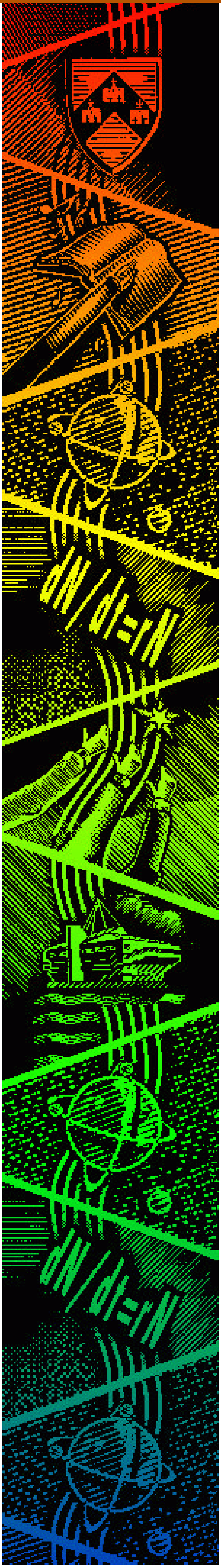

THE UNIVERSITY of York

Discussion Papers in Economics

No. 2000/05

Empirical Likelihood Inference with Applications to Some Econometric Models

by

Francesco Bravo

Department of Economics and Related Studies
University of York
Heslington
York, YO10 5DD

Empirical Likelihood Based Inference with Applications to Some Econometric Models*

Francesco Bravo†

University of York

Abstract

In this paper we analyse the higher order asymptotic properties of the empirical likelihood ratio test, by means of the dual likelihood theory. It is shown that when the econometric model is just identified, these tests are accurate to an order $o(1/n)$, and this accuracy can always be improved to an order $O\left(1/n^2\right)$ by means of a scale correction, as in standard parametric theory. To show this, we first develop a valid Edgeworth expansion for the empirical likelihood ratio test under a local alternative in terms of an "induced" local alternative. As a by-product of the expansion, we find an explicit expression for the Bartlett correction in terms of cumulants of dual likelihood derivatives which is slightly different from the standard adjustment reported in the literature on Bartlett corrections of the empirical likelihood ratio. We then highlight the connection

* This paper is based on Chapter 2 of my Ph. D. dissertation at the University of Southampton. Partial financial support under E.S.R.C. Grant R00429634019 is gratefully acknowledged.
I would like to thank Grant Hillier for many stimulating conversations. Thanks also to Andrew Chesher, Peter Phillips and Jan Podivisnky for some useful suggestions and comments. All remaining errors are my own responsibility.

† Address for correspondence. Francesco Bravo, Department of Economics and Related Studies, University of York, York, YO10 5DD, United kingdom. Email: fb6@york.ac.uk

between the empirical likelihood method and the bootstrap by obtaining a valid Edgeworth expansion for a bootstrap based empirical likelihood ratio test. The theory is then applied to some standard econometric models and illustrated by means of some Monte Carlo simulations.

1 Introduction

The method of empirical likelihood is introduced in Owen (1988) as a semiparametric likelihood technique for testing hypothesis and (by inversion) building con¯dence regions for a vector of parameters characterising a given nonparametric (i.e. distribution free) statistical model. It can be cast in the theory of least favourable family (Stein, 1956) developed for the bootstrap by Efron (1981), who showed that nonparametric inference problems can be reduced to parametric ones by applying parametric techniques to an appropriate smooth sub-family of distributions (assumed to contain the true unknown distribution F generating the data) supported on the sample. The parametric subfamily used by the empirical likelihood is asymptotically least favourable (i.e. the information of the resulting parametric subfamily at the true distribution is no greater than for the original nonparametric one), being in fact a multinomial likelihood assumed to have atoms at the observed data (see Section 2, below). This fact implies that in terms of distributional approximation, empirical and parametric likelihood are very similar, as it will become clear in the remainder of the paper. By pro¯ling this multinomial likelihood with respect to certain constraints representing the only information available about F, one gets pro¯led or implied (using Back and Brown's (1993) terminology) probabilities which can then be used to construct a nonparametric likelihood ratio test which shares many higher order asymptotics properties of its fully parametric analog, such as Bartlett correctabilty as shown in DiCiccio, Hall and Romano (1991).

In this paper we make the following contributions: ¯rstly we develop a valid Edgeworth expansion for the empirical likelihood ratio test under a local alternative in terms of an \induced" local alternative. As a by-product of this expansion, we ¯nd an explicit expression for the Bartlett correction under the null hypothesis. Secondly, we emphasise the connection between the empirical likelihood method and the bootstrap by obtaining a valid Edgeworth expansion for a bootstrap based empirical likelihood ratio. The approach we follow is based on an artificial likelihood[1] characterisation of the empirical likelihood ratio which allows to fully exploit the higher order asymptotic machinery developed for regular parametric models.

The class of econometric models that can be cast into the empirical likelihood framework, is quite wide: as it will become clear in the next sections, the method of empirical likelihood is in fact theoretically justifiable provided that a set of (arbitrary) estimating equations can be specified. We shall refer to this set of estimating equations as a generalised score function: moment based estimates, least squares, and more generally M and Z type estimators are examples of generalised scores, GS henceforth.

It should be noted that our arguments are valid for exactly identified models (i.e. the dimension of the GS equals the dimension of the unknown parameter); the introduction of nuisance parameters or considering overidentified models do change the higher order asymptotics quite dramatically as recently shown by Lazar and Mykland (1999) and Bravo (1999), respectively.

The remainder of the paper is organised as follows: in the next section, after a brief review of the basic empirical likelihood method, we emphasise its interpretation as an artificial likelihood by using Mykland's (1995) dual likelihood theory and develop the necessary stochastic expansions for analysing the higher order asymptotic properties of the empirical likelihood ratio test. The coverage and power properties are then derived via standard Edgeworth expansion theory, as shown in Section 3. In Section 4 we develop the bootstrap approach to empirical likelihood inference and show the effectiveness of the proposed higher order asymptotics corrections with some Monte Carlo experiments. Section 5 is a conclusion.

Our arguments are based on the assumption that the data come in the form of an i.i.d. random sample. We can relax this assumption by allowing the data to be sampled conditionally on some finite $k \times 1$ vector of weakly exogenous variables, say $w_i = \left(w_{i1} \ldots w_{ik}\right)^T$ (with ^{T} denoting transpose), by considering empirical likelihood ratios for triangular arrays. All the following results are still valid under some additional moments conditions.

Notice that throughout the rest of the paper we use (unless otherwise stated) tensor notation and the summation convention (i.e. for any two repeated indices, their sum is understood). All the indices $r, s, \ldots$ run from 1 to q, and the sum $\sum$ is always intended as $\sum_{i=1}^{n}$ unless otherwise stated.

2 The Relationship between Empirical and Dual Likelihood

Suppose that $\{z_i\}_{i=1}^{n}$ is a sequence of independent $m \times 1$ random vectors from an unknown distribution F_θ depending on an unknown parameter vector $\theta \in \Theta \subseteq R^q$. Let P_θ, P_n be the probability measures associated with F_θ and F_n (where $F_n = 1/n$ is the empirical distribution function), and assume that $P_\theta \ll P_n$.

The information about F_θ is available in the form

$$E_{F_\theta} f_r(z_i; \theta) = 0;$$

for some specified value θ_0 of θ, with the GS $f_r(z_i; \theta) : R^q \times \Theta \to R^q$, $q \times 1$ vector of known measurable functions. Assume that the following conditions hold with probability 1 (w.p.1 henceforth):

GS1 $E_{F_\theta} f_r(z_i; \theta) = 0$ for a unique $\theta^* \in \text{int}\{\Theta\}$,

GS2 i) $E_{F_{\theta_0}} f_r(z_i; \theta_0) f_s(z_i; \theta_0)$ is positive definite and ii) $E_{F_\theta} \partial f_r(z_i; \theta)/\partial\theta^s$ is of full column rank q:

Assumptions GS1 and GS2 are standard; in particular,GS2ii) is made in order to assure (local) identifiability for the underlying parameter of interest θ.

The problem of testing the hypothesis $H_0 : \theta = \theta_0$ can be formulated in terms of finding a probability measure P_θ which is consistent with the constrain $E_{P_{\theta_0}} f_r(z_i;\theta_0) = 0$ and is closest, as measured by the Kullback-Liebler divergence, to the empirical measure P_n. This is essentially what the empirical likelihood technique does. By the restriction $P_\theta \ll P_n$, it turns out that the original constrained estimation of P_θ can be expressed as a simple maximisation of a multinomial likelihood supported on the data over the empirical counterpart of the constraint $E_{F_\theta} f_r(z_i;\theta_0) = 0$.

Let p_i denote the ith element of the unit simplex in R^n and $\hat{p}_i = F_n$ be the nonparametric maximum likelihood estimator for p_i. The empirical likelihood ratio function for testing the hypothesis $H_0 : \theta = \theta_0$ is then given by solving the following program:

$$LR(\theta_0)/2 = \sup_{p_i} \left\{ \sum \log np_i \;|\; p_i \geq 0; \; \sum p_i f_r(z_i;\theta_0) = 0; \; \sum p_i = 1 \right\}: \qquad (1)$$

Let $ch\{S\}$ denote the convex hull for the set $S \subseteq R^q$ and $\|\cdot\|$ be the Euclidean norm; assume now that w.p.1:

E1 $0 \in ch \left\{ f_r(z_1;\theta_0) \; \ldots \; f_r(z_n;\theta_0) \right\}$ for n sufficiently large,

E2 $E_F \|f_r(z_1;\theta)\|^3 < \infty$;

it then follows (from E1) that $LR(\theta_0)$ exists and it is positive[2] (hence the distribution F_{θ_0} attaining the supremum is unique). The required implied probabilities satisfying (1) can be found by a standard Lagrange multiplier argument, and are given by:

$$p_i(\lambda^r(\theta_0)) = 1/(1 + \lambda^r(\theta_0) f_r(z_i;\theta_0))\, n$$

where $\lambda^r(\theta_0)$ is a $q \times 1$ vector of Lagrange multipliers. Moreover, by assumption E2, following Owen's (1990), we can show that $LR(\theta_0)/2 \xrightarrow{d} \chi^2(q) + o(1)$ that it is a nonparametric version of Wilks' theorem. For notational simplicity, let $\lambda^r(\theta_0) = \lambda^r$. An empirical likelihood ratio test for the hypothesis $H_0 : \theta = \theta_0$ is based on

$$W_{\theta_0}(\lambda)/2 = -\sum \log np_i = \sum \log(1 + \lambda^r f_r(z_i;\theta_0)) \qquad (2)$$

which depends only on the Lagrange multipliers λ^r which become therefore the parameters of interest. This last observation is the starting point of the dual likelihood approach to empirical likelihood inference. Testing the hypothesis $H_0 : \theta = \theta_0$ can be thought of as testing the dual hypothesis $H_0^* : \lambda = 0$; such a test can be formulated as in standard parametric inference with a dual likelihood ratio type test:

$$\max_{\lambda} W_{\theta_0}(\lambda)/2 = \left(W_{\theta_0}(0) - W_{\theta_0}\left(\hat{\lambda}\right) \right)/2 = W_{\theta_0}/2.$$

Indeed, following Mykland's (1995) approach, we can consider the empirical log-likelihood function $W_\theta(\lambda)$ as an artificial log-likelihood: it is in fact a dual likelihood in λ. It is easy to see that, subject to integrability conditions, $E_{F_\lambda} \exp W_{\theta_0}(\lambda) = 1$ (more generally for $\lambda \downarrow 0$, $E_{F_\lambda} \exp W_{\theta_0}(\lambda) \to 1$), and Bartlett type identities as developed by Mykland (1994) hold for λ to first order. Let us introduce some notation: let U_{R_v} denote the vth mixed derivative array with respect to the dual parameter λ:

$$U_{R_v} = \frac{\partial^v W_{\theta_0}(\lambda)}{\partial \lambda^{r_1} \partial \lambda^{r_2} \ldots \partial \lambda^{r_v}}$$

with $W_{\theta_0}(\lambda)$ defined as in (2), for any set of indices $1 \leq r_1; r_2; \ldots; r_v \leq q$ in the set R_v. Evaluating the resulting derivatives at the null dual hypothesis $H_0^* : \lambda = 0$, it is straightforward to see that:

$$U_{r_1 r_2 \ldots r_v}|_{\lambda=0} = (-1)^{v-1}(v-1)! \sum f_{r_1}(z_i;\theta) \ldots f_{r_v}(z_i;\theta). \tag{3}$$

One of the most interesting feature of the dual likelihood approach to empirical likelihood theory is given by the existence of Bartlett type identities for the dual parameter λ which relate, as in parametric likelihood theory, linear combinations of expectations of the U_{R_v} arrays defined in (3). Specifically, under the appropriate regularity conditions (see assumption D3 below), for any set R_v of indices, we have that:

$$\sum_R E U_{R_{v_i}} U_{R_{v_j}} \ldots U_{R_{v_k}} = 0 \tag{4}$$

where the sum is over all partitions $R_{v_1} | \dots | R_{v_k}$ of the set R_v. As shown in the next section, this is an important result, because one way of proving the Bartlett correctability of the statistic (2) can be based on the Bartlett type identities in (4).

