpha^{j_1\dots j_k} = n^{-1}\sum_{i=1}^{n} E[w_i^{j_1}(\theta_0)\dots w_i^{j_k}(\theta_0)], \quad \beta^{j,j_1\dots j_k} = n^{-1}\sum_{i=1}^{n} E\Big(\frac{\partial^k \Gamma^j(0,\theta_0)}{\partial\eta_{j_1}\dots\partial\eta_{j_k}}\Big) \quad \text{and}$$

$$\gamma^{j,j_1\dots j_l;k,k_1\dots k_m;\dots;p,p_1\dots p_n} = \frac{1}{n}\sum_{i=1}^{n} E\Big(\frac{\partial^l w_i^j(\theta_0)}{\partial\theta^{j_1}\dots\partial\theta^{j_l}}\frac{\partial^m w_i^k(\theta_0)}{\partial\theta^{k_1}\dots\partial\theta^{k_m}}\cdots\frac{\partial^n w_i^p(\theta_0)}{\partial\theta^{p_1}\dots\partial\theta^{p_n}}\Big).$$

We need also to re-define V_n as $n^{-1}\sum_{i=1}^{n} Var\{g(X_i,\theta_0)\}$. These forms of α and V_n were employed in Chen (1993) to establish Bartlett correction for linear regression where $r =$

p. The conditions (2.1) should be modified to reflect the independent but not identically distributed nature of data. Similar conditions as those given in Theorem 20.6 of Bhattacharya and Rao (1976) are required, as in Chen (1993, 1994). Then, it may be shown that Theorem 1 is true by employing Skovgaard (1981) on transformation of Edgeworth expansions. Theorem 2 is then a consequence of Theorem 1 as the calculation of the cumulants follows the same spirits given in the Appendix for independent and identically distributed samples.

APPENDIX

We provide some technical details on the log EL ratio $r(\theta_0)$ in A.2, the sign root decomposition in A.3., and the proofs of the two theorems in A.4. More details on these derivations can be found in Chen and Cui (2003).

A.1. Basic formulae.

We first present some basic formulae which will be used throughout the derivations.

Let us define $\Omega = (\omega^{kl})_{p\times p} =: U\Lambda^{-1}$ where $\omega^{kl} = u^{kl}\lambda_l^{-1}$. Please note here that no summation over the subscript l is carried out due to Λ being a diagonal matrix. Since $\Gamma(\eta) = S^{-1}Q(\eta)$ where

$$S^{-1} = \begin{pmatrix} -I + S_{12}(S_{12}^T S_{12})^{-1}S_{12}^T & S_{12}(S_{12}^T S_{12})^{-1} \\ (S_{12}^T S_{12})^{-1}S_{12}^T & (S_{12}^T S_{12})^{-1} \end{pmatrix} = \begin{pmatrix} 0 & 0 & \Omega^T \\ 0 & -I_{r-p} & 0 \\ \Omega & 0 & \Omega\Omega^T \end{pmatrix},$$

it can be checked that

$$\beta^{j,k} = E(\frac{\partial \Gamma^j(0,\theta_0)}{\partial \eta_k}) = \delta^{jk} \quad \text{and} \quad B =: \begin{pmatrix} B^1 \\ \cdots \\ B^r \end{pmatrix} = S^{-1}\begin{pmatrix} A \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -A_2 \\ \Omega A_1 \end{pmatrix}.$$

Here $A^T = (A^1, \cdots A^r)^T =: (A_1^T, A_2^T)^T$, where $A_1 = (A^1, \cdots, A^p)^T$ and $A_2 = (A^{p+1}, \cdots, A^r)^T$ constitute a partition of A. Therefore for positive integers k and a,

$$\text{(A.1)} \quad B^k = 0 \text{ for } k \le p; B^{p+a} = -A^{p+a} \text{ for } a \le r-p, \text{ and } B^{r+k} = \omega^{kl}A^l \text{ for } k \le p.$$

Let $B_1 = (B^1, \cdots, B^r)^\tau$ and $B_2 = (B^{r+1}, \cdots B^{r+p})^\tau$. Since $SB = (A^\tau, 0^\tau_{p\times 1})^\tau$ which means that $-B_1 + S_{12}B_2 = A$. As $S_{12} = (\gamma^{j,k})_{r\times p}$ and from (A.1) we have

$$\gamma^{j,k} B^{r+k} = A^j I(j \le p) \tag{A.2}$$

where I is the indicator function. Since

$$(B^{p,q})_{(r+p)\times(r+p)} = S^{-1} \begin{pmatrix} -(A^{ij}) & (C^{i,l}) \\ (C^{i,l})^T & 0 \end{pmatrix}, \tag{A.3}$$

we have $S_{21}(B^{j,k})_{r\times p} = (C^{k,m})^\tau_{p\times r}$ and $S_{21}(B^{j,r+a})_{r\times p} = 0$. As $S_{21} = (\gamma^{j,k})^\tau$, these mean

$$\gamma^{j,k} B^{j,l} = C^{l,k} \text{ for } l \le r \text{ and } k \le p \text{ and } \gamma^{j,k} B^{j,r+a} = 0. \tag{A.4}$$

Furthermore, (A.3) also implies the following which links the $B^{s,t}$ system with the A^{jm}- and the $C^{j,m}$- systems

$$\begin{pmatrix} (B^{k,l}) & (B^{k,p+b}) & (B^{k,r+l}) \\ (B^{p+a,l}) & (B^{p+a,p+b}) & (B^{p+a,r+l}) \\ (B^{r+k,l}) & (B^{r+k,p+b}) & (B^{r+k,r+l}) \end{pmatrix} \tag{A.5}$$

$$= \begin{pmatrix} (\omega^{mk} C^{l,m}) & (\omega^{mk} C^{p+b,m}) & 0 \\ (A^{p+a\ l}) & (A^{p+a\ p+b}) & -(C^{p+a,l}) \\ (\omega^{km}[\omega^{nm} C^{l,n} - A^{ml}]) & (\omega^{km}[\omega^{nm} C^{p+b,n} - A^{m\ p+b}]) & (\omega^{km} C^{m,l}) \end{pmatrix}.$$

In a similar fashion, we can establish the following links between $\beta^{s,\underline{t}}$ and $(\alpha^{j\underline{i}}, \gamma^{j,\underline{i}})$ systems where $\underline{t}$ and $\underline{i}$ contains either single or double superscripts:

$$\begin{aligned}
\beta^{l,p+a\ p+c} &= -\omega^{ol}[\gamma^{p+c;p+a,o} + \gamma^{p+a;p+c,o}], \qquad \beta^{l,p+m\ p+c} = \omega^{ol}\gamma^{p+c;om}, \\
\beta^{p+a,p+b\ p+c} &= -2\alpha^{p+a\ p+b\ p+c}, \qquad \beta^{p+a,p+m\ p+c} = \gamma^{p+c;p+a,m} + \gamma^{p+a;p+c,m}, \\
\beta^{l,p+a\ p+n} &= \omega^{ol}\gamma^{p+a;on}, \qquad \beta^{l,p+m\ p+n} = 0, \\
\beta^{p+a,p+b\ p+n} &= \gamma^{p+a,n;p+b} + \gamma^{p+a;p+b,n}, \qquad \beta^{p+a,p+m\ p+n} = -\gamma^{p+a,mn}, \\
\beta^{r+k,p+a\ p+c} &= 2\omega^{ko}\alpha^{op+a\ p+c} - \omega^{ko}\omega^{no}[\gamma^{p+c;p+a,n} + \gamma^{p+a;p+c,n}], \\
\beta^{r+k,p+a\ p+n} &= \omega^{ko}\omega^{mo}\gamma^{p+a,mn} - \omega^{ko}[\gamma^{o,n;p+a} + \gamma^{o;p+a,n}], \\
\beta^{r+k,p+m\ p+c} &= \omega^{ko}\omega^{no}\gamma^{p+c,nm} - \omega^{ko}[\gamma^{p+c;o,m} + \gamma^{o,p+c;m}], \ \beta^{r+k,p+m\ p+n} = \omega^{ko}\gamma^{o,mn}.
\end{aligned} \tag{A.6}$$

See Chen and Cui (2003) for details.

