Provided for non-commercial research and education use.
Not for reproduction, distribution or commercial use.

Volume 141 Issue 2 December 2007 ISSN 0304-4076

ELSEVIER

JOURNAL OF Econometrics

Editors:
TAKESHI AMEMIYA
A. RONALD GALLANT
JOHN F. GEWEKE
CHENG HSIAO
PETER ROBINSON
ARNOLD ZELLNER

Associate Editors:
YACINE AÏT-SAHALIA
BADI H. BALTAGI
MICHAEL BRANDT
MARCUS J. CHAMBERS
SONGNIAN CHEN
VICTOR CHERNOZHUKOV
MANFRED DEISTLER
MIGUEL A. DELGADO
JEAN-MARIE DUFOUR
SYLVIA FRUHWIRTH-SCHNATTER
ERIC GHYSELS
JOHN C. HAM
JAVIER HIDALGO
HAN HONG
YONGMIAO HONG
BO E. HONORÉ
MAXWELL L. KING
YUICHI KITAMURA
GARY KOOP
CHUNG-MING KUAN
NAOTO KUNITOMO
KAJAL LAHIRI
QI LI
TONG LI
OLIVER LINTON
ROSA L. MATZKIN
FRANZ C. PALM
DALE J. POIRIER
BENEDIKT PÖTSCHER
INGMAR PRUCHA
PETER C. REISS
ERIC RENAULT
FRANK SCHORFHEIDE
ROBIN SICKLES
FALLAW SOWELL
GERARD VAN DEN BERG
HERMAN VAN DIJK
QUANG H. VUONG
EDWARD VYTLACIL
TOM WANSBEEK
TAO ZHA

CONTENTS

Contents continued on back cover

This article was published in an Elsevier journal. The attached copy is furnished to the author for non-commercial research and education use, including for instruction at the author's institution, sharing with colleagues and providing to institution administration.

Other uses, including reproduction and distribution, or selling or licensing copies, or posting to personal, institutional or third party websites are prohibited.

In most cases authors are permitted to post their version of the article (e.g. in Word or Tex form) to their personal website or institutional repository. Authors requiring further information regarding Elsevier's archiving and manuscript policies are encouraged to visit:

http://www.elsevier.com/copyright

Available online at www.sciencedirect.com
ScienceDirect
Journal of Econometrics 141 (2007) 492–516
JOURNAL OF Econometrics
www.elsevier.com/locate/jeconom

On the second-order properties of empirical likelihood with moment restrictions

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

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

Abstract

This paper considers the second-order properties of empirical likelihood (EL) for a parameter defined by moment restrictions, which is the inferential framework of the generalized method of moments. It is shown that the EL defined for this general framework still admits the delicate second-order property of Bartlett correction. This represents a substantial extension of all the established cases of Bartlett correction for the EL. An empirical Bartlett correction is proposed, which is shown to work effectively in improving the coverage accuracy of confidence regions for the parameter.

JEL classification: C14; C30; C40

Keywords: Bartlett correction; Coverage accuracy; Empirical likelihood; Generalized method of moments

---

1. Introduction

Generalized method of moments (GMM) introduced by Hansen (1982) are 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 $\mathrm{E}\{g(X,\theta)\} = 0$ which constitutes moment restrictions on the relationship between X and θ. The power of GMM is in its allowing $r \geqslant p$, namely the number of moment restrictions (instruments) can be larger than the number of parameter, which leads to a full exploration of inference opportunities

---

[*]Corresponding author. Tel.: +1 515 294 2729; fax: +1 515 294 4040.
E-mail address: songchen@iastate.edu (S.X. Chen).

0304-4076/$ - see front matter
doi:10.1016/j.jeconom.2006.10.006

provided by the given model. There is a vast pool of literatures on GMM. Here, we only cite the reviews by 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 non-parametric 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 moment restrictions with weakly dependent observations.

There have been comprehensive studies of EL 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, Tripathi and Kitamura (2004) for conditional moment restrictions and Chen and Gao (2006) for constructing an adaptive and rate-optimal test for a regression model. Estimation and testing with conditional moment restrictions are studied in Donald et al. (2003) and Kitamura et al. (2004). They found that EL possesses 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).

Another key property of the EL is Bartlett correction, which is a delicate second-order property that implies 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 et al. (1991) for smoothed functions of means and Chen (1993, 1994) for linear regression. Baggerly (1998) showed that EL is the only member within the Cressie–Read power divergence family that is Bartlett correctable. Jing and Wood (1996) revealed that the exponentially tilted EL for the means does not admit Bartlett correction as the tilting alters the delicate second-order mechanism of EL. Recently, Chen and Cui (2006) show that EL is Bartlett correctable in the presence of a nuisance parameter with just-identified moment restrictions.

In this paper we show that the EL with over-identified moment restrictions is Bartlett correctable. The finding represents a substantial extension to all the established cases of Bartlett correction, which all consider just-identified cases in GMM. The establishment of the Bartlett correction for the just-identified case is 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 Appendix A.

2. EL for generalized moment restrictions

Let $X_1, X_2, \ldots, X_n$ 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 $r \geqslant p$ unbiased moment restrictions $g^j(x, \theta)$, $j = 1, 2, \ldots, r$, such that $\mathrm{E}[g^j(X_1, \theta_0)] = 0$ for a unique θ_0, which is the true value of θ. Let

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

We assume the following regularity conditions:

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

(ii) for any j, $1 \leqslant j \leqslant p$, 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) $\limsup_{|t| \to \infty} |\mathrm{E}[\exp\{it^{\mathrm{T}} g(X_1, \theta_0)\}]| < 1$ and $\mathrm{E}\|g(X_1, \theta_0)\|^{15} < \infty$. (2.1)

Conditions (i) and (ii) are standard requirements for establishing Wilks' theorem and higher-order Taylor expansions of the EL ratio. The first part of Condition (iii) is the Cramér's condition on the characteristic function of $g(X, \theta_0)$. Requiring a finite 15th moment is to ensure finite 5th moment for the sign square root of the EL ratio statistic, which together with Cram*é*r's condition are needed in establishing the Edgeworth expansion for the EL ratio statistic.

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

$$\Psi V^{-1/2} \mathrm{E}\left(\frac{\partial g(X_i, \theta_0)}{\partial \theta}\right) U = (\Lambda, 0)^{\mathrm{T}}_{r \times p}. \tag{2.2}$$

Here $U = (u^{kl})_{p \times p}$ is an orthogonal matrix and $\Lambda = diag(\lambda_1, \ldots, \lambda_p)$ is non-singular. Define $\Omega = (\omega^{kl})_{p \times p} =: U\Lambda^{-1}$ where $\omega^{kl} = u^{kl} \lambda_l^{-1}$. Here no summation over the subscript l is carried out due to Λ being a diagonal matrix.

Let $p_1, p_2, \ldots, p_n$ 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 $\sum_{i=1}^{n} p_i w_i(\theta) = 0$. Let $\ell(\theta) = -2\log\{L(\theta)/n^n\}$. Standard derivations in EL show

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

where $\lambda = \lambda(\theta)$ is the solution of $n^{-1}\sum_{i=1}^{n} w_i(\theta)/(1 + \lambda^{\mathrm{T}} w_i(\theta)) = 0$.

According to Qin and Lawless (1994), the maximum EL estimator $\hat{\theta}$ and its corresponding λ, denoted as $\hat{\lambda}$, are solutions of

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

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

The log EL ratio is $r(\theta) = \ell(\theta) - \ell(\hat{\theta})$.

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

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

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

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

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 et al. (1991) for the mean parameter and Chen (1993) for linear regression.

To expand $\ell(\hat{\theta})$ in the general case of $r > p$, two new systems of notations are introduced. Let $\eta = (\lambda, \theta)$, $Q(\eta) = (Q_{1n}^{\mathrm{T}}(\eta), Q_{2n}^{\mathrm{T}}(\eta))^{\mathrm{T}}$, $S_{21} = U(\Lambda, 0)$ and $S_{12} = S_{21}^{\mathrm{T}}$. Due to the early transformation in (2.2),

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

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

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

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

$$\gamma^{j,j_1\ldots j_l;k,k_1\ldots k_m;\ldots;p,p_1\ldots p_t} = \mathrm{E}\left(\frac{\partial^l w_i^j(\theta_0)}{\partial\theta^{j_1}\ldots\partial\theta^{j_l}}\frac{\partial^m w_i^k(\theta_0)}{\partial\theta^{k_1}\ldots\partial\theta^{k_m}}\cdots\frac{\partial^t w_i^p(\theta_0)}{\partial\theta^{p_1}\ldots\partial\theta^{p_t}}\right) \quad \text{and}$$

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

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

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

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

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

From now on, we fix the ranges of the superscripts $a, b, c, d \in \{1, 2, \ldots, r-p\}$, $f, g, h, i, j \in \{1, 2, \ldots, r\}$, $k, l, m, n, o \in \{1, 2, \ldots, p\}$ and $q, s, t, u \in \{1, 2, \ldots, r+p\}$. It is shown in Appendix A.2 by substituting (2.6) into (2.7) that

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

where $[2;j,i]$ indicates there are two terms by exchanging the superscripts i and j, and $[3;j,i,k]$ means three terms such that the three superscripts take turns to occupy the position of j. Expansion (2.8) for $\ell(\hat{\theta})$ is 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 correction 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 analysis.