Let $\kappa_{R_v} = EU_{R_v}$, $\kappa_{R_{v_r};R_{v_s}} = EU_{R_{v_r};}U_{R_{v_s}}$ etc. denote the joint moments and cumulants of the U_R's, all assumed to be $O(n)$. For example, consider the (usual) third Bartlett identity

$$\kappa_{rst} + [3]\kappa_{r;st} + \kappa_{r;s;t} = 0;$$

where $[3]\kappa_{r;st} = \kappa_{r;st} + \kappa_{s;rt} + \kappa_{t;rs}$ and in general the symbol $[k]$ indicates the sum over k similar terms obtained by suitable permutation of indices; using (4) one gets

$$E\left(\sum f_r(z_i;\theta)\sum f_s(z_i;\theta)\sum f_t(z_i;\theta)\right) = [3]E\left(\sum f_r(z_i;\theta)\sum f_s(z_i;\theta)\, f_t(z_i;\theta)\right) - 2E\left(\sum f_r(z_i;\theta)\, f_s(z_i;\theta)\, f_t(z_i;\theta)\right)$$

i.e. a third order Bartlett type identity for the dual likelihood (2).

In order to deal simultaneously with accuracy (i.e. size and coverage probabilities) and power properties of the empirical likelihood ratio test, we consider local analysis by specifying a Pitman alternative $H_n : \theta_n = \theta_0 + \tilde{\theta}/n^{1/2}$ ($\tilde{\theta}$ is a non random $q \times 1$ vector such that: $0 < \tilde{\theta}^T\tilde{\theta} < \infty$). The alternative hypothesis H_n induces a (local) dual alternative of the form $H_n^* : \lambda_n = \tilde{\lambda}/n^{1/2}$, hence in order to examine the local power of an empirical likelihood ratio test we should be considering the augmented hypothesis $H_n^a : \delta_n = \delta_0 + \tilde{\delta}/n^{1/2}$ with the $2p \times 1$ vector $\delta_n = \left[\tilde{\theta}\;\; \tilde{\lambda}\right]^T$. However, given the particular functional form of the dual likelihood (and hence of its derivatives), following Chesher and Smith's (1997, p.636) argument (see also Mykland (1995, p. 411)), it is easy to see that we can focus on the following local alternative $H_n^* : \lambda_n^r = \kappa_{rs}\tilde{\lambda}^s/n^{1/2}$. (It should be noted that we are implicitly assuming that the density of the empirical likelihood test under the alternative is in the same parametric subfamily of the density under the null hypothesis).

We now derive a stochastic expansion for the empirical likelihood ratio test under the sequence of dual local alternatives H_n^*. Note that all the following inequalities should be intended as componentwise.

Let $\Gamma_\tau := \Gamma(0;\tau)$ be an open sphere on A_λ with radius $\tau > 0$, $f(z_1;\theta_0) = \left(U_r \;\; U_{rs} \;\; U_{rst} \;\; U_{rstu} \right)'$ be a vector in R^k with $k = \sum_{a=1} \binom{q+a-1}{a}$ and γ a positive constant; assume that the following regularity conditions hold on some compact set A_λ of the sample space for which assumptions GS1, GS2 and E1 hold w.p.1:

- D1 Interchanging differentiation with respect to λ and integration with respect to z is allowed in $W_{\theta_0}(\lambda)$, and $\sup_{\lambda \in \Gamma_\tau} E\,|f(z_1;\theta_0)|^{4+\gamma} < \infty$
- D2 $\sup_{\lambda \in \Gamma_\tau} \sup_{\|\lambda^* - \lambda\| \le \tau} |\partial^\alpha W_{\theta_0}(\lambda^*)/\partial\lambda^{r_1} \ldots \partial\lambda^{r_v}|^{5+\gamma} < \infty$; $\quad |r_1 + \ldots r_v| = \alpha = 5$;
- D3 $E\left|U_{R_{v_i}} U_{R_{v_j}} \ldots U_{R_{v_k}}\right| < \infty$ for any partition $R_{v_j} | \ldots | R_{v_k}$ of the set R_v (see (4)).

Assumptions D1 and D2 are standard in higher order asymptotics, ensuring that the various error bounds of the asymptotic expansions given in the next section are uniform over compact subsets of A_λ. First notice that under D1 and D2, we have for $\alpha = 1;2;3$ and $0 < \beta < 3/8$

$$P_\lambda\left(|\partial^\alpha W_{\theta_0}(\lambda)/\partial\lambda^{r_1} \ldots \partial\lambda^{r_v} - E(\partial^\alpha W_\theta(\lambda)/\partial\lambda^{r_1} \ldots \partial\lambda^{r_v})|/n^{1/2} > \delta n^{-\beta} \right) = o(1/n); \quad \delta > 0$$

for $|r_1 + \ldots r_v| = \alpha$, uniformly in A_λ. It then follows by Chebyshev's inequality that

$$P_\lambda\left(|\partial^\alpha W_{\theta_0}(\lambda)/\partial\lambda^{r_1} \ldots \partial\lambda^{r_v} - E(\partial^\alpha W_\theta(\lambda)/\partial\lambda^{r_1} \ldots \partial\lambda^{r_v})|/n^{1/2} > \delta' n^{-\beta} \right) = o(1/n); \quad \delta' > 0$$

$$P_\lambda\left(|R_n| > |\lambda^r|^4 \delta'' n^{-\beta} \right) = o(1/n); \quad \delta'' > 0$$

for

$$|R_n| \le |\lambda^r|^4 \sup_{\|\lambda^* - \lambda\| \le \|\lambda\|} \left| \partial^4 W_{\theta_0}(\lambda^*)/\partial\lambda^{r_1} \ldots \partial\lambda^{r_v} \right| / 24n; \quad |r_1 + \ldots + r_v| = 4.$$

Then, following Bhattacharya and Ghosh's (1978) approach, it is possible to show that on the set A_λ, the maximum dual likelihood estimator $\hat\lambda_{MDL}$ satisfies the dual likelihood equations $\partial W_{\theta_0}(\lambda)/\partial\lambda^r = 0$ with P_{A_λ} probability $1 - o(1/n)$ (by von Bahr's inequality). Also notice that the resulting maximiser is unique given the concavity of the objective function in λ:

Finally, assumption D3 implies that the Bartlett type identities hold to first order.

As we are dealing with asymptotic expansions under a local alternative, it is convenient to express the dual likelihood derivatives (3) in terms of nomalised derivatives at θ_n; let

$$Z_r = (U_r - n\kappa_r)/n^{1/2}; \quad Z_{rs} = (U_{rs} - n\kappa_{rs})/n^{1/2}; \quad Z_{rst} = (U_{rst} - n\kappa_{rst})/n^{1/2}; \quad \ldots \tag{5}$$

be a sequence of $O_p(1)$ centered random arrays, under D1 and D2.

Some straightforward algebra shows that $\hat{\lambda}_{MDL}$ admits the following stochastic expansion:

$$\begin{aligned}
\lambda^r &= -\kappa^{rs}Z_s + \tilde{\lambda}^r + \left(\kappa^{rs}\kappa^{tu}Z_{st}Z_u - \kappa_{uvw}\kappa^{ru}\kappa^{sv}\kappa^{tw}Z_sZ_t/2\right)/n^{1/2} + \qquad (6)\\
&\quad \left(-\kappa^{rs}\kappa^{tu}\kappa^{vw}Z_{st}Z_{uv}Z_w + \kappa_{uvw}\kappa^{ru}\kappa^{sv}\kappa^{tw}Z_sZ_{tu}Z_v - \right.\\
&\quad \kappa^{rs}\kappa^{tu}\kappa^{vw}Z_{suw}Z_tZ_v/2 - \kappa^{rs}\kappa^{tw}\kappa^{uz}\kappa^{vz}\kappa_{wzz}Z_{st}Z_uZ_v/2 + \\
&\quad \kappa_{vwzz}\kappa^{rv}\kappa^{sw}\kappa^{tz}\kappa^{uz}Z_sZ_tZ_u/6 - \\
&\quad \left.\kappa_{r's't'}\kappa_{u'v'w'}\kappa^{rr'}\kappa^{ss'}\kappa^{t'u'}\kappa^{vv'}\kappa^{ww'}Z_sZ_vZ_w/2\right)/n + \zeta/n^{3/2};\\
&= \lambda^* + \zeta/n^{3/2}
\end{aligned}$$

where $\lambda^r = n^{1/2}\left(\hat{\lambda}^r_{MDL}\right)$ and $\kappa^{r;s}$ is the matrix inverse of $\kappa_{r;s}$ and the remainder ζ satisfies $P_\lambda\left(|\zeta| > \xi_n n^{1/2}\right) = o(1/n)$ for some sequence $\xi_n \to 0; \xi_n n^{1/2} \to \infty$ as $n \to \infty$. We can then use Čibisov's (1972) general result to show that the Edgeworth expansion for λ (up to the order $o(1/n)$) is equal to that for λ^* on the set B defined in Appendix B.

