



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<sup>¤</sup>

Francesco Bravo<sup>¥</sup>  
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  $O(1/n^2)$  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

---

<sup>¤</sup> 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 Podivinsky for some useful suggestions and comments. All remaining errors are my own responsibility.

<sup>¥</sup>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 confidence 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 profiling this multinomial likelihood with respect to certain constraints representing the only information available about  $F$ , one gets profiled or implied (using Barlow 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 correctability as shown in DiCiccio, Hall and Romano (1991).

In this paper we make the following contributions: firstly 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 find 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<sup>1</sup> 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  $k \in 1$  vector of weakly exogenous variables,

say  $w_i = (w_{i1} \dots w_{ik})^T$  (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  $r, s, \dots$  run from 1 to  $q$ , and the sum  $\sum \mathbf{P}$  is always intended as  $\sum_{i=1}^n \mathbf{P}_i$  unless otherwise stated.

## 2 The Relationship between Empirical and Dual Likelihood

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

The information about  $F_\mu$  is available in the form

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

for some specified value  $\mu_0$  of  $\mu$ , with the GS  $f_r(z_i; \mu) : \mathbb{R}^m \rightarrow \mathbb{R}^q$ ,  $q \geq 1$  vector of known measurable functions. Assume that the following conditions hold with probability 1 (w.p.1 henceforth):

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

GS2 i)  $E_{F_{\mu_0}} f_r(z_i; \mu_0) f_s(z_i; \mu_0)$  is positive definite and ii)  $E_{F_\mu} f_r(z_i; \mu) = 0$  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  $\mu$ .

The problem of testing the hypothesis  $H_0 : \mu = \mu_0$  can be formulated in terms of finding a probability measure  $P_\mu$  which is consistent with the constraint  $E_{P_{\mu_0}} f_r(z_i; \mu_0) = 0$  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  $P_\mu \in P_n$ , it turns out that the original constrained estimation of  $P_\mu$  can be expressed as a simple maximisation of a multinomial likelihood supported on the data over the empirical counterpart of the constraint  $E_{F_\mu} f_r(z_i; \mu_0) = 0$ .

Let  $p_i$  denote the  $i$ th 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  $H_0 : \mu = \mu_0$  is then given by solving the following program:

$$LR(\mu_0) = 2 \sup_{p_i} \sum_{i=1}^n \log np_i \quad \text{s.t.} \quad \sum_{i=1}^n p_i f_r(z_i; \mu_0) = 0; \quad \sum_{i=1}^n p_i = 1 \quad (1)$$

Let  $ch(S)$  denote the convex hull for the set  $S \subset R^q$  and  $\|k\|$  be the Euclidean norm; assume now that w.p.1:

$$E1 \quad \frac{1}{2} \leq ch \left( \frac{1}{n} \sum_{i=1}^n f_r(z_i; \mu_0) \right) \subset \frac{3}{4} \quad \text{for } n \text{ sufficiently large,}$$

$$E2 \quad E_F \|f_r(z_i; \mu)\|^3 < 1;$$

it then follows (from E1) that  $LR(\mu_0)$  exists and it is positive<sup>2</sup> (hence the distribution  $F_{\mu_0}$  attaining the supremum is unique). The required implied probabilities satisfying (1) can be found by a standard Lagrange multiplier argument, and are given by:

$$p_i(\lambda^r(\mu_0)) = \frac{1}{n} (1 + \lambda^r(\mu_0) f_r(z_i; \mu_0))$$

where  $\lambda^r(\mu_0)$  is a  $q \times 1$  vector of Lagrange multipliers. Moreover, by assumption E2, following Owen's (1990), we can show that  $LR(\mu_0) = 2 \log \hat{A}^2(q) + o(1)$  that it is a nonparametric version of Wilks' theorem. For notational simplicity, let  $\lambda^r(\mu_0) = \lambda^r$ . An empirical likelihood ratio test for the hypothesis  $H_0 : \mu = \mu_0$  is based on

$$W_{\mu_0}(\lambda^r) = 2 \sum_{i=1}^n \log np_i = 2 \sum_{i=1}^n \log (1 + \lambda^r f_r(z_i; \mu_0)) \quad (2)$$

which depends only on the Lagrange multipliers  $\lambda$  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  $H_0 : \mu = \mu_0$  can be thought of as testing the dual hypothesis  $H_0^\mu : \lambda = 0$ ; such a test can be formulated as in standard parametric inference with a dual likelihood ratio type test:

$$\max_{\lambda} W_{\mu_0}(\lambda) = \frac{W_{\mu_0}(\lambda)}{W_{\mu_0}(0)} = \frac{W_{\mu_0}(\lambda)}{W_{\mu_0}(0)} = 2 = W_{\mu_0}(2):$$