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

$$\begin{aligned} n^{-1}r(\theta_0) = {} & A^l A^l - A^{kl} A^k A^l - 2A^{l,p+a} A^{p+a} A^l + \tfrac{2}{3}\alpha^{klm} A^k A^m A^l + 2\omega^{kl} C^{p+a,k} A^{p+a} A^l \\ & + (2\alpha^{kl,p+a} - \gamma^{p+a,mn}\omega^{mk}\omega^{nl}) A^{p+a} A^k A^l + A^{ji}(A^{hi} A^j A^h - B^{i,q} B^q B^j [2; i,j]) \\ & + 2(\alpha^{l,p+a,p+b} - \gamma^{p+a;p+b,k}\omega^{kl}) A^{p+a} A^{p+b} A^l + B^{j,u} B^{j,q} B^u B^q + 2C^{j,k} B^{j,q} B^{r+k} B^q \\ & - \gamma^{j,kl} B^{r+k} B^{r+l} B^{j,q} B^q - 2\gamma^{j,kl} B^j B^{r+l} B^{r+k,q} B^q - 2\alpha^{jih} B^j B^i B^{h,q} B^q \\ & + 2\gamma^{j;i,l}(B^j B^i B^{r+l,q} B^q + B^{r+l} B^i B^{j,q} B^q [2; j, i]) + (\gamma^{j;i,lk} + \gamma^{j,l;i,k}) B^j B^i B^{r+l} B^{r+k} \\ & + (\tfrac{1}{4}\beta^{j,uq}\beta^{j,st} - \tfrac{1}{2}\beta^{j,uq}\beta^{r+k,st}\gamma^{j,k}) B^u B^q B^s B^t + (\alpha^{jih}\beta^{h,uq} - \gamma^{j;i,l}\beta^{r+l,uq}) B^j B^i B^u B^q \\ & + (\gamma^{j,kl}\beta^{r+l,uq} - \gamma^{i;j,k}\beta^{i,uq} - \gamma^{j;i,k}\beta^{i,uq}) B^u B^q B^j B^{r+k} + \tfrac{1}{2}\gamma^{j,kl}\beta^{j,uq} B^u B^q B^{r+l} B^{r+k} \\ & - \tfrac{1}{3}\gamma^{j,klm} B^j B^{r+k} B^{r+l} B^{r+m} - 2\gamma^{j;i;h,k} B^j B^i B^h B^{r+k} + \tfrac{1}{2}\alpha^{jihg} B^j B^i B^h B^g \\ & + C^{j,kl} B^j B^{r+k} B^{r+l} - 2C^{j;i,l} B^j B^i B^{r+l} + \tfrac{2}{3} A^{jih}(B^j B^i B^h + A^j A^i A^h) \\ & - 2\alpha^{jih} A^{gh} A^j A^i A^g + \alpha^{jgf}\alpha^{ihf} A^j A^i A^h A^g - \tfrac{1}{2}\alpha^{jihg} A^j A^i A^h A^g + O_p(n^{-5/2}). \end{aligned} \tag{2.9}$$

This expansion leads to the following signed root decomposition:

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

where $R = R_1 + R_2 + R_3$ and $R_i = O_p(n^{-i/2})$ for $i = 1, 2$ and 3. By matching $R_1^l R_1^l$ with the only term $A^l A^l$ of order n^{-1} in (2.9), we have

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

Then we match $2R_1^l R_2^l$ with the terms of order $n^{-3/2}$ and obtain

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

and $R_1^j = R_2^j = 0$ for $j \in \{p+1, \ldots, r\}$. Similarly, by matching $2R_2^l R_3^l$ with the rest of the terms in (2.9) after removing terms contributing to $(R_1^l + R_2^l)(R_1^l + R_2^l)$, the form of R_3^l is given in Appendix A.2.

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

3. The second-order properties

This section considers the accuracy of the chi-square approximation to the distribution of the EL ratio statistic $r(\theta_0)$ and improving this approximation by the Bartlett correction.

We first introduce some quantities used to define the accuracy of the chi-square approximation. Let

$$\begin{aligned} J^{lo} = {} & \tfrac{1}{4}(\alpha^{lokk} - \delta^{lo}) + \tfrac{1}{36}\alpha^{lkk}\alpha^{omm} - \tfrac{7}{36}\alpha^{lkm}\alpha^{okm} + \alpha^{lo,p+a,p+a} \\ & - \alpha^{l,p+a,p+b}\alpha^{o,p+a,p+b} - \alpha^{lk,p+a}\alpha^{ok,p+a} + \tfrac{1}{2}\omega^{kl}\gamma^{m;p+a,k}\alpha^{om,p+a}[2;l,o] \\ & + \gamma^{p+a;p+b,k}\alpha^{o,p+a,p+b}\omega^{kl}[2;l,o] + \tfrac{1}{4}\gamma^{p+a,mn}\omega^{mk}\omega^{nl}\alpha^{ok,p+a}[2;l,o] \\ & + \tfrac{1}{4}\gamma^{p+a,mn}\gamma^{p+a,m'n'}\omega^{mk}\omega^{nl}\omega^{m'k}\omega^{n'o} - \tfrac{1}{2}\gamma^{p+a,mn}\gamma^{k;p+a,v}\omega^{mk}\omega^{nl}\omega^{vo}[2,l,o] \\ & - \omega^{kl}\gamma^{o;p+a;p+a,k}[2;l,o] + (\gamma^{p+a,k;p+a,v} - \gamma^{p+a;p+b,k}\gamma^{p+a;p+b,v})\omega^{kl}\omega^{vo} \quad \text{and} \end{aligned} \tag{3.1}$$

$$\begin{aligned} K^{lo} = {} & \tfrac{1}{8}(\alpha^{lokk} + \delta^{lo}) - \tfrac{5}{72}\alpha^{lkm}\alpha^{okm} - \tfrac{1}{72}\alpha^{lok}\alpha^{kmm} - \tfrac{1}{2}\alpha^{lok}\alpha^{k,p+a,p+a} \\ & - \tfrac{1}{2}\omega^{km}\gamma^{p+a,k;o}\alpha^{lm,p+a} - \tfrac{2}{3}\omega^{km}\gamma^{p+a,k;p+a}\alpha^{lom} + \tfrac{1}{4}\omega^{ml}\omega^{nk}\alpha^{ok,p+a}\gamma^{p+a,mn} \\ & + \tfrac{1}{2}\omega^{nl}\omega^{km}(\gamma^{p+a,k;o}\gamma^{p+a,n;m} - \gamma^{p+a,n;o}\gamma^{p+a,k;m}) \\ & + \tfrac{1}{2}\omega^{kn}\omega^{vl}\omega^{mo}(\gamma^{p+a,mv}\gamma^{p+a,k;n} - \gamma^{p+a,kv}\gamma^{p+a,m;n}) \\ & + \omega^{mo}\omega^{vl}\omega^{kn}(\gamma^{p+a,km}\gamma^{p+a;n,v} - \gamma^{p+a,kv}\gamma^{p+a;n,m}) \\ & + \tfrac{1}{8}\omega^{m'l}\omega^{mo}\omega^{n'k}\omega^{nk}(\gamma^{p+a,mn}\gamma^{p+a,m'n'} - \gamma^{p+a,mm'}\gamma^{p+a,nn'}). \end{aligned} \tag{3.2}$$

Let $B_c = p^{-1}(\sum_{l=1}^{p} \Delta^{ll} + \frac{1}{36}\alpha^{lkk}\alpha^{lmm})$ where $\Delta^{ll} = 2K^{ll} + J^{ll} - \mu^l\mu^l$ and $\mu^l = -\frac{1}{6}n^{-1}\alpha^{lkk}$. The accuracy of the chi-square approximation is evaluated in the following theorem.

Theorem 1. *Under Condition* (2.1),

$$\sup_{x\in R} |\mathrm{P}\{r(\theta_0) < x\} - P(\chi_p^2 < x) + n^{-1}B_c x f_p(x)| = \mathrm{O}(n^{-2}),$$

where $f_p(\cdot)$ is the density of χ_p^2 distribution.

Theorem 1 indicates that the coverage error of the EL confidence region I_α is $\mathrm{O}(n^{-1})$, 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; and (ii) the shape and the orientation of the region are naturally determined by the likelihood ratio surface, free of any subjective intervention.