To derive now an asymptotic expansion for W, we Taylor expand the empirical likelihood ratio about $\hat{\lambda}_{MDL}$, obtaining (after a further Taylor expansion about the normalised deviation $n^{1/2}\eta^r = \tilde{\eta}^r = \left(\lambda^r - \tilde{\lambda}^r\right)$),

$$W/2 = -\kappa_{rs}\lambda^r\lambda^s/2 - \left(Z_{rs} + \kappa_{rst}\tilde{\lambda}^t + \kappa_{rst}\lambda^t/3\right)\lambda^r\lambda^s/\left(2n^{1/2}\right) - \left(Z_{rst}\tilde{\lambda}^t - \right.$$

$$Z_{rst}\lambda^t/3 + \kappa_{rstu}\lambda^t\lambda^u/2 - \kappa_{rstu}\lambda^t\lambda^u/3 + \kappa_{rstu}\lambda^t\lambda^u/12\Big)\lambda^r\lambda^s/(2n) + o_p(1/n):$$

Plugging the stochastic expansion for the maximum dual likelihood estimator (cf. 6) into this last expansion, we obtain the required stochastic expansion for the empirical likelihood ratio under a local alternative:

$$\begin{aligned} W = {} & -\left(Z_r - \kappa_{rr'}\lambda^{r'}\right)\left(Z_s - \kappa_{ss'}\lambda^{s'}\right)\kappa^{rs} + \Big(-\kappa_{rst}\kappa^{ru}\kappa^{sv}\kappa^{tw}Z_uZ_vZ_w/3 + {} \\ & \kappa^{rt}\kappa^{su}Z_{rs}Z_tZ_u - Z_{rs}\lambda^r\lambda^s + \kappa_{rst}\lambda^r\lambda^s\lambda^t/3\Big)/n^{1/2} + {} \\ & \Big(-\left(\kappa^{tu}Z_{rt}Z_u + \kappa_{rtu}\kappa^{tt'}\kappa^{uu'}Z_{t'}Z_{u'}/2\right)\left(\kappa^{vw}Z_{sv}Z_w + \kappa_{svw}\kappa^{vv'}\kappa^{ww'}Z_{v'}Z_{w'}/2\right)\kappa^{rs} + {} \\ & \kappa_{rstu}\kappa^{rr'}\kappa^{ss'}\kappa^{tt'}\kappa^{uu'}Z_{r'}Z_{s'}Z_{t'}Z_{u'}/12 - \kappa^{ru}\kappa^{sv}\kappa^{tw}Z_{rst}Z_uZ_vZ_w/3 + {} \\ & Z_{rst}\lambda^r\lambda^s\lambda^t/3 - \kappa_{rstu}\lambda^r\lambda^s\lambda^t\lambda^u/12\Big)/n + o_p(1/n) \end{aligned} \qquad (7)$$

In order to characterise the higher order asymptotic behaviour of W, we consider its (signed) square root version W_r: working with W_r is in fact extremely convenient for the justification of Bartlett corrections. As $\kappa_{rs} = -\sum f_r(z_i;\theta) f_s(z_i;\theta)$ we can replace $-\kappa_{rs}$ with $\kappa_{r;s}$, whence we can find a $q \times 1$ vector W_r

$$\begin{aligned} W_r = {} & Z_{r'}\left(\kappa^{r';r}\right)^{1/2} + \lambda^{r'}\kappa^{1/2}_{r';r} + \Big(Z_{r's}Z^s/2 + \kappa_{r'st}Z^sZ^t/6 - {} \\ & Z_{r's}\lambda^s/2 - \kappa_{r'st}Z^s\lambda^t/6 + \kappa_{r'st}\lambda^s\lambda^t/6\Big)\left(\kappa^{r';r}\right)^{1/2}/n^{1/2} + {} \\ & \Big(Z_{r'st}Z^sZ^t/6 + 3Z_{r's}Z_{uv}\kappa^{s;u}\kappa^{t;v}Z_t/8 + {} \\ & \left(\kappa_{r'stu}/8 + \kappa_{r'st'}\kappa_{u'tu}\kappa^{t';u'}/3\right)Z^sZ^tZ^u/3 + {} \\ & 5\kappa^{s;t}\kappa_{tuv}Z_{r's}Z^uZ^v/12 - \kappa^{s;u}\kappa^{t;v}\kappa_{uvw}Z_{r'}Z_{st}\lambda^w/4 - {} \\ & \kappa^{s;t}Z_{r'}Z_{stu}\lambda^u/6 - \kappa^{s;t}Z_{r's}Z_{tu}\lambda^u/8 - {} \\ & \left(\kappa_{r'stu}/2 + \kappa^{v;v'}\kappa_{r'sv}\kappa_{v'tu}\right)Z^sZ^t\lambda^u/12 + Z_{r'st}\lambda^s\lambda^t/6 + {} \\ & \left(\kappa^{s;t}\kappa_{stuv} + \kappa^{s;t}\kappa^{t';u'}\kappa_{stt'}\kappa_{u'uv}\right)\lambda^u\lambda^vZ_{r'}/24 + \kappa^{s;t}\kappa_{tuv}Z_{r's}\lambda^u\lambda^v/12 - {} \\ & \left(\kappa_{r'stu} + \kappa^{s';t'}\kappa_{r'ss'}\kappa_{t'tu}/3\right)\lambda^s\lambda^t\lambda^u/24\Big)\left(\kappa^{r';r}\right)^{1/2}/n + o_p(1/n); \end{aligned} \qquad (8)$$

with $Z^r = \kappa^{r;s}Z_s$ and $(\kappa^{r;s})^{1/2}$ is the matrix square root of the inverse symmetric matrix $\kappa^{r;s}$, such that $W = W_rW_s\delta^{rs} + o_p(1/n)$.

It is interesting to note that the signed square root of the empirical likelihood ratio test belongs to the general class of (parametric) tests described in Chandra and Joshi (1983).

In the next section we analyse the higher order properties of W. By finding a valid Edgeworth expansion, we show that an α level coverage error for the confidence region $R_\alpha = \{\theta_0 : W \leq c_\alpha\}$ with the constant $c_\alpha : \Pr(\chi^2(q) < c_\alpha) = \alpha$ is $o(1/n)$; we then show that this rate can be improved, by means of a Bartlett correction, to an order $O(1/n^2)$.

3 Higher Order Asymptotics for the Empirical Likelihood Ratio Test

Our higher order asymptotic analysis begins with evaluating the first four cumulants of W_r, from which the cumulants of W are readily obtained. Since the signed square root is approximated by simple functions of the random arrays Z_{R_v}, we need to evaluate their asymptotic moments in order to obtain an asymptotic expansion of W_r. Similarly to Lawley (1956), it is not difficult to see that under our assumptions, the following holds (up to $o(1/n)$):

$$
\begin{aligned}
E\left(Z_{R_v} Z_{R_w}\right) &= \kappa_{R_v;R_w}; \quad E\left(Z_{R_v} Z_{R_w} Z_{R_x}\right) = \kappa_{R_v;R_w;R_x}/n^{1/2}; \\
E\left(Z_{R_v} Z_{R_w} Z_{R_x} Z_{R_y}\right) &= [3]\,\kappa_{R_v;R_w}\kappa_{R_z;R_y} + \kappa_{R_v;R_w;R_x;R_y}/n; \\
E\left(Z_{R_{v_1}} \ldots Z_{R_{v_k}}\right) &= O\left(1/n^{(k-2)/2}\right) \quad \text{for } k \geq 5
\end{aligned}
$$

After lengthy algebra, using the relations between moments and cumulants, (see for example McCullagh (1987, p. 31)), and the Bartlett type identities as defined in (4) to simplify where possible, we obtain the following approximate cumulants:

$$
\begin{aligned}
k_r &= \lambda^{r'} \kappa_{r';r}^{1/2} + \left(k^2_{r'}/n^{1/2} + k^3_{r'}/n\right)\left(\kappa^{r';r}\right)^{1/2} + o(1/n); \qquad (9) \\
k_{r;s} &= \delta^{rs} + \left(k^2_{r';s'}/n^{1/2} + k^3_{r';s'}/n\right)\left(\kappa^{r';r}\right)^{1/2}\left(\kappa^{s';s}\right)^{1/2} + o(1/n);
\end{aligned}
$$

$$k_{r,s,t} = k^3_{r',s',t'}\left(\kappa^{r',r}\right)^{1/2}\left(\kappa^{s',s}\right)^{1/2}\left(\kappa^{r',r}\right)^{1/2}/n + o(1/n);$$

$$k_{r_1,\ldots,r_v} = o(1/n) \quad v \ge 4;$$

where $k_r = E(W_r)$, $k_{r,s} = COV(W_r, W_s)$, etc. with

$$k^2_{r'} = -\kappa_{r',s,t}\kappa^{s,t}/6 + \kappa_{r'st}\lambda^s\lambda^t/3; \quad k^3_{r'} = \left(-\kappa^{s,u}\kappa^{t,v}\kappa_{r',s,t}\kappa_{u,v,w}\lambda^w/6 - \kappa^{s,t}\kappa_{r,s,t,u}\lambda^u/24 + \left(-\kappa_{r'stu}/4 + \kappa^{s',t'}\kappa_{r'ss'}\kappa_{t'tu}/18\right)\lambda^s\lambda^t\lambda^u\right); \quad k^2_{r,s} = \kappa_{r',s',t}\lambda^t/3;$$

$$k^3_{r',s'} = \Big(\left(-\kappa_{r',s',t,u}/6 + \kappa_{r't,s'u}/2 - 4\kappa_{r',s',tu}/3\right)\kappa^{t,u} - \left(10\kappa_{r'tv}\kappa_{s'uw}/3 - 4\kappa_{r',t,v}\kappa_{s'uw}/3 + \kappa_{r',t,v}\kappa_{s',u,w}/36\right)\kappa^{t,u}\kappa^{v,w} + 5\kappa_{r',s',t}\kappa_{uvw}\kappa^{t,u}\kappa^{v,w}/3 + \left(-(5\kappa_{r',s',t,u}/6 + 2\kappa_{r',s',tu}/3 - \kappa_{r',s'tu}/3)\lambda^t\lambda^u - 2\kappa_{r',t,u}\kappa_{s'vw}\kappa^{t,u}\lambda^v\lambda^w/9 - \kappa_{r',s',t}\kappa_{uvw}\kappa^{t,u}\lambda^v\lambda^w/3\right)\Big);$$

$$k^3_{r',s',t'} = -\kappa_{r',s',t',u}\lambda^u/2 + \kappa_{r',s',t'}\kappa_{uvw}\kappa^{u,v}\lambda^w/12 + \kappa_{r',s',u}\kappa_{t'vw}\kappa^{u,v}\lambda^w/2.$$

Notice that the order of magnitude of the higher order cumulants $k_{r_1,\ldots,r_v}$ for $v \ge 5$ is deduced by applying the general formulae of James and Mayne (1962). Also, the third and fourth cumulants are 0 up to $o\left(1/n^{1/2}\right)$ and $o(1/n)$ respectively, as in standard parametric theory.

Having characterised the order of magnitude of the first four cumulants of the signed square root of empirical likelihood ratio test, we can derive its Edgeworth expansion. The expansion for the distribution of the empirical likelihood ratio test under a local alternative is then obtained from W_r by using the transformation χ : $W_r \to W_r W_s \delta^{r,s}$.

Let $g_{q,\tau}(x)$ and $G_{q,\tau}(x)$ denote the density and the distribution function of a noncentral chi-square random variate with q degrees of freedom and non centrality parameter τ. Also, let $\nabla^k g_{q,\tau}(x)$ be the kth (double) difference operator applied to the density $g_{q,\tau}(x)$ (i.e. $\nabla^k g_{q,\tau}(x) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} g_{q+2(k-j),\tau}(x)$. The following lemma will be used in Theorem 2 below; essentially, it expresses the density of noncentral "generalised" quadratic forms in normal vectors in terms of linear combinations of noncentral chi-square random variates and it is of its own interest:

Lemma 1 Let $\phi_q(\gamma^r;\delta^{rs})$ be the multivariate normal distribution with mean vector $\gamma^r = \left(\gamma^1 \; \ldots \; \gamma^q\right)$ and identity covariance matrix δ^{rs}, and b^{R_v} be a q^v dimensional array of constants not depending on n (i.e. $b^r; b^{rs}; \ldots$), $v = 1; \ldots 4$. Also let h_{R_v} be the vth order Hermite tensor defined by $(-1)^v \partial_{r_1} \ldots \partial_{r_v} \phi_q(\gamma^r;\delta^{rs})$ (where $\partial_{r_v} = \partial/\partial w^{r_v}$), whose structure is reported in the Appendix for completeness. Assume that

L1 the dominance condition $\int \sup_{t \in N} |\partial^{\gamma} \phi_q(\gamma^r;\delta^{rs}) \exp(w^r \delta^{rs} t_s)| \, dx < \infty$ holds w.p.1. on an set N of $t = 0$.

Then the following holds:

$$\begin{aligned}
b^r h^r \phi(\gamma^r;\delta^{rs}) &= b^r \gamma^r \nabla g_{q;\tau}(x); \quad (10)\\
b^{rs} h^{rs} \phi(\gamma^r;\delta^{rs}) &= b^{rr} \nabla g_{q;\tau}(x) + b^{rs} \gamma^r \gamma^s \nabla^2 g_{q;\tau}(x);\\
b^{rst} h^{rst} \phi(\gamma^r;\delta^{rs}) &= [3]\, b^{rss} \gamma^r \nabla^2 g_{q;\tau}(x) + b^{rst} \gamma^r \gamma^s \gamma^t \nabla^3 g_{q;\tau}(x);\\
b^{rstu} h^{rstu} \phi(\gamma^r;\delta^{rs}) &= [3]\, b^{rrss} \nabla^2 g_{q;\tau}(x) + [6]\, b^{rrst} \gamma^s \gamma^t \nabla^3 g_{q;\tau}(x) +\\
&\quad b^{rstu} \gamma^r \gamma^s \gamma^s \gamma^u \nabla^4 g_{q;\tau}(x);
\end{aligned}$$

where $\tau = \gamma^r \gamma^r$ and $[k]$ indicates sum over k terms obtained by permuting the indices.