A.2. DERIVATIONS OF (2.8) AND (2.9)

We shall expand each term on the right of (2.7). By ignoring terms of $O_p(n^{-5/2})$, the first term

$$\begin{aligned}
\hat\lambda^T n^{-1}\sum_{i=1}^n w_i(\hat\theta) &= \hat\lambda^j n^{-1}\sum_{i=1}^n [w_i^j(\theta_0) + \frac{\partial w_i^j(\theta_0)}{\partial\theta^k}\hat\theta^k + \tfrac{1}{2}\tfrac{\partial^2 w_i^j(\theta_0)}{\partial\theta^k\partial\theta^l}\hat\theta^k\hat\theta^l + \tfrac{1}{6}\tfrac{\partial^3 w_i^j(\theta_0)}{\partial\theta^k\partial\theta^l\partial\theta^m}\hat\theta^k\hat\theta^l\hat\theta^m] \\
&= \hat\lambda^j A^j + \gamma^{j,k}\hat\lambda^j\hat\theta^k + \hat\lambda^j\hat\theta^k C^{j,k} + \tfrac{1}{2}\gamma^{j,kl}\hat\lambda^j\hat\theta^k\hat\theta^l + \tfrac{1}{2}\hat\lambda^j\hat\theta^k\hat\theta^l C^{j,kl} + \tfrac{1}{6}\gamma^{j,klm}\hat\lambda^j\hat\theta^k\hat\theta^l\hat\theta^m.
\end{aligned}$$

Similarly, the second term

$$\begin{aligned}
&\hat\lambda^T n^{-1}\sum_{i=1}^n w_i(\hat\theta)w_i(\hat\theta)^T\hat\lambda \\
= \; &\hat\lambda^j\hat\lambda^h n^{-1}\sum_{i=1}^n \{w_i^j(\theta_0)w_i^h(\theta_0) + w_i^h(\theta_0)\frac{\partial w_i^j(\theta_0)}{\partial\theta^l}\hat\theta^l[2,j,h] + \tfrac{1}{2}w_i^j(\theta_0)\tfrac{\partial^2 w_i^h(\theta_0)}{\partial\theta^l\partial\theta^k}\} \\
= \; &\hat\lambda^j\hat\lambda^i(A^{ji}+\delta^{ji}) + \hat\lambda^j\hat\lambda^i\hat\theta^l\{(C^{j;i,l}+\gamma^{j;i,l})[2,j,i]\} + \tfrac{1}{2}\hat\lambda^j\hat\lambda^i\hat\theta^l\hat\theta^k\{(C^{j;i,lk}+\gamma^{j;i,lk})[2,j,i]\} \\
+ \; &\hat\lambda^j\hat\lambda^i\hat\theta^l\hat\theta^k\{(C^{j,l;i,k}+\gamma^{j,l;i,k}\} \\
= \; &\hat\lambda^j\hat\lambda^j + \hat\lambda^j\hat\lambda^i A^{ji} + 2\gamma^{j;i,l}\hat\lambda^j\hat\lambda^i\hat\theta^l + 2C^{j;i,l}\hat\lambda^j\hat\lambda^i\hat\theta^l + (\gamma^{j;i,lk}+\gamma^{j,l;i,k})\hat\lambda^j\hat\lambda^i\hat\theta^l\hat\theta^k + O_p(n^{-5/2})
\end{aligned}$$

where $[2,j,i]$ indicates there are two terms by swamping the super-scripts i and j, and the same is understood for similar notations. For the third term

$$\tfrac{2}{3}n^{-1}\textstyle\sum_{i=1}^n[\hat\lambda^T w_i(\hat\theta)]^3 \;=\; \tfrac{2}{3}\hat\lambda^j\hat\lambda^i\hat\lambda^h(A^{jih}+\alpha^{jih}) + 2\hat\lambda^j\hat\lambda^i\hat\lambda^h\hat\theta^k\gamma^{j;i;h,k} + O_p(n^{-5/2}).$$

Finally, $n^{-1}\sum_{i=1}^n[\hat\lambda^T w_i(\hat\theta)]^4 = \hat\lambda^j\hat\lambda^i\hat\lambda^h\hat\lambda^g\alpha^{jihg} + O_p(n^{-5/2})$.

We then have for $a,b,c,d \in \{1,2,\cdots,r-p\}$, $f,g,h,i,j \in \{1,2,\cdots,r\}$, $k,l,m,n,o \in \{1,2,\cdots,p\}$ and $q,s,t,u \in \{1,2,\cdots,r+p\}$ that

$$\begin{aligned}
n^{-1}l(\hat\theta) &= -2B^jA^j - B^jB^j + 2B^{j,q}B^q(A^j+B^j) - \beta^{j,uq}B^uB^q(A^j+B^j) \\
&- 2B^{j,u}B^{u,q}B^q(A^j+B^j) + \beta^{u,qs}B^{j,u}B^qB^s(A^j+B^j) - \beta^{j,uq}\beta^{u,st}B^qB^sB^t(A^j+B^j) \\
&- B^{j,uq}B^uB^q(A^j+B^j) + \tfrac{1}{3}\beta^{j,uqs}B^uB^qB^s(A^j+B^j) + 2\beta^{j,uq}B^{u,s}B^sB^q(A^j+B^j) \\
&+ 2\gamma^{j,k}\{-B^{j,q}B^qB^{r+k} + [\tfrac{1}{2}\beta^{j,uq}B^uB^qB^{r+k} + B^{j,u}B^{u,q}B^qB^{r+k}
\end{aligned}$$

$$
\begin{aligned}
&- \tfrac{1}{2}\beta^{u,qs}B^{j,u}B^qB^sB^{r+k} - \beta^{j,uq}B^{u,s}B^qB^sB^{r+k} + \tfrac{1}{2}\beta^{j,uq}\beta^{u,st}B^qB^sB^tB^{r+k} \\
&+ \tfrac{1}{2}B^{j,uq}B^uB^qB^{r+k} - \tfrac{1}{6}\beta^{j,uqs}B^uB^qB^sB^{r+k} - \tfrac{1}{2}\beta^{j,uq}B^uB^qB^{r+k,s}B^s][2,j,r+k] \\
&+ B^{j,u}B^uB^{r+k,q}B^q + \tfrac{1}{4}\beta^{j,uq}\beta^{r+k,st}B^uB^qB^sB^t\} \\
&+ 2C^{j,k}\{B^jB^{r+k} - B^{j,q}B^qB^{r+k}[2,j,r+k] + \tfrac{1}{2}\beta^{j,uq}B^uB^qB^{r+k}[2,j,r+k]\} \\
&+ \gamma^{j,kl}\{-B^jB^{r+k}B^{r+l} + B^{r+k}B^{r+l}B^{j,q}B^q[3,j,r+k,r+l] \\
&- \tfrac{1}{2}B^jB^{r+k}\beta^{r+l,uq}B^uB^q[3,j,r+k,r+l]\} - C^{j,kl}B^jB^{r+k}B^{r+l} \\
&+ \tfrac{1}{3}\gamma^{j,klm}B^jB^{r+k}B^{r+l}B^{r+m} - B^{j,u}B^uB^{j,q}B^q - \tfrac{1}{4}\beta^{j,uq}\beta^{j,st}B^uB^qB^sB^t \\
&+ \beta^{j,uq}B^uB^qB^{j,s}B^s - B^jB^iA^{ji} + B^jB^{i,q}B^qA^{ji}[2,j,i] - \tfrac{1}{2}\beta^{j,uq}B^uB^qB^iA^{ji}[2,j,i] \\
&+ 2\gamma^{j;i,l}\{B^jB^iB^{r+l} - B^jB^iB^{r+l,q}B^q + \tfrac{1}{2}\beta^{r+l,uq}B^jB^iB^uB^q - B^{r+l}B^iB^{j,q}B^q[2,j,i] \\
&+ \tfrac{1}{2}\beta^{j,uq}B^uB^qB^iB^{r+l}[2,j,i]\} + 2B^jB^iB^{r+l}C^{j;i,l} - (\gamma^{j;i,lk} + \gamma^{j,l;i,k})B^jB^iB^{r+l}B^{r+k} \\
&- \tfrac{2}{3}\alpha^{jih}B^jB^iB^h + 2\alpha^{jih}B^jB^iB^{h,q}B^q - \alpha^{jih}\beta^{j,uq}B^uB^qB^iB^h - \tfrac{2}{3}A^{jih}B^jB^iB^h \\
&+ 2\gamma^{j;i;h,k}B^jB^iB^hB^{r+k} - \tfrac{1}{2}\alpha^{jihg}B^jB^iB^hB^g + O_p(n^{-5/2}).
\end{aligned}
$$

Applying (A.1) and (A.4), it may be shown that the 3rd to the 18th terms on the right hand side cancel each other and the application of (A.4) simplifies the 20th term. Keep all the other terms, we have (2.8).