It can be shown by combining the expressions of $\mathrm{E}(R_i^l R_j^l)$ given in Appendix A.3 that

$$\mathrm{E}\{r(\theta_0)\} = p(1 + B_c n^{-1}) + \mathrm{O}(n^{-2}). \tag{3.3}$$

The rationale of the Bartlett correction is to adjust the mean of the EL ratio $r(\theta_0)$ to make it agreeable with the mean of χ_p^2 to the order of n^{-1}. This mean adjustment, originally proposed by Bartlett (1937), improves the chi-square approximation to the distribution of the likelihood ratio to $\mathrm{O}(n^{-2})$. The following theorem formally establishes the Bartlett correction for the EL ratio $r(\theta_0)$.

Theorem 2. *Under Condition* (2.1),

$$\sup_{x\in R} |\mathrm{P}\{r(\theta_0) < x(1 + B_c n^{-1})\} - \mathrm{P}(\chi_p^2 < x)| = \mathrm{O}(n^{-2}).$$

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.

The Bartlett factor B_c can have a quite 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.14) that

$$B_c = p^{-1}(\tfrac{1}{2}\alpha^{llkk} - \tfrac{1}{3}\alpha^{lkm}\alpha^{lkm}), \tag{3.4}$$

which is the Bartlett factor obtained in DiCiccio et al. (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 $j \leqslant p$ and $a \leqslant r-p$ and (ii) $g^j(x,\theta) = g^j(x)$ does not depend on θ for $j = p+1,\ldots,r$. 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,

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

A special case of the second situation is considered in Cui and Yuan (2001) for a quantile with $r = p+1$.

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

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 B times and obtain $r^{*1}(\hat{\theta}),\ldots,r^{*B}(\hat{\theta})$.

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

$$\hat{\beta}_c = (Bp)^{-1}\sum_{b=1}^{B} r^{*b}(\hat{\theta}).$$

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

$$\mathrm{E}(\hat{\beta}_c|\mathscr{X}_\mathbf{n}) = (1 + B_c n^{-1})\{1 + \mathrm{O_p}(n^{-1/2})\} \tag{3.5}$$

which means that the bootstrap estimate of β_c is $\sqrt{n}$-consistent. Now a practical Bartlett corrected confidence region is $I_{\alpha,\mathrm{bc}} = \{\theta | r(\theta) \leqslant c_\alpha \hat{\beta}_c\}$.

The above use of the bootstrap to estimate β_c naturally leads ones to think of using the bootstrap to calibrate directly on the α-quantile of the EL ratio $r(\theta_0)$. Let $\hat{c}_\alpha$ be the αth quantile of the distribution of $r^*(\hat{\theta})$ given $\mathscr{X}_n$, namely, $\mathrm{P}\{r^*(\hat{\theta}) < \hat{c}_\alpha | \mathscr{X}_n\} = \alpha$. The quantile can be estimated by the $([\alpha B]+1)$th ordered value of $\{r^{*b}(\hat{\theta})\}_{b=1}^B$ where $[\cdot]$ is the integer truncation operator. We will ignore the error of the quantile estimation as it can be made as small as possible by increasing the number of bootstrap simulation B. Then a direct bootstrap confidence interval at a nominal level α is $I_{\alpha,\mathrm{bt}} = \{\theta | r(\theta) \leqslant \hat{c}_\alpha\}$.

Theorem 3. *Under Condition* (2.1),

$$P\{r(\theta_0) < c_\alpha \hat{\beta}_c\} = \alpha + O(n^{-3/2}) \quad \textit{and} \quad P\{r(\theta_0) < \hat{c}_\alpha\} = \alpha + O(n^{-3/2}).$$

The theorem indicates that the coverage errors of the two confidence regions $I_{\alpha,\mathrm{bc}}$ and $I_{\alpha,\mathrm{bt}}$ are at the same order of $n^{-3/2}$, which is one order of magnitude smaller than the original EL region I_α. However, they are both of larger order than $O(n^{-2})$ when the true B_c is used as conveyed in Theorem 2. The underlying reason for achieving a coverage error at the order of n^{-2} in Theorem 2 is due to a fact that an even/odd order Hermit polynomial is an even/odd function, which makes the $n^{-3/2}$-order term in the Edgeworth expansion vanish. However, when we used a $\sqrt{n}$-consistent estimate of β_c in $I_{\alpha,\mathrm{bc}}$, the $n^{-3/2}$-order term in the Edgeworth expansion involves no longer just Hermite polynomials. The error in the estimation of β_c has an effect at the order of $n^{-3/2}$.

4. Simulation results

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

In the first simulation study, $X_1, \ldots, X_n$ 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 $g_2(X_1, \theta) = X_1^2 - 2\theta^2 - 1$. 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):

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

for $t = 1, \ldots, 4$ and $i = 1, \ldots, n$, where $|\theta| < 1$, $\{\varepsilon_{it}\}_{t=1}^4$ and α_i are mutually independent standard normal random variables, $e_i \sim N(0, (1-\theta^2)^{-1})$ and independent of $\{\varepsilon_{it}\}_{t=1}^4$ 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, 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. Three confidence intervals are evaluated. They are I_α based on the limiting chi-square distribution, $I_{\alpha,\mathrm{bc}}$ via the empirical Bartlett correction and $I_{\alpha,\mathrm{bt}}$ based on the direct bootstrap calibration.

To gain information on the degree of over-identification in the moment restriction on the confidence intervals, we also carry out simulations for each model with only one estimating equation, which happens to be the first estimating equation, respectively. These correspond to just-identified cases, where the Bartlett factor are given explicitly in (3.4). Therefore, in additional to the above three confidence intervals, we also evaluate the theoretical Bartlett corrected interval which replaces $\hat{\beta}_c$ by the real value β_c in $I_{\alpha,\mathrm{bc}}$. We are also able to obtain the real B_c value for the over-identified case in the first simulation model, hence the true Bartlett corrected interval is performed for all cases in the first model.

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 B used is 250 and the number of simulation is 1000.

Tables 1 and 2 contain the empirical coverage and the averaged length of the 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 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 $\mathrm{N}(1,2)$ and for the panel data model. These are the situations where the Bartlett correction is needed. In all the cases considered the empirical Bartlett correction ($I_{\alpha,\mathrm{bc}}$) improves significantly the coverage of I_α. The restoration of coverage by the Bartlett correction is impressive when the sample size is small. We also observed that, as anticipated, the direct bootstrap confidence interval has similar performance with the Bartlett corrected intervals in most of the cases. However, in the

Table 1
Empirical coverage (in percentage) and averaged length of the EL confidence interval I_α, the theoretical Bartlett corrected (TBC) interval I_{tbc} by using the ture Bartlett factors, the empirical Bartlett corrected (EBC) interval $I_{\alpha,\mathrm{bc}}$ and the direct Bootstrap (BT) calibrated confidence interval $I_{\alpha,\mathrm{bt}}$ with $X \sim \mathrm{N}(\theta, \theta^2+1)$ and $\theta = 0$

Nominal level		90%				95%			
sample size		EL	TBC	EBC	BT	EL	TBC	EBC	BT
(a) With two moment restrictions									
20	coverage	84.50	89.50	90.30	87.90	91.20	94.80	93.82	93.40
	length	0.635	0.712	0.835	0.792	0.760	0.847	0.932	0.930
30	coverage	85.75	89.25	89.45	87.32	91.20	94.10	93.50	93.20
	length	0.543	0.602	0.626	0.605	0.672	0.720	0.765	0.761
40	coverage	87.65	90.10	89.80	88.76	92.50	94.80	95.60	94.80
	length	0.490	0.515	0.546	0.542	0.585	0.615	0.627	0.625
60	coverage	87.50	89.50	89.04	88.75	94.00	95.50	94.63	93.87
	length	0.448	0.465	0.425	0.413	0.540	0.560	0.547	0.532
(b) With one moment restriction									
20	coverage	86.50	88.70	90.10	89.50	92.60	94.70	94.30	93.80
	length	0.723	0.753	0.865	0.852	0.878	0.912	0.987	0.983
30	coverage	86.50	89.30	89.32	88.75	93.20	95.10	94.20	94.60
	length	0.595	0.613	0.676	0.662	0.718	0.737	0.794	0.806
40	coverage	88.10	90.20	90.10	89.30	94.0	94.50	94.60	95.20
	length	0.515	0.524	0.563	0.557	0.618	0.630	0.645	0.650
60	coverage	87.80	89.40	89.34	88.90	94.10	94.80	94.60	94.30
	length	0.464	0.472	0.452	0.445	0.552	0.567	0.558	0.550

Table 2
Empirical coverage (in percentage) and averaged length of the EL confidence interval I_α, the theoretical Bartlett corrected (TBC) interval I_{tbc} by using the true Bartlett factors, the empirical Bartlett corrected (EBC) interval $I_{\alpha,\mathrm{bc}}$ and the direct Bootstrap (BT) calibrated confidence interval $I_{\alpha,\mathrm{bt}}$ with $X \sim \mathrm{N}(\theta, \theta^2 + 1)$ and $\theta = 1$