Proof. See Appendix A. ■

Let

$$\begin{aligned}
B^{r,s} = \; & (-\kappa_{r,s,t,u}/6 + \kappa_{rt,su}/2 - 4\kappa_{r,s,tu}/3 - 20\kappa_{rtv}\kappa_{suw}\kappa^{v,w}/3 + \quad (11)\\
& 8\kappa_{r,t,v}\kappa_{suw}\kappa^{v,w}/3 + 10\kappa_{r,s,t}\kappa_{uvw}\kappa^{v,w}/3)\,\kappa^{t,u}.
\end{aligned}$$

We can now prove the following theorem.

Theorem 2 Let $i^2 = -1$; assume that the vector $f(x_1;\theta_n)$ defined in Section 2 satisfies the following Cramér's condition:

$$\limsup_{\|t\| \to \infty} \left| E \exp\left(i t^T f(x_1;\theta_n) \right) \right| < 1$$

Then, there exist constants a_{jk} (not depending on n), such that the following holds (uniformly over compact subsets of $\tilde{\Lambda}$):

$$\sup_{u\in R_+}\left|P_{\lambda n}(W\le u)-\sum_{j=0}\sum_{k=0}\left(1/n^{j/2}\right)a_{jk}\int_{-1}^{u}g_{q+2k;\tau}(x)\,dx\right|=o(1/n) \qquad (12)$$

where $\tau=\gamma^r\gamma^s\kappa_{r,s}$, and

$$a_{00}=1;\quad a_{10}=-(\kappa_{r,s,t}+2\kappa_{rst})\lambda^r\lambda^s\lambda^t/6;\quad a_{11}=(-\kappa_{r,s,t}+\kappa_{rst})\lambda^r\lambda^s\lambda^t/3;$$

$$\begin{aligned}
a_{12}={}&\kappa_{r,s,t}\lambda^r\lambda^s\lambda^t/6;\quad a_{20}=(\kappa_{r,t,v}\kappa_{s,u,w}+5\kappa_{rtv}\kappa_{suw})\lambda^r\lambda^s\lambda^t\lambda^u\lambda^v\lambda^w/72+\\
&(-\kappa_{r,t,v}\kappa_{s,u,w}/18-5\kappa_{r,s,t}\kappa_{uvw}/72+\kappa_{r,t,v}\kappa_{suw}/12)\kappa^{r,s}\lambda^t\lambda^u\lambda^v\lambda^w+\\
&(-\kappa_{r,s,t,u}/2+\kappa_{rstu}/4+\kappa_{r,s,tu}/3-\kappa_{r,stu}/6)\lambda^r\lambda^s\lambda^t\lambda^u+(5\kappa_{r,s,t}\kappa_{u,v,w}/24+\\
&5\kappa_{r,t,v}\kappa_{suw}/12+5\kappa_{r,s,t}\kappa_{uvw}/24)\kappa^{r,s}\kappa^{t,u}\lambda^v\lambda^w+(11\kappa_{r,s,t,u}/24+\\
&+\kappa_{r,s,tu}/3-\kappa_{r,stu}/6)\kappa^{r,s}\lambda^t\lambda^u+B^{r,s}\lambda^r\lambda^s/2-B^{r,s}\kappa^{r,s}/2
\end{aligned}$$

$$\begin{aligned}
a_{21}={}&-(\kappa_{r,s,v}\kappa_{t,u,w}+2\kappa_{rtu}\kappa_{svw})\lambda^r\lambda^s\lambda^t\lambda^u\lambda^v\lambda^w/18+(7\kappa_{r,s,t}\kappa_{u,v,w}/4+\\
&+\kappa_{r,s,t}\kappa_{uvw}/3+\kappa_{r,t,v}\kappa_{s,u,w}/18+2\kappa_{r,t,v}\kappa_{suv}/9+\kappa_{rtv}\kappa_{suw}/9)\kappa^{r,s}\lambda^t\lambda^u\lambda^v\lambda^w+\\
&(5\kappa_{r,s,t,u}/6+\kappa_{rstu}/4+2\kappa_{r,s,tu}/3-\kappa_{r,stu}/3)\lambda^r\lambda^s\lambda^t\lambda^u+\big(-11\kappa_{r,s,t,u}/24-\\
&-\kappa_{r,s,tu}/3+\kappa_{r,stu}/6-\kappa_{r,s,t}\kappa_{uvw}\kappa^{t,u}/3-2\kappa_{r,t,v}\kappa_{suw}\kappa^{t,u}/3\big)\kappa^{r,s}\lambda^t\lambda^u-\\
&B^{r,s}\lambda^r\lambda^s+B^{r,s}\kappa^{r,s}/2;
\end{aligned}$$

$$\begin{aligned}
a_{22}={}&(\kappa_{r,s,v}\kappa_{t,u,w}/12+\kappa_{rtu}\kappa_{svw}/18)\lambda^r\lambda^s\lambda^t\lambda^u\lambda^v\lambda^w+(-5\kappa_{r,s,v}\kappa_{t,u,w}/24+\\
&+\kappa_{r,t,v}\kappa_{s,u,w}/12)\kappa^{r,s}\lambda^t\lambda^u\lambda^v\lambda^w+(-5\kappa_{r,s,v}\kappa_{tuw}/12-7\kappa_{r,t,u}\kappa_{s,v,w}/18)\kappa^{u,w}\lambda^r\lambda^s\lambda^t\lambda^u-\\
&\Big(-\kappa_{r,s,t,u}\lambda^r\lambda^s/6+\kappa_{r,s,tu}\lambda^r\lambda^s/3-\kappa_{r,stu}\lambda^r\lambda^s/6+\kappa_{r,s,t,u}\kappa^{r,s}/3-\kappa_{rt,su}\kappa^{r,s}/4\\
&+2\kappa_{r,s,tu}\kappa^{r,s}/3\Big)\lambda^t\lambda^u+\Big(-5\kappa_{rtv}\kappa_{suw}\kappa^{t,u}\kappa^{v,w}/3-7\kappa_{r,t,v}\kappa_{s,u,w}\kappa^{t,u}\kappa^{v,w}/72+\\
&7\kappa_{r,t,u}\kappa_{svw}\kappa^{t,v}\kappa^{u,w}/72+7\kappa_{r,s,t}\kappa_{uvw}\kappa^{t,u}\kappa^{v,w}/8\Big)\lambda^r\lambda^s;
\end{aligned}$$

$$a_{23}=\Big(-\kappa_{r,s,v}\kappa_{t,u,w}\lambda^v\lambda^w/18+5\kappa_{r,s,t}\kappa_{uvw}\kappa^{v,w}/72+\kappa_{r,s,v}\kappa_{tuw}\kappa^{v,w}/12-\kappa_{r,s,t,u}/12\Big)\lambda^r\lambda^s\lambda^t\lambda^u;$$

$$a_{24}=\kappa_{r,s,v}\kappa_{t,u,w}\lambda^r\lambda^s\lambda^t\lambda^u\lambda^v\lambda^w/72$$

Proof. See Appendix B. ■

Theorem 3.2 above gives a valid (in Bhattacharya and Ghosh's (1978) sense) second order Edgeworth expansion for the empirical likelihood ratio test under a contiguous alternative. Let $T_{z_\alpha} = I\{W > z_\alpha\}$ denote the z_α level empirical likelihood ratio test (where the constant z_α is such that $\Pr\left(\chi^2_q(\cdot) > z_\alpha\right) = \alpha$, and $I\{\cdot\}$ is the indicator function); it then follows that the local power π_W of T_{z_α} is given by: $\pi_W = \Pr(T_{z_\alpha} = 1 \mid H_n^* : \lambda = \lambda_n)$. In the next corollary we give a second order asymptotic expansion for π_W.

Corollary 3 (Local Power Function for W) Assume that the condition set forth in Theorem 3.2 holds. Then the second order power function for the empirical likelihood ratio test is:

$$\pi_W = 1 - G_{q;\tau}(z_\alpha) + C_1(z_\alpha)/n^{1/2} + C_2(z_\alpha)/n + +o(1/n) \tag{13}$$

with the constants $C_1(z_\alpha)$ and $P_2(z_\alpha)$

$$\begin{aligned} C_1(z_\alpha) &= a_{10}\int_{z_\alpha}^{\infty} g_{q;\tau}(x)\,dx + a_{11}\int_{z_\alpha}^{\infty} g_{q+2;\tau}(x)\,dx + a_{12}\int_{z_\alpha}^{\infty} g_{q+4;\tau}(x)\,dx; \\ C_2(z_\alpha) &= a_{20}\int_{z_\alpha}^{\infty} g_{q;\tau}(x)\,dx + a_{21}\int_{z_\alpha}^{\infty} g_{q+2;\tau}(x)\,dx + a_{22}\int_{z_\alpha}^{\infty} g_{q+4;\tau}(x)\,dx + \\ & \quad a_{23}\int_{z_\alpha}^{\infty} g_{q+6;\tau}(x)\,dx + a_{24}\int_{z_\alpha}^{\infty} g_{q+8;\tau}(x)\,dx \end{aligned}$$

and the various a_{jk} are as in Theorem 2.

Proof. Immediate, since (11) is a direct consequence of expansion (12), and

$$\int_{w^r w^r \geq z_\alpha} \phi_q(w^r; \gamma^r)\,dw = 1 - G_{q;\tau}(z_\alpha):$$

■

It is interesting to note that the power function depends (to second order) also on the constant $B^{r,s\cdot r,s}$ defined in (11). As it will become clear in the next theorem and its corollary, this constant can be used to improve the order of the approximation of the distribution of the empirical likelihood ratio through to $O(1/n^2)$. Hence according to the sign and magnitude of this constant, we should expect that the uncorrected test statistic could perform better in terms of power that the corrected one. However, no general conclusion can be drawn from our analysis.

We now focus on the higher order asymptotic behaviour of the empirical likelihood ratio test under the null (dual) hypothesis $H_0^*: \lambda^r = 0$. First note from (9) that the third and fourth order cumulants of the signed square root W_r of W are $O\left(1/n^{3/2}\right)$ and $O(1/n^2)$, respectively. This order of magnitude of the (higher order) cumulants of W_r is the crucial feature of the empirical likelihood method, as the Bartlett correctability for the density of its square depends essentially on the higher order cumulants; in the next theorem, we give an asymptotic expansion for the empirical likelihood ratio test under the null hypothesis.

Theorem 4 Let $\iota^2 = -1$; assume that the vector $f(x_1;\theta_0)$ defined in D2 satisfies the following Cramér's condition:

$$\limsup_{\|t\|\to\infty}\left|E\exp\left(\iota t^T f(x_1;\theta_0)\right)\right| < 1$$

Then there exists constants a'_{jk} (not depending on n) such that the following holds:

$$\sup_{u\in\mathbb{R}_+}\left|P_0(W\le u) - \sum_{j=0}^{1}\sum_{k=0}^{j}\left(1/n\right)^j a'_{jk}\int_{-\infty}^{u} g_{q+2j}(x)\,dx\right| = o(1/n) \qquad (14)$$

where

$$a'_{00} = 1; \quad a'_{11} = -B^{r,s}\kappa^{r,s}/2; \quad a'_{12} = B^{r,s}\kappa^{r,s}/2.$$

Proof. The proof is similar to Theorem 2; the only difference is that under the null hypothesis, we can exploit the symmetry of the standard normal distribution together with the orthogonality property of the Hermite tensors to infer that the $O\left(1/n^{1/2}\right)$ term vanishes, as well as the integral $\int_{R^q} h^{rs}(w)\,\phi_q(w)\,dw$ (for $r \neq s$) The validity of expansion (13) follows by using Chandra's (1985) Theorem which holds for all Borel subset C of $\mathcal{C}$ satisfying:

$$\sup_{C\in\mathcal{C}} \int_{(\partial C)^{\varepsilon}} g_{q;\tau}(x)\,dx = O(\varepsilon)\,; \quad \varepsilon \downarrow 0:$$

■

Theorem 4 gives a valid second order Edgeworth expansion for the empirical likelihood ratio test under the null hypothesis. It is interesting to note that the signed square root of the empirical likelihood is $N(0;\delta^{rs}) + O\left(1/n^{3/2}\right)$ (in terms of a formal Edgeworth expansion) as in standard parametric theory. Moreover, by examining the structure of the expansion, it is easy now to see that adjusting the test statistic W by a scale constant of the form $(1 + B'/n)$ where $B' = B^{r;s}\kappa^{r;s}/q$ makes the second order term vanish.