Indeed, following Mykland's (1995) approach, we can consider the empirical log-likelihood function  $W_{\mu}(\lambda)$  as an artificial log-likelihood: it is in fact a dual likelihood in  $\lambda$ . It is easy to see that, subject to integrability conditions,  $E_{F_{\lambda}} \exp W_{\mu_0}(\lambda) = 1$  (more generally for  $\lambda \neq 0$ ,  $E_{F_{\lambda}} \exp W_{\mu_0}(\lambda) \neq 1$ ), and Bartlett type identities as developed by Mykland (1994) hold for  $\lambda$  to first order. Let us introduce some notation: let  $U_{R_v}$  denote the  $v$ th mixed derivative array with respect to the dual parameter  $\lambda$ :

$$U_{R_v} = \frac{\partial^v W_{\mu_0}(\lambda)}{\partial \lambda_{r_1} \partial \lambda_{r_2} \cdots \partial \lambda_{r_v}}$$

with  $W_{\mu_0}(\lambda)$  defined as in (2), for any set of indices  $1 \leq r_1, r_2, \dots, r_v \leq q$  in the set  $R_v$ . Evaluating the resulting derivatives at the null dual hypothesis  $H_0^\mu : \lambda = 0$ , it is straightforward to see that:

$$U_{r_1 r_2 \dots r_v} \big|_{\lambda=0} = (i-1)^{v_i-1} (v-i-1)! \prod_{i=1}^v f_{r_i}(z_i; \mu) \dots f_{r_v}(z_i; \mu). \quad (3)$$

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  $\lambda$  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:

$$\sum_R E U_{R_{v_1}} U_{R_{v_2}} \dots U_{R_{v_k}} = 0 \quad (4)$$



where the sum is over all partitions  $R_{v_j} \ j :: j \ R_{v_k}$  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  $\cdot_{R_v} = EU_{R_v}$ ,  $\cdot_{R_{v_r}; R_{v_s}} = EU_{R_{v_r}; R_{v_s}}$  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

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

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

$$E \prod_{i=1}^3 f_r(z_i; \mu) \prod_{i=1}^3 f_s(z_i; \mu) \prod_{i=1}^3 f_t(z_i; \mu) = [3] E \prod_{i=1}^3 f_r(z_i; \mu) \prod_{i=1}^3 f_s(z_i; \mu) \prod_{i=1}^3 f_t(z_i; \mu) - 2 E \prod_{i=1}^3 f_r(z_i; \mu) \prod_{i=1}^3 f_s(z_i; \mu) \prod_{i=1}^3 f_t(z_i; \mu)$$

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  $H_n : \mu_n = \mu_0 + \beta n^{-1/2}$  ( $\beta$  is a non random  $q \in 1$  vector such that:  $0 < \beta^T \beta < 1$ ). The alternative hypothesis  $H_n$  induces a (local) dual alternative of the form  $H_n^a : \eta_n = \eta_0 + \beta n^{-1/2}$ , hence in order to examine the local power of an empirical likelihood ratio test we should be considering the augmented hypothesis  $H_n^a : \eta_n = \eta_0 + \beta n^{-1/2}$  with the  $2p \in 1$  vector  $\eta_n = \begin{pmatrix} \beta \\ \eta_0 \end{pmatrix} n^{-1/2}$ . 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  $H_n^a : \eta_n = \eta_0 + \beta n^{-1/2}$ . (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  $H_n^a$ . Note that all the following inequalities

should be intended as componentwise.