Now bring in the expansion for $\ell(\theta_0)$ in (2.5) we have

$$
\begin{aligned}
n^{-1}r(\theta_0) &= (A^j + B^j)(A^j + B^j) - A^{ji}(A^jA^i - B^jB^i) - 2C^{j,k}B^jB^{r+k} - 2\gamma^{j;i,l}B^jB^iB^{r+l} \\
&+ \tfrac{2}{3}\alpha^{jih}[A^jA^iA^h + B^jB^iB^h] + \gamma^{j,kl}B^jB^{r+k}B^{r+l} + A^{ji}(A^{hi}A^jA^h - B^{i,q}B^qB^j[2,i,j]) \\
&- \beta^{j,uq}[C^{j,k}B^{r+k} - B^{r+k,s}B^s\gamma^{j,k} + B^{j,s}B^s - A^{ji}B^i]B^uB^q - 2\alpha^{jih}B^jB^iB^{h,q}B^q \\
&+ B^{j,u}B^{j,q}B^uB^q + 2C^{j,k}B^{j,q}B^{r+k}B^q - \gamma^{j,kl}B^{r+k}B^{r+l}B^{j,q}B^q \\
&- 2\gamma^{j,kl}B^jB^{r+l}B^{r+k,q}B^q + 2\gamma^{j;i,l}(B^jB^iB^{r+l,q}B^q + B^{r+l}B^iB^{j,q}B^q[2,j,i]) \\
&+ (\tfrac{1}{4}\beta^{j,uq}\beta^{j,st} - \tfrac{1}{2}\beta^{j,uq}\beta^{r+k,st}\gamma^{j,k})B^uB^qB^sB^t + \tfrac{1}{2}\gamma^{j,kl}\beta^{j,uq}B^uB^qB^{r+l}B^{r+k} \\
&+ (\gamma^{j,kl}\beta^{r+l,uq} - \gamma^{i;j,k}\beta^{i,uq} - \gamma^{j;i,k}\beta^{i,uq})B^uB^qB^jB^{r+k} - 2\gamma^{j;i;h,k}B^jB^iB^hB^{r+k} \\
&+ (\gamma^{j;i,lk} + \gamma^{j,l;i,k})B^jB^iB^{r+l}B^{r+k} - \tfrac{1}{3}\gamma^{j,klm}B^jB^{r+k}B^{r+l}B^{r+m}
\end{aligned}
$$

$$\begin{aligned} &+ \tfrac{1}{2}\alpha^{jihg}B^jB^iB^hB^g + (\alpha^{jih}\beta^{h,uq} - \gamma^{j;i,l}\beta^{r+l,uq})B^jB^iB^uB^q \\ &+ C^{j,kl}B^jB^{r+k}B^{r+l} - 2C^{j;i,l}B^jB^iB^{r+l} + \tfrac{2}{3}A^{jih}(B^jB^iB^h + A^jA^iA^h) \\ &- 2\alpha^{jih}A^{gh}A^jA^iA^g + \alpha^{jgf}\alpha^{ihf}A^jA^iA^hA^g - \tfrac{1}{2}\alpha^{jihg}A^jA^iA^hA^g + O_p(n^{-5/2}). \end{aligned} \tag{A.7}$$

As shown in Chen and Cui (2003), the terms appeared on the 3rd line of the above equation cancel each other. Applying again (A.5), we can express the terms appeared in the first two lines of (A.7) by

$$\begin{aligned} & A^lA^l - A^{kl}A^kA^l - 2A^{l\ p+a}A^{p+a}A^l + \tfrac{2}{3}\alpha^{klm}A^kA^mA^l + 2\omega^{kl}C^{p+a,k}A^{p+a}A^l \\ + & [2\alpha^{kl\ p+a} - \gamma^{p+a,mn}\omega^{mk}\omega^{nl}]A^{p+a}A^kA^l + 2[\alpha^{l\ p+a\ p+b} - \gamma^{p+a;p+b,k}\omega^{kl}]A^{p+a}A^{p+b}A^l \end{aligned}$$

which leads us to (2.9).

A.3. EXPANSION FOR R_3

We subtract $R_2^lR_2^l$ from all the terms appeared in line 4 and below in (2.9). Fortunately all the terms which do not have A^l appeared cancel out with those appeared in $R_2^lR_2^l$. Otherwise, a signed root decomposition of the EL ratio $r(\theta_0)$ would not be possible. Hence the remaining terms can be written as $2R_1^lR_3^l$.

By repeatedly employing the formulae (A.5) and (A.6) as well as (A.1), (A.2) and (A.4), it is shown in Chen and Cui (2003) after some quite involved algebra that $R_3^l = \sum_{i=1}^6 R_{3i}^l$ where

$$\begin{aligned} R_{31}^l &= \tfrac{3}{8}A^{lm}A^{km}A^k + \tfrac{1}{3}A^{lkm}A^kA^m - \tfrac{5}{12}\alpha^{lkm}A^{nm}A^kA^n \\ &- \tfrac{5}{12}\alpha^{knm}A^{lm}A^kA^n + \tfrac{4}{9}\alpha^{lkn}\alpha^{omn}A^mA^kA^o - \tfrac{1}{4}\alpha^{lknm}A^mA^kA^n, \\ R_{32}^l &= A^{lk\ p+a}A^{p+a}A^k + A^{l\ p+a\ p+b}A^{p+a}A^{p+b} - \tfrac{1}{2}\omega^{kn}\omega^{ml}C^{p+a,km}A^{p+a}A^n \\ &- \omega^{kl}C^{p+a;p+b,k}A^{p+a}A^{p+b} + \tfrac{1}{2}\omega^{km}C^{p+a,k}A^{lm}A^{p+a} + \tfrac{1}{2}A^{lk}A^{k\ p+a}A^{p+a} \\ &+ A^{lp+a}A^{p+a\ p+b}A^{p+b} + \tfrac{1}{2}A^{lp+a}A^{p+ak}A^k - \tfrac{1}{2}\omega^{km}\omega^{nl}C^{p+a,n}C^{p+a,k}A^m \\ &- \omega^{kl}\omega^{mn}C^{n,k}C^{p+a,m}A^{p+a} - \omega^{kl}C^{p+a,k}A^{p+a\ p+b}A^{p+b}, \\ R_{33}^l &= \tfrac{1}{2}[\gamma^{m,vo}\omega^{vn}\omega^{ol}\omega^{km} - \tfrac{2}{3}\alpha^{nml}\omega^{km}]C^{p+a,k}A^{p+a}A^n + \tfrac{1}{2}\gamma^{p+b,ko}\omega^{kn}\omega^{ol}\omega^{mv}C^{p+b,m}A^nA^v \end{aligned}$$