Corollary 5 (Bartlett Correction for W) Under the conditions set forth in Theorem 3.4, then the following holds:

$$\sup_{u\in\mathbb{R}_+}\left|P_0\left(W/(1 + B'/n) \le u\right) - \int_{-\infty}^{u} g_q(x)\,dx\right| = o(1/n)\,: \qquad (15)$$

with B' (cf. (11)) the Bartlett correction factor for the empirical likelihood ratio test.

Proof. Immediate given the expansion (14) and hence omitted. ■

Remark 1 It should be noted that the approximation error $o(1/n)$ is obtained by considering a valid Edgeworth expansion for W. In terms of a formal Edgeworth expansion, the error can be replaced by $O(1/n^2)$ given the odd-even property of the third order Hermite tensors (Barndorff-Nielsen and Hall, 1988). (Of course by an appropriate strengthening of moments, we can still obtain a valid Edgeworth expansion to the order $o\left(1/n^{3/2}\right)$).

Remark 2 It should also be noted that the Bartlett adjustment as given in (11) differs from the "standard" adjustment obtained by DiCiccio et al. (1991), being expressed in terms of expectations of product of derivatives of a dual likelihood (see e.g. Lawley (1956)) as opposed to simple moments. This fact should not come as a surprise though, given the "likelihood" approach adopted in this paper.

In the next section, we analyse some applications of empirical likelihood method to some econometric models.

4 Some Econometric Applications

So far we have seen that for a given data set, if the hypothesised model admits a GS, then the use of empirical likelihood methods is theoretically justifiable. We first discuss how we can estimate the Bartlett correction.

As noted in Remark 2 of the previous section, the Bartlett adjustment (10) is characterised by the presence of expectations of products of derivatives of a dual likelihood. This latter fact implies that the estimation of B' itself is more complicated as we need to estimate these product of derivatives. Under the additional assumption:

- D4 The vector $f(x_1;\theta_0)$ defined in D2 satisfies the following moment condition $E\|f(x_i;\theta_0)\|^8 < \infty$,

we can consistently estimate the Bartlett factor B' (i.e. the consistency follows by a straightforward application of Chebyshev's inequality) by introducing the array ξ^{I_ν} for any set of indices $i, j, \ldots$ in I_ν and $1 \le i, j, \ldots \le n$ such that its coefficients satisfy the criterion of unbiasedness and are given by the general formula $(-1)^{\nu-1}(\nu-1)!/(n-1)^{(\nu-1)}$ with $\nu \le n$ and $(n-1)^{(\nu-1)} = (n-1)(n-2)\ldots(n-\nu+1)$ (see McCullagh (1987, Chapter 4) for more details). Let f_r^i denote the (i, r)th component of the matrix f_r^i; dropping temporarily the summation convention for the indices

$i, j, \ldots,$ the sample analog of the various components of the Bartlett adjustment are

$$\tilde\kappa_{r,s} = \sum_{ij} \xi^{ij} f^i_r f^j_s/n; \quad \tilde\kappa_{rst} = \sum_{i} f^i_r f^i_s f^i_t/n; \quad \tilde\kappa_{r,s,t} = \sum_{ijk} \xi^{ijk} f^i_r f^j_s f^k_t/n$$

$$\tilde\kappa_{rt,su} = \sum_{ij} \xi^{ij} f^i_r f^i_t f^j_s f^j_u/n; \quad \tilde\kappa_{r,s,tu} = \sum_{ijk} \xi^{ijk} f^i_r f^j_s f^k_t f^k_u/n;$$

$$\tilde\kappa_{r,s,t,u} = \sum_{ijkl} \xi^{ijkl} f^i_r f^j_s f^k_t f^l_u/n$$

and are evaluated at any root n consistent estimator of θ. The sample (i.e. the feasible) version of the Bartlett correction is:

$$\tilde B' = \Big(-\tilde\kappa_{r,s,t,u}/6 + \tilde\kappa_{rt,su}/2 - 4\tilde\kappa_{r,s,tu}/3 - 20\tilde\kappa_{r,t,v}\tilde\kappa_{suw}\tilde\kappa^{v,w}/3 + 8\tilde\kappa_{r,t,v}\tilde\kappa_{suw}\tilde K^{v,w}/3 + 10\tilde\kappa_{r,s,t}\tilde\kappa_{uvw}\tilde K^{v,w}/3 \Big) \tilde K^{r,s}\tilde K^{t,u}/2 \tag{16}$$

It should be noted that replacing the theoretical correction with its sample version does not affect the (formal) $O(1/n^2)$ order of approximation. A simple Taylor expansion about θ_0 shows in fact that:

$$\tilde B' = B' + C^r(\theta)\, U^r/n^{1/2} + O_p(1/n)$$

(where $C^r(\theta) = (\partial B'/\partial\theta^r)\,|_{\theta=\theta_0}$) which implies that the difference between $W/(1 + B'/nq)$ and $W/\left(1 + \tilde B'/nq\right)$ is given by the following integral:

$$\int \phi_q(w^r; \delta^{rs})\, w^r w^r C^s(\theta)\, w^s dw/n^{3/2}$$

which is again 0 by symmetry (Barndorff-Nielsen and Hall, 1988).

As the computation of the sample adjustment is rather complicated, an alternative approach to achieve higher order asymptotic refinements to the limiting distribution of the empirical likelihood ratio test seems preferable. We propose to use the bootstrap method. The bootstrap calibration can be implemented in two different ways: we can either bootstrap the distribution of W or the Bartlett correction B' itself. Both methods relies essentially on the following theorem, which shows that bootstrapping the distribution of W under the null hypothesis leads to the same level of accuracy of the Bartlett corrected W. Let $\{x_i^*\}_{i=1}^n$ denote a bootstrap sample obtained by the original sample $\{x_i\}_{i=1}^n$. Let also $f_r^*(x_i^*;\theta)$ denote a centered bootstrap GS; recentering here is essential as it makes the boostrapped GS unbiased conditionally on the original sample. Let $f^*(z_1;\theta_0) = \left(U_r^* \; U_{rs}^* \; U_{rst}^* \; U_{rstu}^*\right)'$ be the bootstrap vector analogous to the one described in Section 2. Assume that with bootstrap probability P^*_λ 1 (w.b.p.1) the following holds

BE1 $0 \in ch\left\{f_r^*(x_1^*;\theta_0) \; \ldots \; f_r^*(x_n^*;\theta_0)\right\}$ for n sufficiently large

which justifies the existence and positiveness of a bootstrapped empirical likelihood ratio for the parameter θ.

Assume also that w.b.p.1.

BD1 $\sup_{\lambda\in\Lambda} E|f^*(z_1;\theta_0)|^{4+\delta} < \infty$

BD2 $\sup_{\lambda\in\Lambda}\sup_{\|\lambda'\|\le\tau} |\partial^\alpha W_{\theta_0}(\lambda')/\partial\lambda^{r_1}\ldots\partial\lambda^{r_v}|^{5+\delta} < \infty; \quad |r_1+\ldots r_v| = \alpha = 5;$

BD3 $E\left|U^*_{R_{v_i}} U^*_{R_{v_j}} \ldots U^*_{R_{v_k}}\right| < \infty$ for any partition $R_{v_j} | \ldots | R_{v_k}$ of the set R_v (see (4)).

We can then prove the following theorem:

Theorem 6 (Bootstrap empirical likelihood test) Under conditions BE1, BD1, BD2 and BD3, assuming that.

$$\limsup_{\|t\|\to\infty}\left|E\exp\left(\iota t^T f^*(x_1;\theta_0)\right)\right| < 1$$

holds, then conditional on the original sample χ, there exist constants a^*_{jk} $(k = 1;2)$ (not depending on n) such that the following holds:

$$\sup_{u\in R_+}\left|P_0^*(W^*\le u) - \sum_{j=0}\sum_{k=0}(1/n)^j a^*_{jk}\int_{-\infty}^u g_{q+2k}(x)\,dx\right| = o(1/n): \qquad (17)$$

Proof. See Appendix C. ■

Let z^*_α be now a constant such that $P^*_\le(W^* \le z^*_\alpha) = \alpha$ (i.e. an α level for the bootstrap test W^*). Recalling Remark 1, we can deduce that the formal approximation error is actually $O(1/n^2)$. We can then state the following corollary:

Corollary 7 (Higher order accuracy for W^*) The level of the empirical likelihood ratio test with bootstrap corrected critical value z^*_α is given by:

$$P_0(W \le z^*_\alpha) = \alpha + O\left(1/n^2\right). \qquad (18)$$

Proof. By direct comparison of expansion (13) with its bootstrap analog, as the difference $\kappa_{jk} - \kappa^*_{jk} = o_p(1)$, it follows that

$$\sup_{z \in R_+} |P_0(W \le z) - P^*_0(W^* \le z)| = O\left(1/n^2\right);$$

and hence the results follows immediately replacing z with z^*_α. ■

As originally suggested by Hall and LaScala (1991), the bootstrap distribution W^* of W can be used to estimate directly a bootstrap based Bartlett correction, say B_b. Specifically, let n_b denote the number of bootstrap replications. Then a bootstrap based Bartlett correction can be found by solving the equation:

$$\sum_{B=1}^{n_b} W^*_B = q(1 + B_b/n) \qquad (19)$$

for B_b. It should be noted that $B_b \xrightarrow{p} B'$ for $n_b \to \infty$, as it can easily deduced by the expansion (17).

We now turn to some examples that will illustrate the applications of the empirical likelihood method to some econometric problems.

EXAMPLE 1. Moment condition models

We consider the case where the assumed (unconditional) moment restrictions are of intrinsic interest (for example they might have been derived by economic theory) and there is no parametric specification of the data generating process; in this set-up the GS is given by the vector of moment conditions itself. Given the unconditional nature of the problem, the bootstrapped centered moment condition is

$$f_r^*(x_i^*;\theta) = f(x_i^*;\theta) - E^* f\left(x_i^*;\hat{\theta}\right)$$

where $\hat{\theta}$ is a simple moment estimator. The following model, which can be related to real business cycle models and is adapted from Burnside and Eichenbaum (1994), is analysed:

$$E f(x_i;\theta_0)^T = E\left[x_{1i}^2 - \theta_{10} \quad x_{2i}^2 - \theta_{20} \quad \cdots \quad x_{qi}^2 - \theta_{q0}\right] \qquad (20)$$

The $q = 5$ elements of the vector x are either standard normal or are $t(5)$ distributed, and the null hypothesis to be tested are $H_0 : \theta_0^T = \left[1 \ \cdots \ 1\right]$ and $\theta_0^T = \left[\mu_2 \ \cdots \ \mu_2\right]$ ($\mu_2 = 5\left(\Gamma(1/2)\,\Gamma(5/2)\right)^{-1}\left(\Gamma(3/2)\right)^2 = 5/3$), respectively. Tables 1 and 2 report the empirical sizes of the original empirical likelihood ratio, the feasible Bartlett corrected analog (16) and its bootstrap based counterpart (18) for 0.10, 0.05 and 0.01 nominal sizes.