Let  $\mathcal{U}_\delta := \mathcal{U}(0; \delta)$  be an open sphere on  $A_\delta$  with radius  $\delta_\delta > 0$ ,  $f(z_1; \mu_0) = U_r \cdot U_{rs} \cdot U_{rst} \cdot U_{rstu}$  be a vector in  $\mathbb{R}^k$  with  $k = \sum_{a=1}^3 q + a_i - 1$  and  $\delta$  a positive constant; assume that the following regularity conditions hold on some compact set  $A_\delta$  of the sample space for which assumptions GS1, GS2 and E1 hold w.p.1:

D1 Interchanging differentiation with respect to  $\delta_\delta$  and integration with respect to  $z$  is allowed in  $W_{\mu_0}(\delta_\delta)$ , and  $\sup_{\delta_\delta} E |f(z_1; \mu_0)|^{4+\delta} < 1$

D2  $\sup_{\delta_\delta} \sup_{k_1, \dots, k_r} |j^{\otimes k} W_{\mu_0}(\delta_\delta)|_{\otimes r_1, \dots, \otimes r_v} j^{5+\delta} < 1$ ;  $|j_{r_1} + \dots + r_v j|_{\otimes} = 5$ ;

D3  $E |U_{R_{v_1}} U_{R_{v_2}} \dots U_{R_{v_k}}| < 1$  for any partition  $R_{v_j} j \dots j R_{v_k}$  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  $A_\delta$ . First notice that under D1 and D2, we have for  $\otimes = 1, 2, 3$  and  $0 < \delta < 3=8$

$$P_{\delta_\delta} |j^{\otimes k} W_{\mu_0}(\delta_\delta)|_{\otimes r_1, \dots, \otimes r_v} | E (|j^{\otimes k} W_{\mu_0}(\delta_\delta)|_{\otimes r_1, \dots, \otimes r_v}) j^{n^{1=2}} > \pm n^{-\delta} = o(1=n); \quad \pm > 0$$

for  $|j_{r_1} + \dots + r_v j|_{\otimes} = \otimes$ , uniformly in  $A_\delta$ . It then follows by Chebyshev's inequality that

$$P_{\delta_\delta} |j^{\otimes k} W_{\mu_0}(\delta_\delta)|_{\otimes r_1, \dots, \otimes r_v} | E (|j^{\otimes k} W_{\mu_0}(\delta_\delta)|_{\otimes r_1, \dots, \otimes r_v}) j^{n^{1=2}} > \pm n^{-\delta} = o(1=n); \quad \pm > 0$$

$$P_{\delta_\delta} |j_{R_n} j|_{\otimes} > j_{\otimes} j^4 \pm n^{-\delta} = o(1=n); \quad \pm > 0$$

for

$$|j_{R_n} j|_{\otimes} \cdot j_{\otimes} j^4 \sup_{k_1, \dots, k_r} |j^{\otimes k} W_{\mu_0}(\delta_\delta)|_{\otimes r_1, \dots, \otimes r_v} = 24n; \quad |j_{r_1} + \dots + r_v j|_{\otimes} = 4.$$

Then, following Bhattacharya and Ghosh's (1978) approach, it is possible to show that on the set  $A_\delta$ , the maximum dual likelihood estimator  $\hat{b}_{MDL}$  satisfies the dual likelihood equations  $|j^{\otimes k} W_{\mu_0}(\delta_\delta)|_{\otimes r} = 0$  with  $P_{A_\delta}$  probability  $1 - o(1=n)$  (by von Bahr's



$$Z_{rst}^t = 3 + \cdot_{rstu} f_u = 2 \quad ; \quad \cdot_{rstu} f_u = 3 + \cdot_{rstu}^t u = 12 \quad ; \quad \cdot_{rst}^r s = (2n) + o_p(1=n) :$$

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:

$$\begin{aligned} W = & \quad i \quad Z_r i \quad \cdot_{rr^0} f_r^0 \quad Z_s i \quad \cdot_{ss^0} f_s^0 \quad \cdot_{rst}^r s + i \quad \cdot_{rst}^r u \quad \cdot_{sv}^t v \quad \cdot_{tw}^u u Z_u Z_v Z_w = 3 + \\ & \quad \cdot_{rt}^r s \quad \cdot_{su}^t v Z_r Z_t Z_u i \quad Z_r s \quad f_r f_s + \cdot_{rst}^r s \quad f_r f_s f_t = 3 \quad = n^{1=2} + \\ & \quad i \quad \cdot_{tu}^r s Z_r t Z_u + \cdot_{rtu}^t t^0 \quad \cdot_{uu^0}^t t^0 Z_t^0 Z_u^0 = 2 \quad \cdot_{vw}^v v \quad Z_s v Z_w + \cdot_{svw}^v v^0 \quad \cdot_{ww^0}^v v^0 Z_v^0 Z_w^0 = 2 \quad \cdot_{rs}^r s + \\ & \quad \cdot_{rstu}^r r^0 \quad \cdot_{ss^0}^t t^0 \quad \cdot_{tt^0}^u u^0 Z_r^0 Z_s^0 Z_t^0 Z_u^0 = 12 \quad i \quad \cdot_{ru}^r u \quad \cdot_{sv}^t v \quad \cdot_{tw}^u u Z_r s t Z_u Z_v Z_w = 3 + \\ & \quad Z_{rst} f_r f_s f_t = 3 \quad i \quad \cdot_{rstu} f_r f_s f_t f_u = 12 \quad = n + o_p(1=n) \end{aligned} \quad (7)$$