$$
\begin{aligned}
&+ && [\tfrac{1}{2}\gamma^{p+b,ko}\omega^{kn}\omega^{ol} - \alpha^{lnp+b}]A^{p+b\ p+a}A^{p+a}A^{n} + \omega^{vk}[\omega^{nl}(\gamma^{k;p+a,n} + \gamma^{p+a;k,n}) \\
&- && \alpha^{lkp+b} - \tfrac{1}{2}\gamma^{p+b,mn}\omega^{ml}\omega^{nk}]C^{p+a,v}A^{p+a}A^{p+b} + \gamma^{p+a,ko}\omega^{on}\omega^{ml}\omega^{kv}C^{v,m}A^{p+a}A^{n} \\
&+ && [\tfrac{1}{2}\gamma^{p+b,mk}\omega^{ml}\omega^{kn} - \alpha^{lnp+b}]A^{n\ p+a}A^{p+a}A^{p+b}, \\
R^{l}_{34} &= && \gamma^{p+a;p+b,n}\omega^{no}\omega^{ml}C^{o,m}A^{p+a}A^{p+b} - \alpha^{p+a\ p+b\ p+c}A^{l\ p+c}A^{p+a}A^{p+b} \\
&+ && [(\gamma^{p+c;p+a,n} + \gamma^{p+a;p+c,n})\omega^{nl} - 2\alpha^{l\ p+a\ p+c}]A^{p+c\ p+b}A^{p+a}A^{p+b} \\
&+ && (\gamma^{p+c;p+a,o} + \gamma^{p+a;p+c,o})\omega^{on}\omega^{kl}C^{p+c,k}A^{p+a}A^{n} - \alpha^{lk\ p+a}A^{m\ p+a}A^{k}A^{m} \\
&+ && \alpha^{p+a\ p+b\ p+c}\omega^{ml}C^{p+c,m}A^{p+a}A^{p+b} - \tfrac{2}{3}\alpha^{lkm}A^{m\ p+a}A^{p+a}A^{k} \\
&- && [\tfrac{3}{2}\alpha^{ko\ p+a} + \tfrac{1}{4}\gamma^{p+a,mn}\omega^{mk}\omega^{no}]A^{lo}A^{p+a}A^{k} - 2\alpha^{k\ p+a\ p+b}A^{lp+b}A^{p+a}A^{k} \\
&- && [\alpha^{m\ p+a\ p+b} + \gamma^{p+a;p+b,k}\omega^{km}]A^{lm}A^{p+a}A^{p+b}, \\
R^{l}_{35} &= && [\tfrac{1}{2}\alpha^{lk\ p+a}\alpha^{mn\ p+a} - \tfrac{1}{8}\omega^{l'l}\omega^{m'm}\omega^{n'n}\omega^{k'k}\gamma^{p+a,l'm'}\gamma^{p+a,n'k'}]A^{m}A^{n}A^{k} \\
&+ && [2\alpha^{p+a\ kf}\alpha^{lmf} - \alpha^{lkm\ p+a} - \tfrac{1}{3}\alpha^{lmn}(\alpha^{kn\ p+a} - \tfrac{1}{2}\gamma^{p+a,m'n'}\omega^{m'k}\omega^{n'n}) \\
&- && \tfrac{1}{2}\omega^{m'm}\omega^{k'k}\omega^{l'l}(\omega^{l'v}\gamma^{p+a,m'l'}\gamma^{v,l'k'} - \tfrac{1}{3}\gamma^{p+a,m'k'l'})]A^{p+a}A^{k}A^{m} \\
&- && \tfrac{1}{2}[3\alpha^{lk\ p+a\ p+b} + \tfrac{2}{3}\alpha^{klv}(\alpha^{v\ p+a\ p+b} - \tfrac{1}{2}\gamma^{p+a;p+b,n}\omega^{nv}) \\
&+ && (\alpha^{kvp+a} - \tfrac{1}{2}\gamma^{p+a,mn}\omega^{mk}\omega^{nv})(\alpha^{lvp+b} - \tfrac{1}{2}\gamma^{p+b,m'n'}\omega^{m'l}\omega^{n'v})]A^{p+a}A^{p+b}A^{k} \\
&+ && [\alpha^{l\ p+a\ f}\alpha^{p+b\ p+c\ f} - \alpha^{l\ p+a\ p+b\ p+c} - (\alpha^{lk\ p+c} - \tfrac{1}{2}\gamma^{p+c,mn}\omega^{ml}\omega^{nk}) \\
&\times && (\alpha^{k\ p+a\ p+b} - \gamma^{p+a;p+b,v})\omega^{vk}]A^{p+a}A^{p+b}A^{p+c} \quad \text{and}
\end{aligned}
$$

$$
\begin{aligned}
R^{l}_{36} &= && \{2\alpha^{l\ p+a\ f}\alpha^{mp+bf} + \alpha^{lmf}\alpha^{p+a\ p+b\ f} + \omega^{m'm}\omega^{n'l}[\tfrac{1}{2}\omega^{ok}\omega^{vk}\gamma^{p+a,om'}\gamma^{p+b,vn'} \\
&- && \tfrac{1}{2}(\gamma^{p+c;p+a,m'} + \gamma^{p+a;p+c,m'})(\gamma^{p+c;p+b,n'} + \gamma^{p+b;p+c,n'}) \\
&- && \omega^{ko}\gamma^{p+b,n'k}(\gamma^{o,m';p+a} + \gamma^{o;p+a,m'}) - \tfrac{1}{2}\alpha^{p+a\ p+b\ p+c}\gamma^{p+c,m'n'} \\
&- && \tfrac{1}{2}\omega^{ko}\gamma^{o,m'n'}\gamma^{p+a;p+b,k} + \tfrac{1}{2}\gamma^{p+a;p+b,m'n'} + \tfrac{1}{2}\gamma^{p+a,n;p+b,m'}]\}A^{p+a}A^{p+b}A^{m}.
\end{aligned}
$$

A.4. PROOF OF THEOREMS 1 AND 2.

The proof of Theorem 1 is divided into two parts. In the first part, we derive the cumulants of $\sqrt{n}R$. In the second part, we establish an Edgeworth expansion for the signed root which then leads to an Edgeworth expansion for the EL ratio $r(\theta_0)$.

Cumulants of the signed root R. Since the cumulants of order higher than four are of $O(n^{-2})$ or smaller, we only need to derive the first four cumulants. As the first and the third cumulants are easier to derive than the second and the fourth, we present them first. From (2.10) and (2.11), and the fact that R_3^j is the product of four zero-mean averages, we have

$$E(R_1^l) = 0,\ E(R_2^l) = n^{-1}\mu^l \text{ and } E(R_3^j) = O(n^{-2})$$

where $\mu^l = -\frac{1}{6}n^{-1}\alpha^{lkk}$. Therefore, the first order cumulant is

$$cum(R^l) = n^{-1}\mu^l + O(n^{-2}). \tag{A.8}$$

The joint third-order cumulants

$$\begin{aligned} cum(R^l, R^o, R^v) &= E(R^l R^o R^v) - E(R^l)E(R^o R^v)[3] + 2E(R^l)E(R^o)E(R^v)[3] \\ &= E(R_1^l R_1^o R_1^v) + E(R_2^l R_1^o R_1^v)[3] - E(R_2^l)E(R_1^o R_1^v)[3] + O(n^{-3}). \end{aligned}$$

We note that

$$\begin{aligned} E(R_1^l R_1^o) &= n^{-1}\delta^{lo} \quad \text{and} \\ E(R_2^l R_1^o) &= n^{-2}[-\tfrac{1}{2}(\alpha^{lokk} - \delta^{lo}) - \alpha^{lo\ p+a\ p+a} + \tfrac{1}{3}\alpha^{lkm}\alpha^{okm} + \omega^{kl}\gamma^{p+a;o;p+a,k} \\ &+ (\alpha^{lk\ p+a} - \tfrac{1}{2}\omega^{mk}\omega^{nl}\gamma^{p+a,mn})\alpha^{ok\ p+a} + (\alpha^{l\ p+a\ p+b} - \omega^{kl}\gamma^{p+a;p+b,k})\alpha^{o\ p+a\ p+b}]. \end{aligned}$$