Nominal level		90%				95%			
sample size		EL	TBC	EBC	BT	EL	TBC	EBC	BT
(a) With two moment restrictions									
20	coverage	81.40	87.50	86.45	85.48	87.00	92.10	91.80	90.20
	length	0.661	0.807	0.792	0.783	0.789	0.956	0.924	0.917
30	coverage	82.50	92.00	89.15	88.63	92.50	95.25	94.21	93.60
	length	0.610	0.705	0.683	0.677	0.707	0.794	0.782	0.764
40	coverage	85.50	89.50	88.40	88.90	91.25	94.25	93.70	93.20
	length	0.462	0.531	0.523	0.526	0.590	0.657	0.643	0.635
60	coverage	88.50	89.75	90.35	89.23	92.25	94.40	94.25	94.50
	length	0.412	0.443	0.452	0.440	0.492	0.531	0.535	0.540
(b) With one moment restriction									
20	coverage	88.90	90.35	89.75	90.10	93.00	94.20	94.50	94.35
	length	1.025	1.065	1.050	1.063	1.228	1.275	1.283	1.280
30	coverage	87.80	89.20	89.50	88.90	93.50	94.65	94.80	94.40
	length	0.843	0.865	0.869	0.860	1.010	1.036	1.041	1.030
40	coverage	88.60	89.35	89.45	89.90	94.25	94.75	94.60	94.40
	length	0.720	0.734	0.736	0.743	0.862	0.876	0.870	0.867
60	coverage	88.90	89.50	90.15	90.25	94.50	95.15	95.10	95.20
	length	0.602	0.610	0.615	0.619	0.721	0.730	0.726	0.732

normal mean model, the coverage of the direct bootstrap interval $I_{\alpha,\mathrm{bt}}$ is not as good as the Bartlett corrected interval $I_{\alpha,\mathrm{bc}}$ in most of the situations. The robust performance of the Bartlett corrected interval may be due to the fact that the estimation of the Bartlett factor β_c, which involves simple bootstrap averaging, is less variable than the bootstrap estimation of an extreme quantile of the distribution of $r(\hat{\theta})$ (see also Tables 3 and 4).

When the number of moment restriction is reduced to one, the lengths of all intervals increase which is expected as the estimation efficiency declines. The theoretical Bartlett correction is slightly better than $I_{\alpha,\mathrm{bc}}$ and $I_{\alpha,\mathrm{bt}}$ in terms of coverage accuracy and length. However, when n is larger ($n = 60$ in the first study and $n = 100$ in the second), the three second-order intervals perform almost the same.

It is noticed that the lengths of the confidence intervals for the panel data model is comparable with those given in Brown and Newey (2002). Finally, we observed 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. Discussions

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

Table 3

Empirical coverage (in percentage) and averaged length of the EL confidence interval I_α, the theoretical Bartlett corrected (TBC) interval I_{tbc} by using the true Bartlett factors (with one moment restriction only), the empirical Bartlett corrected (EBC) EL interval $I_{\alpha,\mathrm{bc}}$ and the direct Bootstrap (BT) calibrated confidence interval $I_{\alpha,\mathrm{bt}}$ for the panel data model (4.1) with $\theta = 0.5$

Nominal level		90%			95%		
sample size		EL	EBC	BT	EL	EBC	BT
(a) With two moment restrictions							
50	coverage	81.25	87.30	89.20	87.10	93.40	94.05
	length	0.917	1.325	1.384	1.088	1.425	1.504
100	coverage	84.50	89.90	89.80	90.20	93.80	94.30
	length	0.706	0.875	0.894	0.818	1.065	1.096

Nominal level		90%				95%			
sample size		EL	TBC	EBC	BT	EL	TBC	EBC	BT
(b) With one moment restriction									
50	coverage	85.70	88.50	88.75	90.10	92.25	94.10	93.82	94.30
	length	1.382	1.457	1.525	1.534	1.574	1.631	1.625	1.685
100	coverage	87.50	89.75	88.90	89.50	89.75	94.75	94.30	94.50
	length	1.162	1.182	1.175	1.180	1.302	1.320	1.335	1.343

Table 4

Empirical coverage (in percentage) and averaged length of the EL confidence interval I_α, the theoretical Bartlett corrected (TBC) interval I_{tbc} by using the true Bartlett factors (with one moment restriction only), the empirical Bartlett corrected (EBC) EL interval $I_{\alpha,\mathrm{bc}}$ and the direct Bootstrap (BT) calibrated confidence interval $I_{\alpha,\mathrm{bt}}$ for the panel data model (4.1) with $\theta = 0.9$

Nominal level		90%			95%		
sample size		EL	EBC	BT	EL	EBC	BT
(a) With two moment restrictions							
50	coverage	80.50	88.45	89.30	87.30	93.80	94.40
	length	1.596	1.819	1.823	1.765	1.950	1.936
100	coverage	81.40	89.20	88.80	89.20	93.50	94.60
	length	1.540	1.623	1.656	1.726	1.905	1.913

Nominal level		90%				95%			
sample size		EL	TBC	EBC	BT	EL	TBC	EBC	BT
(b) With one moment restriction									
50	coverage	86.50	90.20	90.30	89.40	92.30	95.10	94.70	94.80
	length	1.849	1.889	1.894	1.876	1.923	1.940	1.934	1.939
100	coverage	87.25	90.20	90.10	89.40	93.00	94.80	94.30	94.70
	length	1.808	1.829	1.820	1.813	1.910	1.928	1.920	1.925

et al. (1991) and the linear regression by Chen (1993). It shows that the Bartlett property of the EL is still preserved even in the case of over-identification. Although the Bartlett factor can admit a rather involved expression with over-identified moment restrictions, proving that the EL is Bartlett correctable provides the theoretical foundation to the proposed easily implementable empirical Bartlett correction and bootstrap calibration of the EL ratio.

Although we have focused on the coverage accuracy of the EL confidence regions, the results of this paper have implications on hypothesis tests. For testing the simple hypothesis, $\mathrm{H}_0 : \theta = \theta_0$, a size α EL test rejects H_0 if $r(\theta_0) \geqslant c_\alpha$. The Bartlett corrected EL test rejects H_0 if $r(\theta_0) \geqslant \hat{\beta}_c c_\alpha$. The latter test has more accurate size approximation than the original EL test. The maximum EL ratio $\ell(\hat{\theta})$ can be used to test over-identification restrictions $\mathrm{E}\{g(X, \theta)\} = 0$, as proposed in Qin and Lawless (1994) and Kitamura (2001), which mirrors the GMM test of Hansen (1982). The expansion given in (2.8) would be useful in studying the second-order properties of the EL over-identification restrictions test.

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:

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

$$\gamma^{j, j_1 \ldots j_l; k, k_1 \ldots k_m; \ldots; p, p_1 \ldots p_t} = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}\left(\frac{\partial^l w_i^j(\theta_0)}{\partial \theta^{j_1} \ldots \partial \theta^{j_l}} \frac{\partial^m w_i^k(\theta_0)}{\partial \theta^{k_1} \ldots \partial \theta^{k_m}} \cdots \frac{\partial^t w_i^p(\theta_0)}{\partial \theta^{p_1} \ldots \partial \theta^{p_t}} \right).$$

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 $r = p$. 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. 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 Appendix A for independent and identically distributed samples.

Acknowledgments

The authors thank two referees, Professor Gautam Tripathi and the Editor, Professor Takeshi Amemiya, for valuable comments and suggestions which improve the presentation of the paper. The research was supported by a National University of Singapore Academic Research Grant, a National Science Foundation Grant (SES-0518904) and a RFDP of China Grant (20020027010).

Appendix A

We provide some technical details on the log EL ratio $r(\theta_0)$ in Appendix A.2, the sign root decomposition in Appendix A.3, and the proofs of the two theorems in Appendix A.4. More details on these derivations can be found in Chen and Cui (2005).