Tables 1 and 2 here

Notice that both corrected tests improve upon the standard first order asymptotics based empirical likelihood ratio test. The bootstrap based correction seems to perform slightly better that the empirical one, but this fact is hardly surprising given the notorious difficulty to estimating empirical cumulants. Also notice that as the sample size grows the relevance of the correction diminishes.

EXAMPLE 2. Regression models

We consider semiparametric (i.e. with unknown distribution of the innovations) possibly non linear regression models. In regression models, the GS is obtained by considering (see e.g. Newey (1990)) a regression residual function

$$\gamma\left(y_i, x_i^r, \beta_r\right) = y_i - g\left(x_i^r, \beta_r\right)$$

for some known measurable function $g(\cdot)$ such that under the true distribution of the data

$$E\left(\gamma\left(y_i, x_i^r, \beta_r\right) \mid x\right) = 0$$

which implies an unconditional moment restriction of the form

$$EA\left(x_i^r\right)\gamma\left(y_i, x_i^r, \beta_r\right) = 0$$

for some $q \times n$ matrix of instruments $A\left(x_i^r\right)$. The matrix of optimal instruments is $A\left(x_i^r\right) = \Gamma_{ri}\nu^{ij}$ (where $\Gamma_{ri} = \partial\gamma\left(y_i, x_i^s, \beta_s\right)/\partial\beta_r = -\partial g\left(x_i^s, \beta_s\right)/\partial\beta_r$ and $\nu_{ij} = E\left(\gamma\left(y_i, x_i^r, \beta_r\right)\gamma\left(y_j, x_j^r, \beta_r\right) \mid x\right)$); assuming further that the conditional variance takes the following functional specification $\nu_{ii} = h\left(x_i^r\right)$, for some measurable function $h(\cdot) : R^{q'} \to R^{+}$ $(q' \leq q)$ we obtain the optimally weighted GS

$$E\,\Gamma_{ri}\nu^{ii}\gamma\left(y_i, x_i^r, \beta_r\right) = 0.$$

The corresponding bootstrapped GS is then based on

$$E^{*}\partial g\left(x_i^{s*}, \beta_s\right)/\partial\beta_r\,\nu^{ii}\left(y_i^{*} - g\left(x_i^{r*}, \beta_r\right) - \varepsilon_n^{*}\right) = 0,$$

where $y_i^{*} = x_i^{r*}\hat{\beta}_r + \varepsilon_i^{*}$ is the ith bootstrap pseudo-observation, $\hat{\beta}_r$ is a heteroskedasticity corrected non linear least square estimator the unknown $q \times 1$ vector β^r, ε_i^{*} is the bootstrap sample drawn from $\hat{\varepsilon}_i = y_i - x_i^r\hat{\beta}_r$, x_i^{r*} are sampled (independently from ε_i^{*}) from the empirical distribution of x^r, and $\varepsilon_n^{*} = E^{*}\varepsilon_i^{*}$. In the Monte Carlo study, we consider the following specification for $g(\cdot)$: $y_i = \exp\left(\beta_0 + \beta_1 x_i^1\right) + \varepsilon_i$ with innovations $\varepsilon_i \sim N(0,1)$ or $\sim t(4)$ and heteroskedasticity function $\nu_{ii} = x_{1i}^2$. Tables 3 and 4 report the empirical sizes of the original empirical likelihood ratio, the feasible Bartlett corrected analog (16) and its bootstrap based counterpart (18) for 0.10, 0.05 and 0.01 nominal sizes..

Tables 3 and 4 here

EXAMPLE 3. Robust regression models

We consider robust regression models with fixed regressors; the GS in this case is given by

$$E x_i' \psi \left(y_i - x_i' \beta_r \right) = 0$$

for the psi-function $\psi : \mathbb{R} \to \mathbb{R}$ satisfying $E\psi(\varepsilon_i) = 0$. The bootstrapped GS is

$$E^* x_i' \left(\psi \left(y_i^* - x_i' \beta_r \right) - \varepsilon_n^* \right) = 0$$

where $y_i^* = x_i' \hat{\beta}_r + \varepsilon_i^*$ is the bootstrap pseudo-observation, $\hat{\beta}_r$ is an M-estimator for the unknown $q \times 1$ vector β', ε_i^* is the bootstrap sample drawn from $\hat{\varepsilon}_i = y_i - x_i' \hat{\beta}_r$ and $\varepsilon_n^* = E^* \varepsilon_i^*$. Following Huber (1973), we specify the psi-function $\psi(\cdot)$ as

$$\left(y_i - x_i' \beta_r \right) I \left\{ \left| y_i - x_i' \beta_r \right| \le k \right\} + k \cdot \operatorname{sgn} \left(y_i - x_i' \beta_r \right) I \left\{ \left| y_i - x_i' \beta_r \right| > k \right\}$$

with the constant $k = 1.4$, the scale parameter $\sigma^2 = 1$, $\operatorname{sgn}(\cdot)$ and $I\{\cdot\}$ are the sign and indicator function, respectively. Table 5 and 6 report some Monte Carlo results for a simple 2 covariates design with an intercept and a single fixed regressor x_i generated as equally spaced grid of numbers between -1 and 1 and points at -3 and 3, so that we have a rather substantial leverage effect. The innovation process is specified to be $N(0;1)$ and $t(4)$. The null hypothesis is $\beta_0 = \begin{bmatrix} 1 & 1 \end{bmatrix}'$.

Table 5 and 6 here

EXAMPLE 4. Quasi and Pseudo-likelihood models

We consider quasi and pseudo-likelihood models together because the analysis in rather similar from the point of view of empirical likelihood based inference. In the case of "classical" quasi-likelihood approach, the GS is given by the $q \times 1$ quasi-score

$$E\,\partial\eta(x_i^r;\beta_r)/\partial\beta_s \nu^{ij}(\beta_r)\left(y_j - \eta\left(x_j^r;\beta_r\right)\right) = 0;$$

with $Ey_i = \eta(x_i^r;\beta_r)$ for some known link function $\eta(\cdot)$ and $V(y_i) = \nu_{ij}(\beta_r)$(we assume known dispersion parameter $\phi = 1$). For this class of models, the bootstrapped GS is

$$E^*\partial\eta(x_i^r;\beta_r)/\partial\beta_s \nu^{ij}(\beta_r)\,|_{\beta_r=\hat{\beta}_r}\,\varepsilon_j^* = 0$$

where ε_i^* is a bootstrap sample drawn from the centered residuals $\tilde{e}_i = \hat{\varepsilon}_i - \sum_i \hat{\varepsilon}_i/n$ with $\hat{\varepsilon}_i = \nu_{ii}^{-1/2}\left(\hat{\beta}_r\right)\left(y_j - \eta\left(x_j^r;\hat{\beta}_r\right)\right)$ and $\hat{\beta}_r$ is the quasi-maximum likelihood estimator for β_r.

For pseudo-maximum likelihood based models, we need to take into account the possible misspecification of the model (i.e. the second standard Bartlett identity does not hold as in quasi-likelihood models), hence a pseudo-score is

$$E\left(\partial\eta(x_i^r;\beta_r)/\partial\beta_s \nu^{ij}(\beta_r)\,\partial\eta\left(x_j^s;\beta_s\right)/\partial\beta_t\right)\partial\eta(x_k^r;\beta_r)/\partial\beta_t v^{kl}(\beta_r)(y_l - \eta(x_l^r;\beta_r)) = 0$$

for a given specification of the matrix $v_{ij}(\beta_r) = COV\left((y_i - \eta(x_i^r;\beta_r))\left(y_j - \eta\left(x_j^r;\beta_r\right)\right)\right)$. The bootstrap analog is (as in the quasi-likelihood case)

$$E^*\left(\partial\eta(x_i^r;\beta_r)/\partial\beta_s \nu^{ij}(\beta_r)\,\partial\eta\left(x_j^s;\beta_s\right)/\partial\beta_t\right)\partial\eta(x_k^r;\beta_r)/\partial\beta_t v^{kl}\varepsilon_l^* = 0;$$

the only (important) difference being that the estimated residuals are obtained by using the matrix $v_{ii}^{-1/2}$. Notice that we have been implicitly assuming a fixed regressors set-up; as in the case of nonlinear regression, though, we can assume stochastic regressors which implies resampling also from the empirical distribution of the x's in the bootstrap algorithm. More importantly, it is worth noting that in this case the bootstrap calibration is based not on a GS evaluated at the null, but on the (estimated) residuals. This is due to the fact that both quasi and pseudo-likelihood models have not the structure of an expected term plus noise typical of regression type models. However using the same argument used to justify the application of an estimated Bartlett correction, we are still able to achieve higher order accuracy.

We analyse the Poisson model with a specification error discussed in Gourieroux, Monfort and Trognon (1984). Suppose that $y_i \sim Pois(\varphi_i)$ with parameter $\varphi_i = \exp(x_i^{r\prime}\beta_r + \varepsilon_i)$, where x^r is a $q \times 1$ vector of exogenous variables and ε_i is a specification error. We assume that: $E(\exp \varepsilon_i) = 1$ and $VAR(\exp \varepsilon_i) = 1$, and use $N(\varphi_i; 1)$ and $Pois(\varphi_i)$ as kernels for the pseudo-likelihood; the resulting pseudo-scores can be found in Gourieroux et al. (1984). In the Monte Carlo simulation we take $\varepsilon_i \sim N(-0.35; 0.7056)$ (so that $E(y_i) = 1$); Tables 7 and 8 report the results, for the null hypothesis $H_0 : \beta_r = \left[\beta_0 \ \ \beta_1\right] = 0$.

Tables 7 and 8 here

Remark 3 (Using second moment information) So far, we have assumed that the information available is given (essentially) in the form of a moment restriction for the mean of the model. The empirical likelihood framework can easily incorporate additional information, most noticeably information about the second moment. For example, in regression analysis we can augment the residual regression function to allow the conditional variance to depend on an additional $p \times 1$ parameter vector η_a (which may include β_r as well), so that residual specification. The resulting GS is $E(\gamma(y_i; x_i^r; \theta_{r'}) \mid x) = 0$ with the 2×1 vector

$$\gamma(y_i; x_i^r; \theta_{r'}) = \left[y_i - g(x_i^r; \beta_r) \quad (y_i - g(x_i^r; \beta_r))^2 - h(x_i^r; \eta_a) \right]$$

depending on the $r' \times 1$ ($r' = q + s$) vector of parameters, for some measurable function $h(\cdot) : R^s \to R^+$. A straightforward calculation shows that the optimal instrument matrix is given in this case by $\Gamma^{r'}_i = \begin{bmatrix} \partial \gamma(y_i; x_i^r; \theta_{r'})/\partial \beta_r & 0 \\ \partial h(x_i^r; \eta_a)/\partial \beta_r & \partial h(x_i^r; \eta_a)/\partial \eta_a \end{bmatrix}$ and joint covariance matrix of ε_i and ε_i^2. In the quasi-likelihood case we can introduce unknown overdispersion via an extended quasi-likelihood , while for the robust regression model we could introduce an estimating equation for the scale parameter σ^2.

5 Conclusions

We have shown how the empirical likelihood method can be applied to inferential problems based on moment restrictions, emphasising the interpretation of the empirical likelihood ratio test statistics as a dual likelihood. Provided that the econometric model is identified, it is easy to test a simple hypothesis about the parameters of interest by means of the dual empirical likelihood ratio test: one needs just to specify a constraint (which in the case of moment conditions based models is given by the empirical counterpart of the moment condition itself) and maximise the dual empirical log-likelihood ratio with respect to the dual parameter. The accuracy of the resulting test can be improved to third order by applying a Bartlett correction factor to the test statistic itself; this latter feature is, possibly, the most interesting property of empirical likelihood based inference, as no other nonparametric technique is known to be Bartlett correctable. The dual likelihood approach gives a simple explanation of this peculiar phenomenon. We have also investigated analytically the power properties of the dual empirical log-likelihood: from our analysis, it is clear that any loss in power is typically a second order effect and hence its impact can be considered negligible when the sample size is reasonably big, however no general conclusion can be drawn.