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  $\cdot_{rs} = i \quad \mathbf{P} \quad f_r(z_i; \mu) f_s(z_i; \mu)$  we can replace  $i \quad \cdot_{rs}$  with  $\cdot_{r;s}$ , whence we can find a  $q \in 1$  vector  $W_r$

$$\begin{aligned} W_r = & \quad Z_r^0 \quad \cdot_{r^0;r}^{r^0} r^{1=2} + f_r^0 \quad \cdot_{r^0;r}^{1=2} + Z_{r^0s} Z^s = 2 + \cdot_{r^0st} Z^s Z^t = 6 \quad i \\ & \quad Z_{r^0s} f_s = 2 \quad i \quad \cdot_{r^0st} Z^s f_t = 6 + \cdot_{r^0st} f_s f_t = 6 \quad \cdot_{r^0;r}^{r^0} r^{1=2} = n^{1=2} + \\ & \quad Z_{r^0st} Z^s Z^t = 6 + 3 Z_{r^0s} Z_{uv} \cdot_{s;u}^s u \quad \cdot_{t;v}^t v Z_t = 8 + \\ & \quad \cdot_{r^0stu} = 8 + \cdot_{r^0st^0} \quad \cdot_{u^0tu}^t t^0 \quad \cdot_{u^0}^t t^0 Z^s Z^t Z^u = 3 + \\ & \quad 5 \cdot_{s;t}^s t \quad \cdot_{tuv} Z_{r^0s} Z^u Z^v = 12 \quad i \quad \cdot_{s;u}^s u \quad \cdot_{t;v}^t v \quad \cdot_{uvw} Z_{r^0} Z_{st} f_w = 4 \quad i \\ & \quad \cdot_{s;t}^s t Z_{r^0} Z_{stu} f_u = 6 \quad i \quad \cdot_{s;t}^s t Z_{r^0s} Z_{tu} f_u = 8 \quad i \\ & \quad \cdot_{r^0stu} = 2 + \cdot_{v;v^0}^v v^0 \quad \cdot_{r^0sv} \quad \cdot_{v^0tu} Z^s Z^t f_u = 12 + Z_{r^0st} f_s f_t = 6 + \\ & \quad \cdot_{s;t}^s t \quad \cdot_{stuv} + \cdot_{s;t}^s t \quad \cdot_{t^0;u^0}^t t^0 \quad \cdot_{stt^0} \quad \cdot_{u^0uv} f_u f_v Z_{r^0} = 24 + \cdot_{s;t}^s t \quad \cdot_{tuv} Z_{r^0s} f_u f_v = 12 \quad i \\ & \quad \cdot_{r^0stu} + \cdot_{s^0;t^0}^s s^0 \quad \cdot_{r^0ss^0} \quad \cdot_{t^0tu}^t t^0 = 3 \quad f_s f_t f_u = 24 \quad \cdot_{r^0;r}^{r^0} r^{1=2} = n + o_p(1=n) ; \end{aligned} \quad (8)$$

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

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  $\alpha$  level coverage error for the confidence region  $R_\alpha = \{ \mu_0 : W \leq c_\alpha \}$  with the constant  $c_\alpha : \Pr(\hat{A}^2(q) < c_\alpha) = \alpha$  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  $o(1/n)$ ):