A.1. Basic formulae

We first present some basic formulae which will be used throughout the derivations. Recall that $\Omega = (\omega^{kl})_{p\times p} =: U\Lambda^{-1}$ where $\omega^{kl} = u^{kl}\lambda_l^{-1}$. Since $\Gamma(\eta) = S^{-1}Q(\eta)$ where

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

it can be checked that

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

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

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

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

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

where I is the indicator function. Since

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

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

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

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

$$\begin{pmatrix} (B^{k,l}) & (B^{k,p+b}) & (B^{k,r+l}) \\ (B^{p+a,l}) & (B^{p+a,p+b}) & (B^{p+a,r+l}) \\ (B^{r+k,l}) & (B^{r+k,p+b}) & (B^{r+k,r+l}) \end{pmatrix}$$
$$= \begin{pmatrix} (\omega^{mk}C^{l,m}) & (\omega^{mk}C^{p+b,m}) & 0 \\ (A^{p+a,l}) & (A^{p+a,p+b}) & -(C^{p+a,l}) \\ (\omega^{km}[\omega^{nm}C^{l,n} - A^{ml}]) & (\omega^{km}[\omega^{nm}C^{p+b,n} - A^{m,p+b}]) & (\omega^{km}C^{m,l}) \end{pmatrix}. \tag{A.5}$$

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:

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

See Chen and Cui (2005) 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

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

Similarly, the second term

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

For the third term

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

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

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

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

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

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

$$\begin{aligned}
n^{-1}r(\theta_0) = {} & (A^j+B^j)(A^j+B^j) - A^{ji}(A^jA^i - B^jB^i) - 2C^{j,k}B^jB^{r+k} - 2\gamma^{j;i,l}B^jB^iB^{r+l} \\
& + \tfrac{2}{3}\alpha^{jih}[A^jA^iA^h + B^jB^iB^h] + \gamma^{j,kl}B^jB^{r+k}B^{r+l} + A^{ji}(A^{hi}A^jA^h - B^{i,q}B^qB^j[2;i,j]) \\
& - \beta^{j,uq}[C^{j,k}B^{r+k} - B^{r+k,s}B^s\gamma^{j,k} + B^{j,s}B^s - A^{ji}B^i]B^uB^q - 2\alpha^{jih}B^jB^iB^{h,q}B^q \\
& + B^{j,u}B^{j,q}B^uB^q + 2C^{j,k}B^{j,q}B^{r+k}B^q - \gamma^{j,kl}B^{r+k}B^{r+l}B^{j,q}B^q \\
& - 2\gamma^{j,kl}B^jB^{r+l}B^{r+k,q}B^q + 2\gamma^{j;i,l}(B^jB^iB^{r+l,q}B^q + B^{r+l}B^iB^{j,q}B^q[2;j,i]) \\
& + (\tfrac{1}{4}\beta^{j,uq}\beta^{j,st} - \tfrac{1}{2}\beta^{j,uq}\beta^{r+k,st}\gamma^{j,k})B^uB^qB^sB^t + \tfrac{1}{2}\gamma^{j,kl}\beta^{j,uq}B^uB^qB^{r+l}B^{r+k} \\
& + (\gamma^{j,kl}\beta^{r+l,uq} - \gamma^{i;j,k}\beta^{i,uq} - \gamma^{j;i,k}\beta^{i,uq})B^uB^qB^jB^{r+k} - 2\gamma^{j;i;h,k}B^jB^iB^hB^{r+k} \\
& + (\gamma^{j;i,lk} + \gamma^{j,l;i,k})B^jB^iB^{r+l}B^{r+k} - \tfrac{1}{3}\gamma^{j,klm}B^jB^{r+k}B^{r+l}B^{r+m} \\
& + \tfrac{1}{2}\alpha^{jihg}B^jB^iB^hB^g + (\alpha^{jih}\beta^{h,uq} - \gamma^{j;i,l}\beta^{r+l,uq})B^jB^iB^uB^q \\
& + C^{j,kl}B^jB^{r+k}B^{r+l} - 2C^{j;i,l}B^jB^iB^{r+l} + \tfrac{2}{3}A^{jih}(B^jB^iB^h + A^jA^iA^h) \\
& - 2\alpha^{jih}A^{gh}A^jA^iA^g + \alpha^{jgf}\alpha^{ihf}A^jA^iA^hA^g - \tfrac{1}{2}\alpha^{jihg}A^jA^iA^hA^g + \mathrm{O_p}(n^{-5/2}).
\end{aligned} \tag{A.7}$$

Now the terms appeared on the third line of the above equation cancel each other. To appropriate this, by applying the relationships implied by (A.5) together with (A.1) and (A.4),

$$\begin{aligned}
&\beta^{j,uq}[C^{j,k}B^{r+k} - B^{r+k,s}B^{s}\gamma^{j,k} + B^{j,s}B^{s} - A^{ji}B^{i}]B^{u}B^{q}\\
&\quad = [\beta^{j,uq}C^{j,k}B^{r+k} - \beta^{l,uq}\gamma^{l,k}B^{r+k,p+a}B^{p+a} - \beta^{l,uq}\gamma^{l,k}B^{r+k,r+m}B^{r+m}\\
&\qquad - \beta^{l,uq}A^{lp+a}B^{p+a} - \beta^{p+b,uq}A^{p+bp+a}B^{p+a} + \beta^{l,uq}B^{l,p+a}B^{p+a}\\
&\qquad + \beta^{p+b,uq}B^{p+b,p+a}B^{p+a} + \beta^{p+a,uq}B^{p+a,r+m}B^{r+m}]B^{p}B^{q}\\
&\quad = [\beta^{j,uq}C^{j,k}B^{r+k} - \beta^{l,uq}(B^{l,p+a} - A^{l,p+a})B^{p+a} - \beta^{l,uq}C^{l,m}B^{r+m}\\
&\qquad + \beta^{l,uq}(B^{l,p+a} - A^{l,p+a})B^{p+a} - \beta^{p+b,uq}A^{p+bp+a}B^{p+a} + \beta^{p+a,uq}B^{p+b,p+a}B^{p+a}\\
&\qquad - \beta^{p+a,uq}C^{p+a,m}B^{r+m}]B^{u}B^{q} = 0.
\end{aligned}$$

Applying again (A.5), we can express the terms appeared in the first two lines of (A.7) by

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

which leads us to (2.9).

A.3. Expansion for R_3

We subtract $R_2^l R_2^l$ 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 $R_2^l R_2^l$. Hence, the remaining terms can be written as $2R_1^l R_3^l$.

The pursuit for an expression of R_3 is done by repeatedly employing formulae (A.5) and (A.6) as well as (A.1), (A.2) and (A.4). For instance, the terms appeared in the fourth line of (2.9)

$$\begin{aligned}
&A^{ji}A^{hi}A^{j}A^{h} + B^{j,u}B^{j,q}B^{u}B^{q} - 2A^{ji}B^{i,q}B^{q}B^{j} + 2C^{j,k}B^{j,q}B^{r+k}B^{q}\\
&\quad = A^{kl}A^{km}A^{m}A^{l} + 2A^{kl}A^{k,p+a}A^{p+a}A^{l} + 2A^{l,p+a}A^{p+a,p+b}A^{p+b}A^{l}\\
&\qquad + A^{k,p+a}A^{l,p+a}A^{k}A^{l} + A^{p+al}A^{p+bl}A^{p+a}A^{p+b} + \omega^{ml}\omega^{nl}C^{p+a,m}C^{p+b,n}A^{p+a}A^{p+b}\\
&\qquad + \omega^{nl}\omega^{km}C^{p+a,n}C^{p+a,k}A^{m}A^{l} - 2\omega^{mk}C^{p+b,m}A^{p+a,k}A^{p+b}A^{p+a}\\
&\qquad - 2\omega^{mn}\omega^{kl}C^{n,k}C^{p+a,m}A^{p+a}A^{l} - 2\omega^{kl}C^{p+a,k}A^{p+a,p+b}A^{p+b}A^{l}\\
&\qquad - 2\omega^{ml}\omega^{kn}C^{p+a,k}C^{p+a,m}A^{n}A^{l},
\end{aligned}$$

and the terms in the fifth line

$$\begin{aligned}
&- \gamma^{j,kl}B^{r+k}B^{r+l}B^{j,q}B^{q} - 2\gamma^{j,kl}B^{j}B^{r+l}B^{r+k,q}B^{q}\\
&\quad = -\gamma^{m,kl}B^{p+a}B^{r+k}B^{r+l}B^{m,p+a} - \gamma^{p+b,kl}B^{p+a}B^{r+k}B^{r+l}B^{p+b,p+a}\\
&\qquad - \gamma^{p+b,kl}B^{r+k}B^{r+l}B^{r+n}B^{p+b,r+n} - 2\gamma^{p+b,kl}B^{p+a}B^{p+b}B^{r+l}B^{r+k,p+a}\\
&\qquad - 2\gamma^{p+a,kl}B^{p+a}B^{r+l}B^{r+m}B^{r+k,r+m}\\
&\quad = \gamma^{m,kl}\omega^{kn}\omega^{lo}\omega^{vm}C^{p+a,v}A^{p+a}A^{n}A^{o} + \gamma^{p+b,kl}\omega^{kn}\omega^{lo}A^{p+b,p+a}A^{p+a}A^{n}A^{o}\\
&\qquad + \gamma^{p+b,kl}\omega^{kn}\omega^{lo}\omega^{mv}C^{p+b,m}A^{n}A^{o}A^{v} - 2\gamma^{p+b,kl}\omega^{kn}\omega^{lo}\omega^{vn}C^{p+a,v}A^{p+a}A^{p+b}A^{o}\\
&\qquad + 2\gamma^{p+b,kl}\omega^{kn}\omega^{lo}A^{n,p+a}A^{p+a}A^{p+b}A^{o} + 2\gamma^{p+a,kl}\omega^{kv}\omega^{ln}\omega^{mo}C^{v,m}A^{p+a}A^{n}A^{o},
\end{aligned}$$

and so on for the other terms.