Empirical likelihood can also be adapted to dependent processes (Kitamura, 1997). In particular for smooth functions of α mixing processes, Kitamura (1997) proves that it is still possible to obtain higher order accuracy (specifically up to $O\left(n^{-5/6}\right)$) for the empirical likelihood ratio test statistic by using blockwise resampling techniques analog to those used in the bootstrap literature. This should be of particular relevance for time series based models.

6 Notes

1 We use the term artificial to stress the fact that we are dealing with a mathematical object which shares some properties of a parametric likelihood but it cannot be defined as a formal Radon-Nikodym derivative with respect to some dominating measure.

2 To show this essential point, we restrict the collections of sets $\mathcal{C}$ on the Borel field $(\mathbb{R}^q; \mathcal{B}^q)$ supporting the unknown measure P_θ to some pointwise separable, (to ensure measurability) Vapnik-Červonenkis classes of sets (see e.g. Gaenssler (1983)). Let e and E be a unit and the set of unit vectors in $\mathbb{R}^q$ respectively. By the classical Glivenko-Cantelli theorem generalised to uniform convergence to half spaces (Ranga Rao, 1962) we get

$$\sup_{e \in E} \left| (P - P_n)\, e^T f(z;\theta_0) \right| \to 0 \quad a.s.;$$

this implies (Owen, 1990) that for any $\varepsilon > 0$,

$$\Pr \left\{ \inf_{e \in E} P_n \left(e^T f(z;\theta_0) > 0 \right) > \varepsilon/2 \right\} \quad \text{all but finitely often w.p.1} \qquad (21)$$

and as we are considering VC classes of sets, we can conclude that the latter probability converges to 0 at an exponential rate by the Vapnik-Červonenkis inequality (see for example Gaenssler (1983, Lemma 10)). This fact in turns implies that $0 \in ch\{f(z_1;\theta_0); \ldots; f(z_2;\theta_0)\}$ as an interior point (as in ED1 above), whence the empirical likelihood ratio exists and its positive.

References

Back, K. and Brown, D.: 1993, Implied probabilities in GMM estimation, Econometrica 61, 971{975.

Barndor®-Nielsen, O. and Hall, P.: 1988, On the level-error after Bartlett adjustment of the likelihood ratio statistics, Biometrika 75, 374{378.

Bhattacharya, R.: 1987, Some aspects of Edgeworth expansions in statistics and probability, in J. V. M. L. Puri and W. Wertz (eds), New Perspectives in Theoretical and Applied Statistics, New York: John Wiley & Sons, pp. 157{170.

Bhattacharya, R. and Ghosh, J.: 1978, On the validity of formal Edgeworth expansion, Annals of Statistics 6, 434{451.

Bravo, F.: 1999, Higher order asymptotics for the empirical likelihood ratio J test, Mimeo, University of Southampton.

Burnside, C. and Eichenbaum, M.: 1994, Small sample properties of generalized method of moments based wald tests, Technical Working Paper 155, National Bureau of Economic Research.

Chandra, T.: 1985, Asymptotic expansion of perturbed chi-square variables, Sankhyā A 47, 100{110.

Chandra, T. and Ghosh, J.: 1980, Valid asymptotic expansions for the likelihood ratio and other statistics under contiguous alternatives, Sankhyā A 42, 170{184.

Chandra, T. S. and Joshi, S.: 1983, Comparison of the likelihood ratio, Rao's and Wald's tests and a conjecture of C.R.Rao, Sankhyā A 45, 226{246.

Chesher, A. and Smith, R.: 1997, Likelihood ratio speci¯cation tests, Econometrica 65, 627{646.

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

Efron, B.: 1981, Nonparametric standard errors and con¯dence intervals (with discussion), Canadian Journal of Statistics, 9, 139{172.

Gaenssler, P.: 1983, Empirical Processes, IMS , Hayward CA.

Gourieroux, C., Monfort, A. and Trognon, A.: 1984, Pseudo maximum likelihood methods: Applications to poisson models, Econometrica 52, 701{720.

Hall, P. and LaScala, B.: 1991, Methodology and algorithms of empirical likelihood, International Statistical Review 58, 109{127.

Huber, P.: 1973, Robust regression: Asymptotics, conjectures and Monte Carlo, Annals of Statistics 1, 799{821.

James, G. and Mayne, A.: 1962, Cumulants of functions of random variables, Sankhyā A 24, 47{54.

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

Lawley, D.: 1956, A general method for approximating to the distribution of the likelihood-ratio criteria, Biometrika 43, 295{303.

Lazar, N. and Mykland, P.: 1999, Empirical likelihood in the presence of nuisance parameters, Biometrika 86, 203{211.

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

Muirhead, R.: 1982, Aspects of Multivariate Statistical Theory, New York: Wiley.

Mykland, P.: 1994, Bartlett type of identities, Annals of Statistics 22, 21{38.

Mykland, P.: 1995, Dual likelihood, Annals of Statistics, 23, 396{421.

Newey, W.: 1990, E±cient instrumental variables estimation of nonlinear models, Econometrica 58, 809{837.

Owen, A.: 1988, Empirical likelihood ratio con¯dence intervals for a single functional, Biometrika 36, 237{249.

Owen, A.: 1990, Empirical likelihood ratio con¯dence regions, Annals of Statistics 18, 90{120.

Ranga Rao, R.: 1962, Relations between weak and uniform convergence of measures with applications, Annals of Mathematical Statistics 33, 659{680.

Stein, C.: 1956, E±cient nonparametric testing and estimation, in J. Neyman (ed.), Proceedings of Third Berkley Symposium in Mathematical Statistics and Probability, Vol. 1, University of California Press, Berkeley, pp. 187{195.

Čibisov, D.: 1972, An asymptotic expansion for the distribution of a statistic admitting a stochastic expansion, Theory of Probability and its Applications 17, 620{630.

APPENDIX

A Proof of Lemma 1

The first four Hermite tensors are:

$$h^r = w^r - \gamma^r; \quad h^{rs} = h^r h^s - \delta^{rs}; \quad h^{rst} = h^r h^s h^t - [3]\, h^r \delta^{st}; \qquad (22)$$
$$h^{rstu} = h^r h^s h^t h^u - [6]\, h^r h^s \delta^{tu} + [3]\, \delta^{rs} \delta^{tu}.$$

Let $\tau = \gamma^r \gamma^r$, $\phi_q(\gamma^r, \delta^{rs})$ be the q variate normal density with mean γ^r and identity covariance matrix δ^{rs}, and t_r be a vector of auxiliary real variables. Also, let $w^{R_\nu} = w^{r_1} w^{r_2} \ldots w^{r_\nu}$ and $b^{R_\nu} = b^{r_1 r_2 \ldots r_\nu}$. To prove the lemma, we use the transformation $T : w^r \to (x, v^r)$ (with $x = w^r w^r$, $v^r = w^r/(w^s w^s)^{1/2}$ and Jacobian $J = x^{q/2-1}/2$) and the following identity:

$$w^{R_\nu} b^{R_\nu} \phi_q(\gamma^r, \delta^{rs}) \equiv \sum_{R_\nu} b^{R_\nu} \partial^{\gamma} \left(\phi_q(\gamma^r, \delta^{rs}) \exp(w^r \delta^{rs} t_s)\right) |_{t_r=0}$$

where $\sum_{R_\nu}$ indicates summation over the partition $\Upsilon = \{\nu_1, \ldots, \nu_p\}$ of ν indices into p non-empty blocks such that the resulting homogeneous polynomial in γ^{R_ν} is even or odd according to the number of indices in the set R_ν.(i.e.e the components of the b array). Using T, the density for x is obtained by integrating out the vector v^r over the unit sphere $v^r v^r = 1$ in R^q, that is:

$$(2\pi)^{-q/2} \sum_{R_\nu} \int_{v^r v^r = 1} b^{R_\nu} \exp\{-(x+\tau)/2\}\, J\, \partial^{\gamma} \exp\left\{ x^{1/2} v^r \delta^{rs} t_s \right\}_{t_r=0} (v^r dv^r). \qquad (23)$$

Interchanging differentiation and integration which is permissible by assumptions D1, D2 and L1 (note that the transformation T is essentially a polar coordinate type transformation), we can then use Theorem 7.4.1 in Muirhead (1982), to get:

$$C(x, \tau) \sum_{R_\nu} b^{R_\nu} \partial^{\gamma}\, {}_0F_1(; q/2; x(\tau + t^r t^r + 2\gamma^r t^r)/4) |_{t_r=0}$$

with $C(x;\tau) = x^{q/2-1}\exp\{-(x+\tau)/2\}/2^{q/2}\Gamma(q/2)$, ${}_0F_1(;a;z) = \sum_{k=0}^{\infty} z^k/(a)_k k!$ and $(a)_k = \Gamma(a+k)/\Gamma(a)$. Differentiating now ${}_0F_1(;\cdot;\cdot)$ (up to $\nu = 4$), evaluating the resulting derivatives at $t_r = 0$, and taking into account the symmetric structure of the b arrays, we obtain:

$$\sum_{r=1} b^r \partial\, {}_0F_1(;\cdot;\cdot)/\partial t_r \,|_{t=0} = b^r\gamma^r x\, {}_0F_1(;q/2+1;x\tau/4)/2(q/2);$$

$$\sum_{r,s=1} b^{rs}\partial^2\, {}_0F_1(;\cdot;\cdot)/\partial t_r\partial t_s \,|_{t=0} = b^{rr} x\, {}_0F_1(;q/2+1;x\tau/4)/2(q/2) + b^{rs}\gamma^r\gamma^s x^2\, {}_0F_1(;q/2+2;x\tau/4)/2(q/2)_2;$$

$$\sum_{r,s,t=1} b^{rst}\partial^3\, {}_0F_1(;\cdot;\cdot)/\partial t_r\partial t_s\partial t_t \,|_{t=0} = [3]\, b^{rss}\gamma^r x^2\, {}_0F_1(;q/2+2;x\tau/4)/2(q/2)_2 + b^{rst}\gamma^r\gamma^s\gamma^t x^3\, {}_0F_1(;q/2+3;x\tau/4)/8q(q/2)_3;$$

$$b^{rstu}\partial^4\, {}_0F_1(;\cdot;\cdot)/\partial t_r\partial t_s\partial t_t\partial t_u \,|_{t=0} = [3]\, b^{rrss}x^2\, {}_0F_1(;q/2+2;x\tau/4)/2(q/2)_2 + [6]\, b^{rstt}\gamma^r\gamma^s x^3\, {}_0F_1(;q/2+3;x\tau/4)/8q(q/2)_3 + b^{rstu}\gamma^r\gamma^s\gamma^t\gamma^u x^4\, {}_0F_1(;q/2+4;x\tau/4)/16q(q/2)_4;$$

from which the following can be easily deduced

$$\begin{aligned}
b^r w^r \psi_q(\gamma^r;\delta^{rs}) &= b^r\gamma^r g_{q+2;\tau}(x); \\
b^{rs}w^r w^s \psi_q(\gamma^r;\delta^{rs}) &= b^{rs}\gamma^r\gamma^s g_{q+4;\tau}(x) + b^{rr}g_{q+2;\tau}(x); \\
b^{rst}w^r w^s w^t \psi_q(\gamma^r;\delta^{rs}) &= b^{rst}\gamma^r\gamma^s\gamma^t g_{q+6;\tau}(x) + 3b^{rrs}\gamma^s g_{q+4;\tau}(x); \\
b^{rstu}w^r w^s w^t w^u \psi_q(\gamma^r;\delta^{rs}) &= b^{rstu}\gamma^r\gamma^s\gamma^t\gamma^u g_{q+8;\tau}(x) + 6b^{rstt}\gamma^r\gamma^s g_{q+6;\tau}(x) + 3b^{rrss}g_{q+4;\tau}(x).
\end{aligned} \tag{24}$$