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

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:

$$\begin{aligned} k_r &= \gamma_{r^0; r^0}^{1/2} + k_{r^0}^2 = n^{1/2} + k_{r^0}^3 = n^{1/2} \gamma_{r^0; r^0}^{1/2} + o(1/n); \\ k_{r;s} &= \gamma_{r^0; s^0}^{1/2} + k_{r^0; s^0}^2 = n^{1/2} + k_{r^0; s^0}^3 = n^{1/2} \gamma_{r^0; r^0}^{1/2} \gamma_{s^0; s^0}^{1/2} + o(1/n); \end{aligned} \quad (9)$$

$$k_{r;s;t} = k_{r^0;s^0;t^0}^3 \cdot r^0;r^{1=2} \cdot s^0;s^{1=2} \cdot r^0;r^{1=2} = n + o(1=n);$$

$$k_{r_1;\dots;r_v} = o(1=n) \quad v \geq 4;$$

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

$$k_{r^0}^2 = i \cdot r^0;s;t \cdot s;t=6 + \cdot r^0;st \cdot f_s f_t=3; \quad k_{r^0}^3 = i \cdot s;u \cdot t;v \cdot r^0;s;t \cdot u;v;w \cdot f_w=6 i \cdot s;t \cdot r;s;t;u \cdot f_u=24 +$$

$$i \cdot r^0;stu=4 + \cdot s^0;t^0 \cdot r^0;ss^0 \cdot t^0;tu=18 f_s f_t f_u; \quad k_{r;s}^2 = \cdot r^0;s^0;t \cdot f_t=3;$$

$$k_{r^0;s^0}^3 = (i \cdot r^0;s^0;t;u=6 + \cdot r^0;t;s^0;u=2 i \cdot 4 \cdot r^0;s^0;t;u=3) \cdot t;u i \cdot (10 \cdot r^0;t;v \cdot s^0;uw=3 i$$

$$4 \cdot r^0;t;v \cdot s^0;uw=3 + \cdot r^0;t;v \cdot s^0;u;w=36) \cdot t;u \cdot v;w + 5 \cdot r^0;s^0;t \cdot uvw \cdot t;u \cdot v;w=3 +$$

$$i \cdot (5 \cdot r^0;s^0;t;u=6 + 2 \cdot r^0;s^0;t;u=3 i \cdot r^0;s^0;t;u=3) f_t f_u i \cdot 2 \cdot r^0;t;u \cdot s^0;vw \cdot t;u f_v f_w=9$$

$$i \cdot r^0;s^0;t \cdot uvw \cdot t;u f_v f_w=3;$$

$$k_{r^0;s^0;t^0}^3 = i \cdot r^0;s^0;t^0;u \cdot f_u=2 + \cdot r^0;s^0;t^0 \cdot uvw \cdot u;v \cdot f_w=12 + \cdot r^0;s^0;u \cdot t^0;vw \cdot u;v \cdot f_w=2;$$

Notice that the order of magnitude of the higher order cumulants  $k_{r_1;\dots;r_v}$  for  $v \geq 5$  is deduced by applying the general formulae of James and Mayne (1962). Also, the third and fourth cumulants are 0 up to  $o(1=n^{1=2})$  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  $\hat{A} : W_r \rightarrow W_r W_s \pm r;s$ .

Let  $g_{q;\zeta}(x)$  and  $G_{q;\zeta}(x)$  denote the density and the distribution function of a noncentral chi-square random variate with  $q$  degrees of freedom and non centrality parameter  $\zeta$ . Also, let  $r^k g_{q;\zeta}(x)$  be the  $k$ th (double) difference operator applied to the density  $g_{q;\zeta}(x)$  (i.e.  $r^k g_{q;\zeta}(x) = \sum_{j=0}^k (-1)^j \binom{k}{j} g_{q+2(k-j);\zeta}(x)$ ). 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  $\hat{A}_q(\circ^r; \pm^{rs})$  be the multivariate normal distribution with mean vector  $\circ^r = \circ^1 \dots \circ^q$  and identity covariance matrix  $\pm^{rs}$ , and  $b^{R_v}$  be a  $q^v$  dimensional array of constants not depending on  $n$  (i.e.  $b^r; b^{rs}; \dots$ ),  $v = 1; \dots; 4$ . Also let  $h_{R_v}$  be the  $v$ th order Hermite tensor defined by  $(j-1)^v @_{r_1} \dots @_{r_v} \hat{A}_q(\circ^r; \pm^{rs})$  (where  $@_{r_v} = @ = @w^{r_v}$ ), whose structure is reported in the Appendix for completeness. Assume that L1 the dominance condition  $\sup_{t \in N} \int \exp(j @ \hat{A}_q(\circ^r; \pm^{rs}) \exp(w^r \pm^{rs} t_s)) j dx < 1$  holds w.p.1. on an set  $N$  of  $t = 0$ .