Let us define

$$\begin{aligned}
R^l_{31} ={}& \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^l_{32} ={}& 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^l_{33} ={}& \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 \\
&+ [\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^kA^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^mA^nA^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^kA^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}
\end{aligned}$$

and

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

It may be shown after some algebra that $R^l_3 = \sum_{i=1}^{6} R^l_{3i}$.

A.4. Proof of the Theorems

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 $\mathrm{O}(n^{-2})$ 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

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

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

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

The joint third-order cumulants

$$\begin{aligned} cum(R^l,R^o,R^v) &= \mathrm{E}(R^lR^oR^v)-\mathrm{E}(R^l)\mathrm{E}(R^oR^v)[3;l,o,v]+2\mathrm{E}(R^l)\mathrm{E}(R^o)\mathrm{E}(R^v)\\ &= \mathrm{E}(R_1^lR_1^oR_1^v)+\mathrm{E}(R_2^lR_1^oR_1^v)[3;l,o,v]-\mathrm{E}(R_2^l)\mathrm{E}(R_1^oR_1^v)[3;l,o,v]\\ &\quad +\mathrm{O}(n^{-3}). \end{aligned}$$

We note that

$$\begin{aligned} &\mathrm{E}(R_1^lR_1^o)=n^{-1}\delta^{lo}\quad \text{and}\\ &\mathrm{E}(R_2^lR_1^o)=n^{-2}[-\tfrac{1}{2}(\alpha^{lokk}-\delta^{lo})-\alpha^{lo,p+a,p+a}+\tfrac{1}{3}\alpha^{lkm}\alpha^{okm}+\omega^{kl}\gamma^{p+a;o;p+a,k}\\ &\qquad +(\alpha^{lk,p+a}-\tfrac{1}{2}\omega^{mk}\omega^{nl}\gamma^{p+a,mn})\alpha^{ok,p+a}+(\alpha^{l,p+a,p+b}-\omega^{kl}\gamma^{p+a;p+b,k})\alpha^{o,p+a,p+b}]. \end{aligned}$$

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

$$\begin{aligned} &\mathrm{E}(R_{21}^l)=-\tfrac{1}{6}n^{-1}\alpha^{lkk},\\ &\mathrm{E}(R_{22}^lR_1^oR_1^v)=\mathrm{E}(R_{22}^l)\delta^{ov}+\mathrm{O}(n^{-3})=\mathrm{O}(n^{-3})\quad \text{and}\\ &\mathrm{E}(R_{21}^lR_1^oR_1^v)=n^{-2}(-\tfrac{1}{6}\alpha^{lkk}\delta^{ov}-\tfrac{1}{3}\alpha^{lov})+\mathrm{O}(n^{-3}). \end{aligned}$$

Thus,

$$\mathrm{E}(R_2^lR_1^oR_1^v)=\mathrm{E}(R_2^l)\mathrm{E}(R_1^oR_1^v)-\tfrac{1}{3}\mathrm{E}(R_1^lR_1^oR_1^v)+\mathrm{O}(n^{-3}), \tag{A.9}$$

which means that

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

To compute the second cumulants, we have to derive the expectation $R_2^lR_2^l$ which involves 21 terms. After deriving the expectation of each term and combine them as given in Chen and Cui (2005), we have

$$\mathrm{E}(R_2^lR_2^o)=n^{-2}J^{lo}+\mathrm{O}(n^{-3}), \tag{A.11}$$

where J^{lo} is defined in (3.1).

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

$$\begin{aligned}
n^2\mathrm{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\mathrm{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}\\
&\quad + \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}\\
&\quad + \tfrac{1}{2}\omega^{km}(\gamma^{p+a,k;o}\alpha^{lm,p+a} + \gamma^{p+a,k;p+a}\alpha^{lom}) - \tfrac{1}{2}\omega^{nl}\omega^{ko}\gamma^{p+a,n;p+a,k}\\
&\quad - \tfrac{1}{2}\omega^{km}\omega^{nl}(\gamma^{p+a,n;m}\gamma^{p+a,k;o} + \gamma^{p+a,n;o}\gamma^{p+a,k;m}) + \alpha^{lo,p+a}\alpha^{p+a,p+b,p+b}\\
&\quad - \omega^{kl}\omega^{mn}(\gamma^{n,k;p+a}\gamma^{p+a,m;o} + \gamma^{n,k;o}\gamma^{p+a,m;p+a})\\
&\quad - \omega^{kl}(\gamma^{p+a,k;p+b}\alpha^{o,p+a,p+b} + \gamma^{p+a,k;o}\alpha^{p+a,p+b,p+b}),\\
n^2\mathrm{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}\\
&\quad - \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}\\
&\quad + \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}\\
&\quad - \omega^{vk}\alpha^{lk,p+a}\gamma^{p+a,v;o} + \omega^{vk}\omega^{nl}(\gamma^{k;p+a,n} + \gamma^{p+a;k,n})\gamma^{p+a,v;o}\\
&\quad + \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^2\mathrm{E}(R_{34}^l R_1^o) &= -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}\\
&\quad - \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}\\
&\quad + \omega^{nl}(\gamma^{p+b;p+a,n} + \gamma^{p+a;p+b,n})\alpha^{o,p+a,p+b}\\
&\quad + \omega^{no}\omega^{kl}(\gamma^{p+b;p+a,n} + \gamma^{p+a;p+b,n})\gamma^{p+b,k;p+a} + \omega^{ml}\gamma^{p+b,m;o}\alpha^{p+a,p+a,p+b}\\
&\quad - \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}$$

$$\begin{aligned}
n^2\mathrm{E}(R_{35}^l R_1^o) &= \tfrac{1}{2}\alpha^{lo,p+a}\alpha^{kk,p+a} + \alpha^{lk,p+a}\alpha^{ok,p+a} - \tfrac{3}{2}\alpha^{lo,p+a,p+a} - \tfrac{1}{3}\alpha^{lok}\alpha^{k,p+a,p+a}\\
&\quad + \tfrac{1}{6}\omega^{nv}\alpha^{lov}\gamma^{p+a;p+a,n} - \tfrac{1}{2}\alpha^{lk,p+a}\alpha^{ok,p+a} + \tfrac{1}{4}\omega^{mo}\omega^{nv}\alpha^{lvp+a}\gamma^{p+a,mn}\\
&\quad - \tfrac{1}{8}\omega^{l'l}\omega^{m'o}\omega^{n'k}\omega^{k'k}\gamma^{p+a,l'm'}\gamma^{p+a,n'k'}\\
&\quad - \tfrac{3}{8}\omega^{mo}\omega^{nv}\omega^{m'l}\omega^{n'v}\gamma^{p+a,mn}\gamma^{p+a,m'n'},\\
n^2\mathrm{E}(R_{36}^l R_1^o) &= 2\alpha^{lk,p+a}\alpha^{ok,p+a} + 2\alpha^{l,p+a,p+b}\alpha^{o,p+a,p+b} + \alpha^{lok}\alpha^{k,p+a,p+b}\\
&\quad + \alpha^{lo,p+a}\alpha^{p+a,p+b,p+b} + \tfrac{1}{2}\omega^{m'o}\omega^{n'l}\omega^{mk}\omega^{vk}\gamma^{p+a,mm'}\gamma^{p+a,vn'}\\
&\quad - \tfrac{1}{2}\omega^{m'o}\omega^{n'l}(\gamma^{p+b;p+a,m'} + \gamma^{p+a;p+b,m'})(\gamma^{p+b;p+a,n'} + \gamma^{p+a;p+b,n'})\\
&\quad - \omega^{m'o}\omega^{n'l}\omega^{km}(\gamma^{m,m';p+a} + \gamma^{m;p+a,m'})\gamma^{p+a,n'k}\\
&\quad - \tfrac{1}{2}\omega^{m'o}\omega^{n'l}\alpha^{p+a,p+a,p+b}\gamma^{p+b,m'n'} - \tfrac{1}{2}\omega^{m'o}\omega^{n'l}\omega^{km}\gamma^{m,m'n'}\gamma^{p+a;p+a,k}\\
&\quad + \tfrac{1}{2}\omega^{m'o}\omega^{n'l}(\gamma^{p+a;p+a,m'n'} + \gamma^{p+a,n;p+a,m'}),\\
n^2\mathrm{E}(R_2^l R_1^o) &= -\tfrac{1}{2}(\alpha^{lokk} - \delta^{lo}) - \alpha^{lo,p+a,p+a} + \tfrac{1}{3}\alpha^{lkm}\alpha^{okm} + \alpha^{l,p+a,p+b}\alpha^{o,p+a,p+b}\\
&\quad + \alpha^{lk,p+a}\alpha^{ok,p+a} + \omega^{kl}\gamma^{p+a;o;p+a,k} - \tfrac{1}{2}\omega^{mk}\omega^{nl}\gamma^{p+a,mn}\alpha^{ok,p+a}\\
&\quad - \omega^{kl}\gamma^{p+a;p+b,k}\alpha^{o,p+a,p+b}.
\end{aligned}$$

In summary,

$$E(R_2^l R_1^o) + E(R_3^l R_1^o) = n^{-2} K^{lo} + O(n^{-3}), \tag{A.12}$$

where K^{lo} is defined in (3.2).