Expression (10) follows immediately after simple algebra, and applying the difference operator ∇^k $(k = 1,\ldots,4)$ to the various $g_{q+\cdot;\tau}(\cdot)$. ■

B Proof of Theorem 2

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

$$\sup_{B\in\mathcal{B}} \int_{(\partial B)^{\varepsilon}} \phi_{q,\gamma,\kappa}(w)\,dw = O(\varepsilon^{a}); \quad \varepsilon \downarrow 0$$

where (∂B), $(\partial B)^{\varepsilon}$ are the boundary of B and ε-neighbourhood of (∂B), respectively $a \in R_+$ and $\phi_{q,\gamma,\kappa}(\cdot)$ is the q dimensional multivariate normal distribution with mean γ^r and covariance matrix $\kappa_{r,s}$. Using Bhattacharya and Ghosh's (1978, Theorem 2(b)), it follows that a formal Edgeworth expansion for the distribution of W_r is given as follows:

$$\sup_{B\in\mathcal{B}} \left| P_{\lambda}(W_r \in B) - \int_B H(w)\,\phi_{q,\gamma,\kappa}(w)\,dw \right| = o(1/n) \tag{25}$$

where $H(w) = (-1)^{v}\,\partial_{r_1}\ldots\partial_{r_v}\phi_{q,\gamma,\kappa}(\cdot)/\partial w^{r_1}\partial w^{r_2}\ldots\partial w^{r_v}$ (see e.g. (22)) is the (fourth order) Edgeworth polynomial:

$$\begin{aligned} H(w) = 1 &+ \left(k^2_r h^r(w) + k^2_{r,s}h^{rs}(w)/2\right)/n^{1/2} + \left(k^3_r h^r(w) + \left(k^3_{r,s}/2 + k^2_r k^2_s\right) h^{rs}(w) + \left(k^2_{r,s,t}/6 + k^2_r k^2_{s,t}/2\right) h^{rst}(w) + k^2_{r,s}k^2_{t,u}h^{rstu}(w)/8\right)/n. \end{aligned} \tag{26}$$

In (26), the k's are the approximate cumulants obtained by the delta method as in (9). Hence a valid Edgeworth expansion for W_r is given by:

$$P_{\lambda}(W_r \leq u) = \int_{-\infty}^{u} H(w)\,\phi_{q,\gamma,\kappa}(w)\,dw + o(1/n): \tag{27}$$

which can be shown to be valid by the standard argument of Bhattacharya and Ghosh (1978) as the set A_{λ} is such that $Pr_{\theta_n}\left(A^C_{\lambda}\right) = o(1/n)$ given the assumptions. We can then apply Lemma 3.1 to the integral in (27), by considering the approximate cumulants k as the constants b^{R_v} appearing in the lemma and replacing γ with $\kappa_{r,s}^{1/2}\sqrt{\lambda}$. After some algebra it follows that:

$$\sup_{C\in\mathcal{C}} \left| P_{\theta_0}(W_r W_s \delta^{rs} \in C) - \sum_{j=0}\sum_{k=0} 1/n^{j/2}\,a_{jk} \int_C g_{q+2k,\tau}(x)\,dx \right| \tag{28}$$

for all Borel subset C of $\mathcal{C}$ satisfying:

$$\sup_{C \in \mathcal{C}} \int_{(\partial C)^{\varepsilon}} g_{q;\tau}(x)\, dx = O(\varepsilon); \quad \varepsilon \downarrow 0$$

where Ψ_{jk} is as in (12)and is obtained after some simplifications in $H(w)$. The validity of (28) follows from the classical result of Chandra and Ghosh (1980). ■

C Proof of Theorem 4

Under the assumptions, it is not difficult to show that the bootstrap maximum dual likelihood estimator $\hat{\lambda}_{BMDL}$ satisfies the bootstrapped dual likelihood equations $\partial W^*_{\theta_0}(\lambda)/\partial \lambda^r = 0$ with bootstrap probability $Pr^*_{\lambda}\ 1 - o(1/n)$ and it admits a stochastic expansion of the same form of (6) admits a stochastic expansion of the form (2:6) (under the null hypothesis i.e. $\widetilde{\lambda}^r = 0$); also, we can derive the stochastic expansion for the bootstrap empirical likelihood ratio and its signed square root version as in Section 2. Next, let $\mathcal{B}$ be the class of Borel sets satisfying :

$$\sup_{B \in \mathcal{B}} \int_{(\partial B)^{\varepsilon}} \phi_{q;\gamma;\kappa}(w)\, dw = O(\varepsilon^{a}); \quad \varepsilon \downarrow 0$$

for some $a > 0$. We can then use Bhattacharya's (1987, Theorem3.3) to deduce that

$$\sup_{B \in \mathcal{B}} \left| P(W^*_r \in B) - \int_B H^*(w)\, \phi_{q;\gamma;\kappa}(w)\, dw \right| = o(1/n) \qquad (29)$$

where $H^*(w)$ is the Hermite polynomial as in (26) with coefficients replaced by their bootstrap analog. Proceeding then as in Theorem 3.2 we obtain the required result. The validity of the expansion follows by using Chandra's (1985) theorem. ■

D Tables

TABLE 1[†]. Moment condition model, $N(0;1)$ observations

Nominal size	0.100			0.050			0.010		
n = 50	0.145[a]	0.127[b]	0.121[c]	0.091[a]	0.080[b]	0.073[c]	0.036[a]	0.029[b]	0.015[c]
n = 100	0.132[a]	0.119[b]	0.119[c]	0.079[a]	0.069[b]	0.064[c]	0.028[a]	0.020[b]	0.014[c]
n = 200	0.124[a]	0.110[b]	0.109[c]	0.068[a]	0.061[b]	0.059[c]	0.024[a]	0.019[b]	0.013[c]
n = 500	0.111[a]	0.109[b]	0.105[c]	0.061[a]	0.060[b]	0.058[c]	0.015[a]	0.015[b]	0.013[c]

† Based on 5000 replications. a original, b feasible Bartlett adjusted (16), and c bootstrapped empirical likelihood ratio test (18)

Table 2[†]. Moment condition model, $t(5)$ observations

Nominal size	0.100			0.050			0.010		
n = 50	0.192[a]	0.150[b]	0.143[c]	0.121[a]	0.093[b]	0.083[c]	0.052[a]	0.042[b]	0.032[c]
n = 100	0.151[a]	0.130[b]	0.123[c]	0.093[a]	0.083[b]	0.075[c]	0.039[a]	0.030[b]	0.021[c]
n = 200	0.132[a]	0.126[b]	0.119[c]	0.081[a]	0.076[b]	0.070[c]	0.024[a]	0.020[b]	0.018[c]
n = 500	0.129[a]	0.120[b]	0.110[c]	0.073[a]	0.069[b]	0.065[c]	0.021[a]	0.018[b]	0.015[c]

† Based on 5000 replications. a original, b feasible Bartlett adjusted (16), and c bootstrapped empirical likelihood ratio. ratio test (18)

Table 3[†]. Nonlinear heteroskedastic regression model with $N(0;1)$ innovations

Nominal size	0.100			0.050			0.010		
n = 50	0.131[a]	0.118[b]	0.113[c]	0.083[a]	0.73[b]	0.067[c]	0.035[a]	0.031[c]	0.026[c]
n = 100	0.115[a]	0.111[b]	0.109[c]	0.072[a]	0.065[b]	0.062[c]	0.026[a]	0.022[b]	0.020[c]
n = 200	0.109[a]	0.109[b]	0.108[c]	0.065[a]	0.062[b]	0.059[c]	0.021[a]	0.017[b]	0.013[c]
n = 500	0.105[a]	0.105[b]	0.105[c]	0.061[a]	0.060[b]	0.057[c]	0.018[a]	0.016[b]	0.016[c]

† Based on 5000 replications. a original, b feasible Bartlett adjusted (16), and c bootstrapped empirical likelihood ratio ratio test (18)

Table 4[y]. Nonlinear heteroskedastic regression model with t(4) innovations

Nominal size	0:100			0:050			0:010		
n = 50	0:178[a]	0:164[b]	0:156[c]	0:113[a]	0:090[b]	0:072[c]	0:043[a]	0:034[b]	0:026[c]
n = 100	0:142[a]	0:134[b]	0:125[c]	0:092[a]	0:081[b]	0:069[c]	0:037[a]	0:030[b]	0:023[c]
n = 200	0:135[a]	0:129[b]	0:114[c]	0:073[a]	0:067[b]	0:063[c]	0:027[a]	0:021[b]	0:019[c]
n = 500	0:129[a]	0:120[b]	0:116[c]	0:070[a]	0:065[b]	0:062[c]	0:024[a]	0:021[b]	0:018[c]

y Based on 5000 replications. a original, b feasible Bartlett adjusted (16), and c bootstrapped empirical likelihood ratio ratio test (18)

Table 5[y]: Robust regression model, N (0; 1) innovations

Nominal size	0:100			0:050			0:010		
n = 50	0:167[a]	0:149[b]	0:121[c]	0:096[a]	0:082[b]	0:076[c]	0:036[a]	0:030[b]	0:025[c]
n = 100	0:142[a]	0:0131[b]	0:119[c]	0:079[a]	0:071[b]	0:064[c]	0:029[a]	0:023[b]	0:019[c]
n = 200	0:130[a]	0:121[b]	0:114[c]	0:068[a]	0:061[b]	0:059[c]	0:026[a]	0:021[b]	0:017[c]
n = 500	0:120[a]	0:118[b]	0:113[c]	0:061[a]	0:059[b]	0:058[c]	0:020[a]	0:018[b]	0:016[c]

y Based on 5000 replications. a original, b feasible Bartlett adjusted (16), and c bootstrapped empirical likelihood ratio test (18)

Table 6[y]: Robust regression model, t(4) innovations

Nominal size	0:100			0:050			0:010		
n = 50	0:187[a]	0:169[b]	0:158[c]	0:116[a]	0:093[b]	0:086[c]	0:057[a]	0:043[b]	0:039[c]
n = 100	0:152[a]	0:141[b]	0:132[c]	0:099[a]	0:084[b]	0:072[c]	0:043[a]	0:034[b]	0:030[c]
n = 200	0:141[a]	0:135[b]	0:127[c]	0:081[a]	0:075[b]	0:068[c]	0:036[a]	0:027[b]	0:023[c]
n = 500	0:132[a]	0:126[b]	0:120[c]	0:075[a]	0:064[b]	0:060[c]	0:029[a]	0:021[b]	0:021[c]

y Based on 5000 replications. a original, b feasible Bartlett adjusted (16), and c bootstrapped empirical likelihood ratio test (18)

Table 6[y]: Pseudo-likelihood model with N ($^{1\,i}$;1)

Nominal size	0:100		0:050		0:010	
n = 50	0:128[a]	0:116[b]	0:081[a]	0:077[b]	0:033[a]	0:021[b]
n = 100	0:112[a]	0:110[b]	0:072[a]	0:061[b]	0:027[a]	0:016[b]
n = 200	0:111[a]	0:104[b]	0:067[a]	0:059[b]	0:019[a]	0:013[b]
n = 500	0:109[a]	0:105[b]	0:061[a]	0:058[b]	0:016[a]	0:012[b]

y Based on 5000 replications. a original and b bootstrapped empirical likelihood ratio test

Table 7[y]: Pseudo-likelihood model with Pois ($^{1\,i}$)

Nominal size	0:100		0:050		0:010	
n = 50	0:138[a]	0:121[b]	0:084[a]	0:076[b]	0:038[a]	0:025[b]
n = 100	0:121[a]	0:115[b]	0:081[a]	0:065[b]	0:030[a]	0:019[b]
n = 200	0:120[a]	0:116[b]	0:069[a]	0:061[b]	0:021[a]	0:015[b]
n = 500	0:114[a]	0:110[b]	0:065[a]	0:059[b]	0:020[a]	0:014[b]

y Based on 5000 replications. a original and b bootstrapped empirical likelihood ratio test