Write $R_2^l = R_{21}^l + R_{22}^l$ where $R_{21}^l = -\frac{1}{2}A^{kl}A^k + \frac{1}{3}\alpha^{klm}A^kA^m$ and R_{22}^l contains the rest of the terms appeared in (2.11). We have

$$\begin{aligned} E(R_{21}^l) &= -\tfrac{1}{6}n^{-1}\alpha^{lkk}, \\ E(R_{22}^l R_1^o R_1^v) &= E(R_{22}^l)\delta^{ov} + O(n^{-3}) = O(n^{-3}) \quad \text{and} \\ E(R_{21}^l R_1^o R_1^v) &= n^{-2}(-\tfrac{1}{6}\alpha^{lkk}\delta^{ov} - \tfrac{1}{3}\alpha^{lov}) + O(n^{-3}). \end{aligned}$$

Thus,

$$E(R_2^l R_1^o R_1^v) = E(R_2^l)E(R_1^o R_1^v) - \tfrac{1}{3}E(R_1^l R_1^o R_1^v) + O(n^{-3}), \tag{A.9}$$

which means that

$$cum(R^l, R^o, R^v) = O(n^{-3}). \tag{A.10}$$

To compute the second cumulants, we have to derive the expectation $R_2^l R_2^l$ which contains 21 terms. After deriving the expectation of each term and combine them as in Chen and Cui (2003), we have

$$E(R_2^l R_2^o) = n^{-2} J^{lo} + O(n^{-3}) \tag{A.11}$$

where

$$\begin{aligned} J^{lo} &= \tfrac{1}{4}(\alpha^{lokk} - \delta^{lo}) + \tfrac{1}{36}\alpha^{lkk}\alpha^{omm} - \tfrac{7}{36}\alpha^{lkm}\alpha^{okm} + \alpha^{lo\ p+a\ p+a} \\ &- \alpha^{l\ p+a\ p+b}\alpha^{o\ p+a\ p+b} - \alpha^{lk\ p+a}\alpha^{ok\ p+a} + \tfrac{1}{2}\omega^{kl}\gamma^{m;p+a,k}\alpha^{om\ p+a}[2] \\ &+ \gamma^{p+a;p+b,k}\alpha^{o\ p+a\ p+b}\omega^{kl}[2] + \tfrac{1}{4}\gamma^{p+a,mn}\omega^{mk}\omega^{nl}\alpha^{ok\ p+a}[2] \\ &+ \tfrac{1}{4}\gamma^{p+a,mn}\gamma^{p+a,m'n'}\omega^{mk}\omega^{nl}\omega^{m'k}\omega^{n'o} - \tfrac{1}{2}\gamma^{p+a,mn}\gamma^{k;p+a,v}\omega^{mk}\omega^{nl}\omega^{vo}[2] \\ &- \omega^{kl}\gamma^{o;p+a;p+a,k}[2] + [\gamma^{p+a,k;p+a,v} - \gamma^{p+a;p+b,k}\gamma^{p+a;p+b,v}]\omega^{kl}\omega^{vo}. \end{aligned} \tag{A.12}$$

We also need to compute $E(R_2^l R_1^o) + E(R_3^l R_1^o)$. It may be shown that, with remainder terms of $O(n^{-1})$,

$$\begin{aligned} n^2 E[R_{31}^l R_1^o] &= \tfrac{5}{8}\alpha^{lokk} - \tfrac{3}{8}\delta^{lo} - \tfrac{29}{72}\alpha^{lkm}\alpha^{okm} - \tfrac{1}{72}\alpha^{lok}\alpha^{kmm}, \\ n^2 E(R_{32}^l R_1^o) &= \tfrac{5}{2}\alpha^{lo\ p+a\ p+a} + \alpha^{lk\ p+a}\alpha^{ok\ p+a} + \tfrac{1}{2}\alpha^{lok}\alpha^{k\ p+a\ p+a} + \alpha^{l\ p+a\ p+b}\alpha^{o\ p+a\ p+b} \\ &+ \tfrac{1}{2}\alpha^{lo\ p+a}\alpha^{kk\ p+a} - \tfrac{1}{2}\omega^{ko}\omega^{ml}\gamma^{p+a,km;p+a} - \omega^{kl}\gamma^{p+a;p+a,k;o} \\ &+ \tfrac{1}{2}\omega^{km}\big(\gamma^{p+a,k;o}\alpha^{lm\ p+a} + \gamma^{p+a,k;p+a}\alpha^{lom}\big) - \tfrac{1}{2}\omega^{nl}\omega^{ko}\gamma^{p+a,n;p+a,k} \\ &- \tfrac{1}{2}\omega^{km}\omega^{nl}\big(\gamma^{p+a,n;m}\gamma^{p+a,k;o} + \gamma^{p+a,n;o}\gamma^{p+a,k;m}\big) + \alpha^{lo\ p+a}\alpha^{p+a\ p+b\ p+b} \\ &- \omega^{kl}\omega^{mn}\big(\gamma^{n,k;p+a}\gamma^{p+a,m;o} + \gamma^{n,k;o}\gamma^{p+a,m;p+a}\big) \\ &- \omega^{kl}\big(\gamma^{p+a,k;p+b}\alpha^{o\ p+a\ p+b} + \gamma^{p+a,k;o}\alpha^{p+a\ p+b\ p+b}\big), \\ n^2 E(R_{33}^l R_1^o) &= -\alpha^{lk\ p+a}\alpha^{ok\ p+a} - \alpha^{lop+b}\alpha^{p+a\ p+a\ p+b} + \tfrac{1}{2}\omega^{vo}\omega^{nl}\omega^{km}\gamma^{m,vn}\gamma^{p+a,k;p+a} \\ &- \tfrac{1}{3}\omega^{km}\gamma^{p+a,k;p+a}\alpha^{oml} + \tfrac{1}{2}\omega^{ko}\omega^{nl}\gamma^{p+b,kn}\alpha^{p+a\ p+a\ p+b} \\ &+ \tfrac{1}{2}\omega^{kn}\omega^{vl}\omega^{mo}\gamma^{p+b,kv}\gamma^{p+b,m;n} + \tfrac{1}{2}\omega^{ko}\omega^{nl}\omega^{mv}\gamma^{p+b,kn}\gamma^{p+b,m;v} \end{aligned}$$

$$\begin{aligned}
&- && \omega^{vk}\alpha^{lk\ p+a}\gamma^{p+a,v;o} + \omega^{vk}\omega^{nl}\big(\gamma^{k;p+a,n} + \gamma^{p+a;k,n}\big)\gamma^{p+a,v;o} \\
&+ && \tfrac{1}{2}\omega^{ml}\omega^{kn}\gamma^{p+a,mk}\alpha^{on\ p+a} + \omega^{no}\omega^{ml}\omega^{kv}\gamma^{p+a,kn}\gamma^{v,m;p+a}, \\
n^2E(R^l_{34}R^o_1) &= && -4\alpha^{l\ p+a\ p+b}\alpha^{o\ p+a\ p+b} - \tfrac{5}{3}\alpha^{lom}\alpha^{m\ p+a\ p+a} - \tfrac{7}{2}\alpha^{lk\ p+a}\alpha^{ok\ p+a} \\
&- && \alpha^{lo\ p+a}\alpha^{kk\ p+a} - \alpha^{lo\ p+a}\alpha^{p+a\ p+b\ p+b} + \omega^{ml}\omega^{nk}\gamma^{k,m;o}\gamma^{p+a;p+a,n} \\
&+ && \omega^{nl}\big(\gamma^{p+b;p+a,n} + \gamma^{p+a;p+b,n}\big)\alpha^{o\ p+a\ p+b} \\
&+ && \omega^{no}\omega^{kl}\big(\gamma^{p+b;p+a,n} + \gamma^{p+a;p+b,n}\big)\gamma^{p+b,k;p+a} + \omega^{ml}\gamma^{p+b,m;o}\alpha^{p+a\ p+a\ p+b} \\
&- && \tfrac{1}{4}\omega^{mo}\omega^{nk}\gamma^{p+a,mn}\alpha^{lk\ p+a} - \omega^{km}\gamma^{p+a;p+a,k}\alpha^{lom},
\end{aligned}$$