In light of (A.11) and (A.12), we have

$$cum(R^l, R^o) = n^{-1}\delta^{lo} + n^{-2}\Delta^{lo} + O(n^{-3}), \tag{A.13}$$

where

$$\Delta^{lo} = K^{lo}[2; l, o] + J^{lo} - \mu^l \mu^o. \tag{A.14}$$

The joint fourth-order cumulants of R is

$$\begin{aligned} cum(R^l, R^k, R^m, R^t) &= E(R^l R^k R^m R^t) - E(R^l R^k) E(R^m R^t)[3; l, k; m, t] \\ &\quad - E(R^l) E(R^k R^m R^t)[4; l, k, m, t] + 2E(R^l)E(R^k)E(R^m R^t)[6] \\ &\quad - 6E(R^l)E(R^k)E(R^m)E(R^t) \\ &= E(R_1^l R_1^k R_1^m R_1^t) + E(R_2^l R_1^k R_1^m R_1^t) + E(R_3^l R_1^k R_1^m R_1^t) \\ &\quad + E(R_2^l R_2^k R_1^m R_1^t)[6] - E(R_1^l R_1^k]E(R_1^m R_1^t)[3; l, k; m, t] \\ &\quad - E(R_2^l R_1^k)E(R_1^m R_1^t)[12] - E(R_3^l R_1^k)E(R_1^m R_1^t)[12] \\ &\quad - E(R_2^l R_2^k)E(R_1^m R_1^t)[6] - E(R_2^l)E(R_1^k R_1^m R_1^t)[4] \\ &\quad - E(R_2^l)E(R_2^k R_1^m R_1^t)[12] + 2E(R_2^l)E(R_2^k)E(R_1^m R_1^t)[6] \\ &\quad + O(n^{-4}). \end{aligned} \tag{A.15}$$

In the above, [6] means six terms by choosing two superscripts from a total of four; and [12] implies 12 terms by choosing l from the four superscripts first and then choosing k from the rest of the three. From now on, we will just use, say, [6] without the details of the superscript rotations to save space.

From (A.9), we immediately have

$$E(R_2^l)\{E(R_1^k R_1^m R_1^t)[4] + E(R_2^k R_1^m R_1^t)[12] - 2E(R_2^k)E(R_1^m R_1^t)[6]\} = O(n^{-4})$$

which means the sum of the last three terms in (A.15) is negligible.

To facilitate easy expressions, let us define

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

It is relatively easy to show that

$$E(R_1^l R_1^k R_1^m R_1^t) - E(R_1^l R_1^k)E(R_1^m R_1^t)[3; l, k, m, t] = n^{-3}(t_1 - t_2) + O(n^{-4}). \tag{A.16}$$

Derivations in Chen and Cui (2005) show that

$$\begin{aligned}
&\mathrm{E}(R_2^l R_1^k R_1^m R_1^t)[4] - \mathrm{E}(R_2^l R_1^k)\mathrm{E}(R_1^m R_1^t)[12]\\
&\quad = n^{-3}\{\omega^{ol}(\gamma^{p+a,o;k}\alpha^{mn,p+a} + \gamma^{p+a,o;m}\alpha^{kn,p+a} + \gamma^{p+a,o;n}\alpha^{km,p+a})][4]\\
&\qquad - 6t_1 + 2t_2 - \tfrac{1}{6}t_3 + \tfrac{2}{3}t_4 - [\omega^{m'l}\omega^{n'k}\gamma^{p+a,m'n'}\alpha^{mn,p+a}][6]\} + \mathrm{O}(n^{-4}),
\end{aligned} \tag{A.17}$$

$$\begin{aligned}
&\mathrm{E}(R_2^l R_2^k R_1^m R_1^t)[6] - \mathrm{E}(R_2^l R_2^k)\mathrm{E}(R_1^m R_1^t)[6]\\
&\quad = n^{-3}\{3t_1 - t_2 + \tfrac{1}{6}t_3 - \tfrac{5}{9}t_4 + [\omega^{m'm}\omega^{n'l}\gamma^{p+a,m'n'}\alpha^{kn,p+a}][6]\\
&\qquad - \tfrac{1}{2}[\omega^{ol}(\gamma^{p+a,o;m}\alpha^{kn,p+a} + \gamma^{p+a,o;n}\alpha^{km,p+a})\\
&\qquad + \omega^{ok}(\gamma^{p+a,o;m}\alpha^{ln,p+a} + \gamma^{p+a,o;n}\alpha^{lm,p+a})][6]\\
&\qquad + [\omega^{k'l}\omega^{m'k}(\gamma^{p+a,m';n}\gamma^{p+a,k';m} + \gamma^{p+a,k';n}\gamma^{p+a,m';m})][6]\\
&\qquad - \tfrac{1}{2}\gamma^{p+a,m'n'}[(\omega^{ol}\omega^{n'k} + \omega^{ok}\omega^{n'l})(\omega^{m'n}\gamma^{p+a,o;m} + \omega^{m'm}\gamma^{p+a,o;n})][6]\\
&\qquad + \tfrac{1}{4}[\omega^{l'l}\omega^{m'm}(\omega^{n'k}\omega^{k'n} + \omega^{n'n}\omega^{k'k})\gamma^{p+a,l'm'}\gamma^{p+a,n'k'}][6]\} + \mathrm{O}(n^{-4})
\end{aligned} \tag{A.18}$$

and

$$\begin{aligned}
&\mathrm{E}(R_3^l R_1^k R_1^m R_1^t)[4] - \mathrm{E}(R_3^l R_1^k)\mathrm{E}(R_1^m R_1^t)[12]\\
&\quad = n^{-3}\{2t_1 - \tfrac{1}{9}t_4 + [\omega^{n'l}\omega^{k'k}(\gamma^{p+a,k';n}\gamma^{p+a,n';m} + \gamma^{p+a,n';n}\gamma^{p+a,k';m})][6]\\
&\qquad + \tfrac{1}{2}[\gamma^{p+a,m'n'}(\omega^{ol}\omega^{n'k} + \omega^{ok}\omega^{n'l})(\omega^{m'n}\gamma^{p+a,o;m} + \omega^{m'm}\gamma^{p+a,o;n})][6]\\
&\qquad - \tfrac{1}{4}[\omega^{l'l}\omega^{m'm}(\omega^{n'k}\omega^{k'n} + \omega^{n'n}\omega^{k'k})\gamma^{p+a,l'm'}\gamma^{p+a,n'k'}][6]\} + \mathrm{O}(n^{-4}).
\end{aligned} \tag{A.19}$$

Combining (A.16)–(A.19) it may be shown that

$$cum(R^l, R^k, R^m, R^t) = \mathrm{O}(n^{-4}). \tag{A.20}$$

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 jth order joint cumulant of $n^{1/2}R$. From (A.8), (A.13), (A.10) and (A.20),

$$\kappa_1 = n^{-1/2}\mu + \mathrm{O}(n^{-3/2}), \quad \kappa_2 = I_p + n^{-1}\Delta + \mathrm{O}(n^{-3}),$$
$$\kappa_3 = \mathrm{O}(n^{-3/2}), \quad \kappa_4 = \mathrm{O}(n^{-2}),$$

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

Let

$$\bar{U}_A = (A^1, \ldots, A^r, A^{11}, \ldots, A^{rr}, A^{111}, \ldots, A^{rrr})^{\mathrm{T}},$$
$$\bar{U}_C = (C^{1,1}, \ldots, C^{1,p}, \ldots, C^{r,1}, \ldots, C^{r,p}, C^{1;1,1}, \ldots, C^{r;r,p})^{\mathrm{T}}$$

and $\bar{U} = (U_A^{\mathrm{T}}, U_C^{\mathrm{T}})^{\mathrm{T}}$ 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 Condition (2.1). In particular, let $\mathscr{B}$ be a class of Boreal sets in R^p satisfying

$$\sup_{B\in\mathscr{B}} \int_{(\partial B)^{\varepsilon}} \phi(v)\,\mathrm{d}v = \mathrm{O}(\varepsilon), \quad \varepsilon \downarrow 0,$$