Then the following holds:

$$\begin{aligned} b^r h^r \hat{A}(\circ^r; \pm^{rs}) &= b^{r \circ r} r g_{q; \zeta}(x); \\ b^{rs} h^{rs} \hat{A}(\circ^r; \pm^{rs}) &= b^{rr} r g_{q; \zeta}(x) + b^{rs \circ r \circ s} r^2 g_{q; \zeta}(x); \\ b^{rst} h^{rst} \hat{A}(\circ^r; \pm^{rs}) &= [3] b^{rss \circ r} r^2 g_{q; \zeta}(x) + b^{rst \circ r \circ s \circ t} r^3 g_{q; \zeta}(x); \\ b^{rstu} h^{rstu} \hat{A}(\circ^r; \pm^{rs}) &= [3] b^{rrss} r^2 g_{q; \zeta}(x) + [6] b^{rrst \circ s \circ t} r^3 g_{q; \zeta}(x) + \\ &\quad b^{rstu \circ r \circ s \circ s \circ u} r^4 g_{q; \zeta}(x); \end{aligned} \tag{10}$$

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

**Proof.** See Appendix A. ■

Let

$$\begin{aligned} B^{r;s} &= (j \cdot r; s; t; u = 6 + \cdot r; t; s; u = 2 j \cdot 4 \cdot r; s; t; u = 3 j \cdot 20 \cdot r; t; v \cdot s; u; w \cdot v; w = 3 + \\ &\quad 8 \cdot r; t; v \cdot s; u; w \cdot v; w = 3 + 10 \cdot r; s; t \cdot u; v; w \cdot v; w = 3) \cdot t; u; \end{aligned} \tag{11}$$

We can now prove the following theorem.

**Theorem 2** Let  $\mathbb{P} = j-1$ ; assume that the vector  $f(x_1; \mu_n)$  defined in Section 2 satisfies the following Cramér's condition:

$$\limsup_{k \rightarrow \infty} \frac{1}{k!} E \exp \left( \mathbb{P}^T f(x_1; \mu_n) \right) < 1$$

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

$$\sup_{u \in \mathbb{R}_+} \left| P_{\beta, n}(W \cdot u) \prod_{j=0}^{\infty} \prod_{k=0}^{\infty} \frac{1}{n^{j+2k}} a_{jk} \int_{\mathbb{R}_+} g_{q+2k, \beta}(x) dx \right| = o(1) \quad (12)$$

where  $\beta = \sigma \circ \sigma$ ,  $\sigma_{r,s}$ , and

$$a_{00} = 1; \quad a_{10} = \int (\sigma_{r,s;t} + 2 \cdot \sigma_{rst}) f_r f_s f_t = 6; \quad a_{11} = (\int \sigma_{r,s;t} + \sigma_{rst}) f_r f_s f_t = 3;$$

$$a_{12} = \int \sigma_{r,s;t} f_r f_s f_t = 6; \quad a_{20} = (\int \sigma_{r;t,v} \sigma_{s;u,w} + 5 \cdot \sigma_{rtv} \sigma_{suw}) f_r f_s f_t f_u f_v f_w = 72 +$$

$$(\int \sigma_{r;t,v} \sigma_{s;u,w} = 18 \int 5 \cdot \sigma_{r,s;t} \sigma_{uvw} = 72 + \int \sigma_{r;t,v} \sigma_{suw} = 12) \cdot \sigma_{r,s} f_t f_u f_v f_w +$$

$$(\int \sigma_{r,s;t;u} = 2 + \int \sigma_{rstu} = 4 + \int \sigma_{r;s,tu} = 3 \int \sigma_{rstu} = 6) f_r f_s f_t f_u + (5 \cdot \int \sigma_{r,s;t} \sigma_{u;v,w} = 24 +$$

$$5 \cdot \int \sigma_{r;t,v} \sigma_{suw} = 12 + 5 \cdot \int \sigma_{r,s;t} \sigma_{uvw} = 24) \cdot \sigma_{r,s} \sigma_{t,u} f_v f_w + (11 \cdot \int \sigma_{r;s,tu} = 24 +$$