$$\begin{aligned}
n^2E(R^l_{35}R^o_1) &= && \tfrac{1}{2}\alpha^{lo\ p+a}\alpha^{kk\ p+a} + \alpha^{lk\ p+a}\alpha^{ok\ p+a} - \tfrac{3}{2}\alpha^{lo\ p+a\ p+a} - \tfrac{1}{3}\alpha^{lok}\alpha^{k\ p+a\ p+a} \\
&+ && \tfrac{1}{6}\omega^{nv}\alpha^{lov}\gamma^{p+a;p+a,n} - \tfrac{1}{2}\alpha^{lk\ p+a}\alpha^{ok\ p+a} + \tfrac{1}{4}\omega^{mo}\omega^{nv}\alpha^{lvp+a}\gamma^{p+a,mn}[2] \\
&- && \tfrac{1}{8}\omega^{l'l}\omega^{m'o}\omega^{n'k}\omega^{k'k}\gamma^{p+a,l'm'}\gamma^{p+a,n'k'} \\
&- && \tfrac{3}{8}\omega^{mo}\omega^{nv}\omega^{m'l}\omega^{n'v}\gamma^{p+a,mn}\gamma^{p+a,m'n'}, \\
n^2E(R^l_{36}R^o_1) &= && 2\alpha^{lk\ p+a}\alpha^{ok\ p+a} + 2\alpha^{l\ p+a\ p+b}\alpha^{o\ p+a\ p+b} + \alpha^{lok}\alpha^{k\ p+a\ p+b} \\
&+ && \alpha^{lo\ p+a}\alpha^{p+a\ p+b\ p+b} + \tfrac{1}{2}\omega^{m'o}\omega^{n'l}\omega^{mk}\omega^{vk}\gamma^{p+a,mm'}\gamma^{p+a,vn'} \\
&- && \tfrac{1}{2}\omega^{m'o}\omega^{n'l}\big(\gamma^{p+b;p+a,m'} + \gamma^{p+a;p+b,m'}\big)\big(\gamma^{p+b;p+a,n'} + \gamma^{p+a;p+b,n'}\big) \\
&- && \omega^{m'o}\omega^{n'l}\omega^{km}\big(\gamma^{m,m';p+a} + \gamma^{m;p+a,m'}\big)\gamma^{p+a,n'k} \\
&- && \tfrac{1}{2}\omega^{m'o}\omega^{n'l}\alpha^{p+a\ p+a\ p+b}\gamma^{p+b,m'n'} - \tfrac{1}{2}\omega^{m'o}\omega^{n'l}\omega^{km}\gamma^{m,m'n'}\gamma^{p+a;p+a,k} \\
&+ && \tfrac{1}{2}\omega^{m'o}\omega^{n'l}\big(\gamma^{p+a;p+a,m'n'} + \gamma^{p+a,n;p+a,m'}\big), \\
n^2E(R^l_2R^o_1) &= && -\tfrac{1}{2}(\alpha^{lokk} - \delta^{lo}) - \alpha^{lo\ p+a\ p+a} + \tfrac{1}{3}\alpha^{lkm}\alpha^{okm} + \alpha^{l\ p+a\ p+b}\alpha^{o\ p+a\ p+b} \\
&+ && \alpha^{lk\ p+a}\alpha^{ok\ p+a} + \omega^{kl}\gamma^{p+a;o;p+a,k} - \tfrac{1}{2}\omega^{mk}\omega^{nl}\gamma^{p+a,mn}\alpha^{ok\ p+a} \\
&- && \omega^{kl}\gamma^{p+a;p+b,k}\alpha^{o\ p+a\ p+b}.
\end{aligned}$$

In summary,

$$E(R^l_2R^o_1) + E(R^l_3R^o_1) = n^{-2}K^{lo} + O(n^{-3}) \tag{A.13}$$

where

$$K^{lo} = \tfrac{1}{8}(\alpha^{lokk} + \delta^{lo}) - \tfrac{5}{72}\alpha^{lkm}\alpha^{okm} - \tfrac{1}{72}\alpha^{lok}\alpha^{kmm} - \tfrac{1}{2}\alpha^{lok}\alpha^{k\ p+a\ p+a}$$

$$\begin{aligned}
&- \tfrac{1}{2}\omega^{km}\gamma^{p+a,k;o}\alpha^{lm\ p+a} - \tfrac{2}{3}\omega^{km}\gamma^{p+a,k;p+a}\alpha^{lom} + \tfrac{1}{4}\omega^{ml}\omega^{nk}\alpha^{ok\ p+a}\gamma^{p+a,mn} \\
&+ \tfrac{1}{2}\omega^{nl}\omega^{km}\big(\gamma^{p+a,k;o}\gamma^{p+a,n;m} - \gamma^{p+a,n;o}\gamma^{p+a,k;m}\big) \\
&+ \tfrac{1}{2}\omega^{kn}\omega^{vl}\omega^{mo}\big(\gamma^{p+a,mv}\gamma^{p+a,k;n} - \gamma^{p+a,kv}\gamma^{p+a,m;n}\big) \\
&+ \omega^{mo}\omega^{vl}\omega^{kn}\big(\gamma^{p+a,km}\gamma^{p+a;n,v} - \gamma^{p+a,kv}\gamma^{p+a;n,m}\big) \\
&+ \tfrac{1}{8}\omega^{m'l}\omega^{mo}\omega^{n'k}\omega^{nk}\big(\gamma^{p+a,mn}\gamma^{p+a,m'n'} - \gamma^{p+a,mm'}\gamma^{p+a,nn'}\big).
\end{aligned}$$

In light of (A.11) and (A.13), we have

$$cum(R^l, R^o) = n^{-1}\delta^{lo} + n^{-2}\Delta^{lo} + O(n^{-3}) \tag{A.14}$$

where

$$\Delta^{lo} = K^{lo}[2] + J^{lo} - \mu^l\mu^o. \tag{A.15}$$

The joint fourth-order cumulants of R is

$$\begin{aligned}
cum(R^l, R^k, R^m, R^n) &= E(R^lR^kR^mR^n) - E(R^lR^k)E(R^mR^n)[3] \\
&- E(R^l)E(R^kR^mR^n)[4] + 2E(R^l)E(R^k)E(R^mR^n)[6] \\
&- 6E(R^l)E(R^k)E(R^m)E(R^n) \\
&= E(R^l_1R^k_1R^m_1R^n_1) + E(R^l_2R^k_1R^m_1R^n_1)[4] + E(R^l_3R^k_1R^m_1R^n_1)[4] \\
&+ E(R^l_2R^k_2R^m_1R^n_1)[6] - E(R^l_1R^k_1]E(R^m_1R^n_1)[3] \\
&- E(R^l_2R^k_1)E(R^m_1R^n_1)[12] - E(R^l_3R^k_1)E(R^m_1R^n_1)[12] \\
&- E(R^l_2R^k_2)E(R^m_1R^n_1)[6] - E(R^l_2)E(R^k_1R^m_1R^n_1)[4] \\
&- E(R^l_2)E(R^k_2R^m_1R^n_1)[12] + 2E(R^l_2)E(R^k_2)E(R^m_1R^n_1)[6] + O(n^{-4}).
\end{aligned} \tag{A.16}$$

From (A.9), we immediately have

$$E(R^l_2)\{E(R^k_1R^m_1R^n_1)[4] + E(R^k_2R^m_1R^n_1)[12] - 2E(R^k_2)E(R^m_1R^n_1)[6]\} = O(n^{-4})$$

which means the sum of the last three terms in (A.16) is negligible.