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

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

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

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

$$\begin{aligned}
\mathrm{P}\{\ell(\beta)<c_\alpha\} &= \mathrm{P}\{(n^{1/2}R)^{\mathrm{T}}(n^{1/2}R)<c_\alpha\}+\mathrm{O}(n^{-2})\\
&= \int_{\|v\|<c_\alpha^{1/2}} \pi(v)\phi(v)\,\mathrm{d}v+\mathrm{O}(n^{-2})\\
&= \mathrm{P}(\chi_p^2<c_\alpha)+\frac{1}{2}n^{-1}\int_{\|v\|<c_\alpha^{1/2}}\left\{\sum_{i=1}^{p}h_{ii}(v_i^2-1)+\sum_{i\neq j}h_{ij}v_iv_j\right\}\phi(v)\,\mathrm{d}v\\
&\quad +\mathrm{O}(n^{-2})\\
&= \alpha - B_c c_\alpha f_p(c_\alpha)n^{-1}+\mathrm{O}(n^{-2}),
\end{aligned}\tag{A.21}$$

where $B_c=\sum_{i=1}^p h_{ii}=\sum_{l=1}^p(\mu^l\mu^l+\Delta^{ll})$. We note in particular that the remainder term in (A.21) is $\mathrm{O}(n^{-2})$ rather than $\mathrm{O}(n^{-3/2})$. This is due to a fact that an even/odd order Hermit polynomial is an even/odd function. This completes the proof.

Proof of Theorem 2. Based on the Edgeworth expansion given in (A.21),

$$\begin{aligned}
&\mathrm{P}\{\ell(\beta)<c_\alpha(1+n^{-1}B_c)\}\\
&\quad= \mathrm{P}\{\chi_p^2<c_\alpha(1+n^{-1}B_c)\}-B_c c_\alpha f_p\{c_\alpha(1+n^{-1}B_c)\}n^{-1}+\mathrm{O}(n^{-2})\\
&\quad= \alpha+\mathrm{O}(n^{-2})
\end{aligned}$$

since the chi-square distribution satisfies

$$\mathrm{P}\{\chi_p^2<c_\alpha(1+n^{-1}B_c)\}=\mathrm{P}(\chi_p^2<c_\alpha)+B_c c_\alpha f_p(c_\alpha)n^{-1}+\mathrm{O}(n^{-2})$$

and $f_p\{c_\alpha(1+n^{-1}B_c)\}=f_p(c_\alpha)+\mathrm{O}(n^{-1})$. This proves the theorem. □

Proof of Theorem 3. Note that the bootstrap estimate for $\beta_c=1+B_cn^{-1}$ is $\sqrt{n}$-consistent. Applying the delta method (Bhattacharya and Ghosh, 1978) on top of the derivation in the proof of Theorem 2, we have

$$\begin{aligned}
\mathrm{P}\{\ell(\beta)<c_\alpha\hat{\beta}_c\} &= \mathrm{P}\{\ell(\beta)<c_\alpha(1+B_cn^{-1})\}+\mathrm{O}(n^{-3/2})\\
&= \alpha+\mathrm{O}(n^{-2})+\mathrm{O}(n^{-3/2})=\alpha+\mathrm{O}(n^{-3/2}).
\end{aligned}\tag{A.22}$$

We note here that the $\mathrm{O}(n^{-3/2})$ term is entirely due to using a $\sqrt{n}$-consistent estimator of β_c. This completes the first part of the theorem.

Let R^* be the sign root decomposition of $r^*(\hat{\theta})$. It can be shown that its expression is given by replacing in R, the sign root of $r(\theta)$: (i) all the population moments by their sample averages based on χ_n and (ii) all A, B and C's by their corresponding quantities based on the bootstrap resample $\{X_i^*\}_{i=1}^n$. As shown in Hall (1992), the cumulants of R^* have forms which essentially replace all those population moments by their sample counterparts based on the original sample χ. These lead to, for any $x^* \in R$,

$$\mathrm{P}\{r^*(\hat{\theta}) < x^* | \chi_n\} = \mathrm{P}(\chi_p^2 < x^*) - p^{-1}\hat{B}_c x^* f_p(x^*) + \mathrm{O_p}(n^{-2}), \tag{A.23}$$

where $\hat{B}_c$ is a version of B_c by replacing all the population moments by their sample counterparts. It may be shown that under the conditions of the theorem, $\hat{B}_c = B_c + \mathrm{O_p}(n^{-1/2})$.

As $\hat{c}_\alpha$ satisfies $\mathrm{P}\{r^*(\hat{\theta}) < \hat{c}_\alpha | \chi_n\} = \alpha$, we can carry out a Cornish–Fisher type expansion of $\hat{c}_\alpha$ based on (A.23) and obtain

$$\hat{c}_\alpha = c_\alpha(1 + B_c n^{-1}) + \mathrm{O_p}(n^{-3/2}).$$

This together with the Edgeworth expansion for $r(\theta)$ given in the proof of Theorem 1 implies that

$$\begin{aligned}\mathrm{P}\{r(\theta) < \hat{c}_\alpha\} &= \mathrm{P}\{r(\theta) < c_\alpha(1 + B_c n^{-1})\} + \mathrm{O_p}(n^{-3/2}) \\ &= \alpha + \mathrm{O}(n^{-3/2})\end{aligned}$$

by repeating the derivation of (A.22). In the first equation above, we use again the delta method whose validity is given in Bhattacharya and Ghosh (1978). □

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.

Bartlett, M.S., 1937. Properties of sufficiency and statistical tests. Proceeding of the Royal Society A 160, 268–282.

Bhattacharya, R.N., Ghosh, J.K., 1978. On the validity of the formal Edgeworth expansion. Annals of Statistics 6, 434–451.

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

Brown, B.W., 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. Annals of the Institute of Statistical Mathematics 45, 621–637.

Chen, S.X., 1994. Empirical likelihood confidence intervals for linear regression coefficients. Journal of Multivariate Analysis 49, 24–40.

Chen, S.X., Cui, H.J., 2005. On the Bartlett properties of empirical likelihood with moment restrictions. Technical Report, Department of Statistics, Iowa State University, Iowa. ⟨http://www.public.iastate.edu/~songchen/workingpapers.html/⟩.

Chen, S.X., Cui, H.J., 2006. On Bartlett correction of empirical likelihood in the presence of nuisance parameters. Biometrika 93, 215–220.

Chen, S.X., Gao, J., 2006. An adaptive empirical likelihood test for parametric time series regression models. Working Paper, Iowa State University, Iowa.

Cui, H.J., Yuan, X.J., 2001. Smoothed empirical likelihood confidence intervals for quantile in the partial symmetric auxiliary information. Journal of System Science and Mathematical Sciences 21, 172–181.

DiCiccio, T.J., Hall, P., Romano, J.P., 1991. Empirical likelihood is Bartlett correctable. Annals of Statistics 19, 1053–1061.

Donald, S.G., Imbens, G.W., Newey, W.K., 2003. Empirical likelihood estimation and consistent tests with conditional moment restrictions. Journal of Econometrics 117, 55–93.

Hansen, B.E., 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, New York.

Hall, P., 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.

Jing, B.Y., Wood, A.T.A., 1996. Exponential empirical likelihood is not Bartlett correctable. Annals of Statistics 24, 365–369.

Kitamura, Y., 1997. Empirical likelihood methods with weakly dependent processes. Annals of Statistics 25, 2084–2102.

Kitamura, Y., 2001. Asymptotic optimality of empirical likelihood for testing moment restrictions. Econometrica 69, 1661–1672.

Kitamura, Y., Stutzer, S.M., 1997. An information-theoretic alternative to generalized method of moments estimation. Econometrica 65, 861–874.

Kitamura, Y., Tripathi, G., Ahn, H., 2004. Empirical likelihood-based inference in conditional moment restriction models. Econometrica 72, 1667–1714.

Newey, W.K., Smith, R.J., 2004. Higher order properties of GMM and generalized empirical likelihood estimators. Econometrica 72, 219–255.

Owen, A.B., 1988. Empirical likelihood ratio confidence intervals for a single functional. Biometrika 75, 237–249.

Owen, A.B., 2001. Empirical Likelihood. Chapman & Hall, London.

Qin, J., Lawless, J., 1994. Empirical likelihood and general estimation equations. Annals of Statistics 22, 300–325.

Skovgaard, I.B.M., 1981. Transformation of an Edgeworth expansion by a sequence of smooth functions. Scandinavian Journal of Statistics 8, 207–217.

Smith, R.J., 1997. Alternative semi-parametric likelihood approaches to generalized method of moments estimation. Economic Journal 107, 503–519.

Tripathi, G., Kitamura, Y., 2004. Testing conditional moment restrictions. Annals of Statistics 31, 2059–2095.