To facilitate easy expressions, let us define

$$\begin{aligned}
t_1 &= \alpha^{lkmn}, t_2 = \delta^{lk}\delta^{mn} + \delta^{lm}\delta^{kn} + \delta^{ln}\delta^{km}, \\
t_3 &= \alpha^{lkm}\alpha^{noo} + \alpha^{lkn}\alpha^{moo} + \alpha^{lmn}\alpha^{koo} + \alpha^{kmn}\alpha^{loo}, \\
t_4 &= \alpha^{lko}\alpha^{mno} + \alpha^{lmo}\alpha^{kno} + \alpha^{lno}\alpha^{kmo}, \\
t_5 &= \alpha^{kmn}\alpha^{l\ p+a\ p+a} + \alpha^{lmn}\alpha^{k\ p+a\ p+a} + \alpha^{lkn}\alpha^{m\ p+a\ p+a} + \alpha^{lkm}\alpha^{n\ p+a\ p+a}, \\
t_6 &= \alpha^{lk\ p+a}\alpha^{mn\ p+a} + \alpha^{lmo\ p+a}\alpha^{kn\ p+a} + \alpha^{ln\ p+a}\alpha^{km\ p+a}.
\end{aligned}$$

It is relatively easy to show that

$$E(R_1^l R_1^k R_1^m R_1^n) - E(R_1^l R_1^k)E(R_1^m R_1^n)[3] = n^{-3}(t_1 - t_2) + O(n^{-4}). \tag{A.17}$$

Derivations in Chen and Cui (2003) show that

$$\begin{aligned}
&E(R_2^l R_1^k R_1^m R_1^n)[4] - E(R_2^l R_1^k)E(R_1^m R_1^n)[12] \\
= \;& n^{-3}\{\omega^{ol}(\gamma^{p+a,o;k}\alpha^{mn\ p+a} + \gamma^{p+a,o;m}\alpha^{kn\ p+a} + \gamma^{p+a,o;n}\alpha^{km\ p+a})][4] \\
- \;& 6t_1 + 2t_2 - \tfrac{1}{6}t_3 + \tfrac{2}{3}t_4 - [\omega^{m'l}\omega^{n'k}\gamma^{p+a,m'n'}\alpha^{mn\ p+a}][6]\} + O(n^{-4}),
\end{aligned} \tag{A.18}$$

$$\begin{aligned}
&E(R_2^l R_2^k R_1^m R_1^n)[6] - E(R_2^l R_2^k)E(R_1^m R_1^n)[6] \\
= \;& n^{-3}\{3t_1 - t_2 + \tfrac{1}{6}t_3 - \tfrac{5}{9}t_4 + [\omega^{m'm}\omega^{n'l}\gamma^{p+a,m'n'}\alpha^{kn\ p+a}][6] \\
- \;& \tfrac{1}{2}[\omega^{ol}(\gamma^{p+a,o;m}\alpha^{kn\ p+a} + \gamma^{p+a,o;n}\alpha^{km\ p+a}) \\
+ \;& \omega^{ok}(\gamma^{p+a,o;m}\alpha^{ln\ p+a} + \gamma^{p+a,o;n}\alpha^{lm\ p+a})][6] \\
+ \;& [\omega^{k'l}\omega^{m'k}(\gamma^{p+a,m';n}\gamma^{p+a,k';m} + \gamma^{p+a,k';n}\gamma^{p+a,m';m})][6] \\
- \;& \tfrac{1}{2}\gamma^{p+a,m'n'}[(\omega^{ol}\omega^{n'k} + \omega^{ok}\omega^{n'l})(\omega^{m'n}\gamma^{p+a,o;m} + \omega^{m'm}\gamma^{p+a,o;n})][6] \\
+ \;& \tfrac{1}{4}[\omega^{l'l}\omega^{m'm}(\omega^{n'k}\omega^{k'n} + \omega^{n'n}\omega^{k'k})\gamma^{p+a,l'm'}\gamma^{p+a,n'k'}][6]\} + O(n^{-4})
\end{aligned} \tag{A.19}$$

and

$$E(R_3^l R_1^k R_1^m R_1^n)[4] - E(R_3^l R_1^k)E(R_1^m R_1^n)[12]$$

$$\begin{aligned} &= n^{-3}\{2t_1 - \tfrac{1}{9}t_4 + [\omega^{n'l}\omega^{k'k}(\gamma^{p+a,k';n}\gamma^{p+a,n';m} + \gamma^{p+a,n';n}\gamma^{p+a,k';m})][6] \\ &+ \tfrac{1}{2}[\gamma^{p+a,m'n'}(\omega^{ol}\omega^{n'k} + \omega^{ok}\omega^{n'l})(\omega^{m'n}\gamma^{p+a,o;m} + \omega^{m'm}\gamma^{p+a,o;n})][6] \\ &- \tfrac{1}{4}[\omega^{l'l}\omega^{m'm}(\omega^{n'k}\omega^{k'n} + \omega^{n'n}\omega^{k'k})\gamma^{p+a,l'm'}\gamma^{p+a,n'k'}][6]\} + O(n^{-4}). \end{aligned} \tag{A.20}$$

Combining (A.17), (A.18), (A.19) and (A.20) it may be shown that

$$cum(R^l, R^k, R^m, R^n) = O(n^{-4}). \tag{A.21}$$

Edgeworth Expansion for $r(\theta_0)$. We first derive an Edgeworth expansion for the distribution of $n^{1/2}R$. Let κ_j be the j-th order joint cumulant of $n^{\frac{1}{2}}R$. From (A.8), (A.14), (A.10) and (A.21),

$$\begin{aligned} \kappa_1 &= n^{-1/2}\mu + O\left(n^{-3/2}\right), \qquad \kappa_2 = I_p + n^{-1}\Delta + O(n^{-3}), \\ \kappa_3 &= O(n^{-3/2}), \qquad \kappa_4 = O(n^{-2}), \end{aligned}$$

where I_p is the $p \times p$ identity matrix, $\mu = (\mu^1, \cdots, \mu^p)^\tau$ with $\mu^l = -\frac{1}{6}\alpha^{lkk}$ and $\Delta = \left(\Delta^{lo}\right)_{p\times p}$.

Let

$$\begin{aligned} \bar{U}_A &= \left(A^1, \cdots, A^r, A^{11}, \cdots, A^{rr}, A^{111}, \cdots, A^{rrr}\right)^\tau, \\ \bar{U}_C &= \left(C^{1,1}, \cdots, C^{1,p}, \cdots, C^{r,1}, \cdots, C^{r,p}, C^{1;1,1}, \cdots, C^{r;r,p}\right)^\tau \end{aligned}$$

and $\bar{U} = (U_A^\tau, U_C^\tau)^\tau$ is a vector of centralized means. From (2.10), (2.11) and the expansion for R_3 given in Appendix A.3, the signed square root $n^{1/2}R$ can be expressed as a smooth function of U, namely there exists a smooth function h such that $n^{1/2}R = h(\bar{U})$. We can then use the results given in Bhattacharya and Ghosh (1978) to formally establish Edgeworth expansion for the distribution of $n^{1/2}R$ under conditions (2.1). In particular, let $\mathcal{B}$ be a class of Boreal sets in R^p satisfying

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

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 $n^{\frac{1}{2}}R$ is

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

where $\pi(v) = 1+n^{-1/2}\mu^T v+\frac{1}{2}n^{-1}\{v^T\left(\mu\mu^T+\Delta\right)v-\mathrm{tr}\left(\mu\mu^T+\Delta\right)\}$, $\phi(v)$ is the p-dimensional standard normal density, and $\mathrm{tr}(\dot{})$ is the trace operation for square matrices.

Let $H = (h_{ij})_{p\times p} =: \mu\mu^T + \Delta$. By the symmetry of $\phi(v)$ we have

$$\begin{aligned} P\{\ell(\beta) < c_\alpha\} &= P\{(n^{\frac{1}{2}}R)^T(n^{\frac{1}{2}}R) < c_\alpha\} + O(n^{-3/2}) \\ &= \int_{\|v\|<c_\alpha^{1/2}} \pi(v)\phi(v)dv + O(n^{-3/2}) \\ &= P(\chi_p^2 < c_\alpha) + \tfrac{1}{2}n^{-1}\textstyle\int_{\|v\|<c_\alpha^{1/2}}\{\sum_{i=1}^p h_{ii}(v_i^2-1) + \sum_{i\neq j} h_{ij}v_iv_j\}\phi(v)d\,v \\ (\text{A.22}) \qquad & \quad +O(n^{-3/2}) \\ &= \alpha - p^{-1}B_c\, c_\alpha\, f_p(c_\alpha)\, n^{-1} + O(n^{-3/2}) \end{aligned}$$

where $B_c = \sum_{i=1}^p h_{ii} = \sum_{l=1}^p(\mu^l\mu^l + \Delta^{ll})$. Due to fact that an even order Hermit polynomial is an even function, the error term in (A.22) is in fact $O(n^{-2})$. This completes the proof.

Proof of Theorem 2. Based on the Edgeworth expansion given in Theorem 1 and follow standard derivations, for instance those given in Chen (1993), we can readily establish the theorem.

ACKNOWLEDGEMENT

The research of this paper was supported by a National University of Singapore Academic Research Grant (R-155-000-018-112) and a RFDP of China grant (20020027010). The authors thank Dr Weidong Lin for computational support, who was supported under the National University of Singapore Academic Research Grant.

REFERENCES

ANDREWS, D. W. K. (2002): Generalized method of moments estimation when a parameter is on a boundary, *Journal of Business and Economic Statistics*, 20, 530-544.

BAGGERLY, K. A. (1998): Empirical likelihood as a goodness-of-fit measure. *Biometrika*, **85**, 535-547.

BHATTACHARYA, R.N. and GHOSH, J.K. (1978): On the validity of the formal Edgeworth expansion. *The Annals of Statistics*, **6**, 434–451.

BHATTACHARYA, R.N. AND RAO, R.R. (1976): *Normal Approximation and Asymptotic Expansions.* Wiley, New York.

BROWN, B. W. AND NEWEY, W. K. (2002): "Generalized method of moments, efficient bootstrapping, and improved inference", *Journal of Business and Economic Statistics*, 20, 507-517.

CHEN S. X. (1993): On the coverage accuracy of empirical likelihood regions for linear regression model, *The Annals of the Institute of Statistical Mathematics*, **45**, 621-637.

CHEN, S. X. (1994). Empirical likelihood confidence intervals for linear regression coefficients. *Journal Multivariate Analysis.* **49**, 24-40.

CHEN, S. X. AND CUI, H. J. (2003): On the Bartlett properties of empirical likelihood with moment restrictions, Technical Report, Department of Statistics, Iowa State University.

DICICCIO, T. J., HALL, P. AND ROMANO, J. P. (1991): Empirical likelihood is Bartlett correctable. *The Annals of Statistics*, **19**, 1053-1061.

DONALD, S. G., IMBENS, G. W. AND NEWEY, W. K. (2003): Empirical likelihood estimation and consistent tests with conditional moment restrictions. *Journal of Econometrics*, 117, 55-93.

HANSEN, B. E. AND WEST, K. D. (2002): Generalized method of moments and macroeconomics, *Journal of Business and Economic Statistics*, 20, 460-469.

HANSEN, L. P. (1982): Large sample properties of generalized method of moments estimators, *Econometrica*, **50**, 1029-1054.

HALL, P. (1992): *The bootstrap and Edgeworth Expansions* Springer-Verlag.

HALL, P. AND LA SCALA, B. (1990): Methodology and algorithms of empirical likelihood. *International Statistical Reviews* **58**, 109–127.

IMBENS, G. W. (1997): One-step estimators for over-identified generalized method of moments models, *Review of Economic Studies*, **64**, 359-383.

IMBENS, G. W. (2002): Generalized method of moments and empirical likelihood equations, *Journal of Business and Economic Statistics*, 20, 493-506.

Jagannathan, R. (2002): "Generalized method of moments: Applications in finance", *Journal of Business and Economic Statistics*, 20, 470-481.

JING, B. Y. and WOOD, A. T. A. (1996). Exponential empirical likelihood is not Bartlett correctable. *Ann. Statist.* **24** 365-369.

KITAMURA, Y. (1997): Empirical likelihood methods with weakly dependent processes. *The Annals of Statistics*, **25**, 2084-2102.

KITAMURA, Y. (2001): Asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica*, **69**, 1661-1672.

KITAMURA, Y. AND STUTZER, M. (1997): An information-theoretic alternative to generalized method of moments estimation. *Econometrica*, **65**, 861-874.

KITAMURA, Y., Tripathi, G. and Ahn, H. (2002): Empirical likelihood-based inference in conditional moment restriction models. Manuscript.

Newey, W. K. and Smith, R. J. (2004): Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* to appear.

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

OWEN, A. B. (2001). *Empirical Likelihood.* Chapman and Hall.

QIN, J. AND LAWLESS, J. (1994): Empirical likelihood and general estimation equations, *The Annals of Statistics*, **22**, 300-325.

Smith, R. J. (1997): Alternative Semi-parametric likelihood approaches to generalized method of moments estimation, *Economic Journal*, **107**, 503-519.

SKOVGAARD, IB. M. (1981): Transformation of an Edgeworth expansion by a sequence of smooth functions. *Scand. J. Statist.* **8**, 207-217.

Tripathi, G. AND KITAMURA, Y. (2002): Testing conditional moment restrictions. Manuscript.

Table 1. Empirical coverage (in percentage) and averaged length of the EL confidence interval I_α, the Bartlett corrected (BC) EL interval $I_{\alpha,bc}$ and the dircet Bootstrap (BT) calibrated confidence interval $I_{\alpha,bt}$ for the normal mean example.

(a) $N(0,1)$

Nominal level		90%			95%		
Sample Size		EL	BC	BT	EL	BCBT	
20	coverage	84.66	89.40	86.88	91.12	93.04	93.34
	length	0.636	0.817	0.791	0.757	0.967	0.967
30	coverage	86.30	89.60	87.80	92.60	93.80	93.10
	length	0.552	0.662	0.640	0.659	0.789	0.781
40	coverage	86.47	89.48	87.98	93.49	95.09	94.79
	length	0.492	0.558	0.548	0.588	0.668	0.665
60	coverage	86.70	89.00	87.90	93.50	94.50	92.90
	length	0.448	0.492	0.437	0.536	0.588	0.527

(b) $N(1,2)$

Nominal level		90%			95%		
Sample Size		EL	BC	BT	EL	BC	BT
20	coverage	80.02	87.99	85.57	86.78	91.62	90.41
	length	0.656	0.898	0.854	0.781	1.061	1.084
30	coverage	82.85	89.47	88.16	88.87	93.68	93.88
	length	0.6547	0.7740	0.752	0.739	0.883	0.879
40	coverage	85.77	89.94	88.92	91.77	93.70	93.70
	length	0.496	0.590	0.572	0.592	0.703	0.701
60	coverage	87.47	90.28	89.68	92.69	94.19	94.29
	length	0.411	0.454	0.446	0.492	0.543	0.540

Table 2. Empirical coverage (in percentage) and averaged length of the EL confidence interval I_α and the Bartlett corrected (BC) EL interval $I_{\alpha,bc}$ for the panel data model (4.1)

(a) $\theta = 0.5$

Nominal level		90%			95%		
Sample Size		EL	BC	BT	EL	BC	BT
50	coverage	81.7	87.3	87.7	88.8	93.6	93.9
	length	0.603	0.723	0.733	0.736	0.876	0.880
100	coverage	87.4	89.9	89.9	94.3	95.7	95.5
	length	0.425	0.461	0.462	0.511	0.556	0.560

(b) $\theta = 0.9$

Nominal level		90%			95%		
Sample Size		EL	BC	BT	EL	BC	BT
50	coverage	84.4	89.6	89.2	90.2	94.9	95.1
	length	0.545	0.648	0.656	0.667	0.787	0.800
100	coverage	87.3	89.1	89.2	92.6	94.4	94.3
	length	0.373	0.401	0.402	0.4494	0.4847	0.487