$$+ \int \sigma_{r;s,tu} = 3 \int \sigma_{rstu} = 6) \cdot \sigma_{r,s} f_t f_u + B^{r,s} f_r f_s = 2 \int B^{r,s} \cdot \sigma_{r,s} = 2$$

$$a_{21} = \int (\sigma_{r;s,v} \sigma_{t;u,w} + 2 \cdot \sigma_{rtu} \sigma_{svw}) f_r f_s f_t f_u f_v f_w = 18 + (7 \cdot \int \sigma_{r;s,t} \sigma_{u;v,w} = 4 +$$

$$+ \int \sigma_{r;s,t} \sigma_{uvw} = 3 + \int \sigma_{r;t,v} \sigma_{s;u,w} = 18 + 2 \cdot \int \sigma_{r;t,v} \sigma_{suw} = 9 + \int \sigma_{rtv} \sigma_{suw} = 9) \cdot \sigma_{r,s} f_t f_u f_v f_w +$$

$$(5 \cdot \int \sigma_{r;s,t;u} = 6 + \int \sigma_{rstu} = 4 + 2 \cdot \int \sigma_{r;s,tu} = 3 \int \sigma_{rstu} = 3) f_r f_s f_t f_u + (\int 11 \cdot \int \sigma_{r;s,t;u} = 24 \int$$

$$\int \sigma_{r;s,tu} = 3 + \int \sigma_{rstu} = 6 \int \sigma_{r;s,t} \sigma_{uvw} \cdot \sigma_{t,u} = 3 \int 2 \cdot \int \sigma_{r;t,v} \sigma_{suw} \cdot \sigma_{t,u} = 3 \cdot \sigma_{r,s} f_t f_u \int$$

$$B^{r,s} f_r f_s + B^{r,s} \cdot \sigma_{r,s} = 2;$$

$$a_{22} = (\int \sigma_{r;s,v} \sigma_{t;u,w} = 12 + \int \sigma_{rtu} \sigma_{svw} = 18) f_r f_s f_t f_u f_v f_w + (\int 5 \cdot \int \sigma_{r;s,v} \sigma_{t;u,w} = 24 +$$

$$+ \int \sigma_{r;t,v} \sigma_{s;u,w} = 12) \cdot \sigma_{r,s} f_t f_u f_v f_w + (\int 5 \cdot \int \sigma_{r;s,v} \sigma_{tuw} = 12 \int 7 \cdot \int \sigma_{r;t,u} \sigma_{s;v,w} = 18) \cdot \sigma_{u,w} f_r f_s f_t f_u \int$$

$$\int \sigma_{r;s,t;u} f_r f_s = 6 + \int \sigma_{r;s,tu} f_r f_s = 3 \int \sigma_{rstu} f_r f_s = 6 + \int \sigma_{r;s,t;u} \cdot \sigma_{r,s} = 3 \int \sigma_{rt;su} \cdot \sigma_{r,s} = 4$$

$$+ 2 \cdot \int \sigma_{r;s,tu} \cdot \sigma_{r,s} = 3) f_t f_u + \int 5 \cdot \int \sigma_{rtv} \sigma_{suw} \cdot \sigma_{t,u} \cdot \sigma_{v,w} = 3 \int 7 \cdot \int \sigma_{r;t,v} \sigma_{s;u,w} \cdot \sigma_{t,u} \cdot \sigma_{v,w} = 72 +$$

$$7 \cdot \int \sigma_{r;t,u} \sigma_{svw} \cdot \sigma_{t,v} \cdot \sigma_{u,w} = 72 + 7 \cdot \int \sigma_{r;s,t} \sigma_{uvw} \cdot \sigma_{t,u} \cdot \sigma_{v,w} = 8 f_r f_s,$$

$$a_{23} = \int \sigma_{r;s,v} \sigma_{t;u,w} f_v f_w = 18 + 5 \cdot \int \sigma_{r;s,t} \sigma_{uvw} \cdot \sigma_{v,w} = 72 + \int \sigma_{r;s,v} \sigma_{tuw} \cdot \sigma_{v,w} = 12 \int \sigma_{r;s,t;u} = 12 f_r f_s f_t f_u,$$

$$a_{24} = \int \sigma_{r;s,v} \sigma_{t;u,w} f_r f_s f_t f_u f_v f_w = 72$$

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  $T_{z_0} = I\{W > z_0\}$  denote the  $z_0$  level empirical likelihood ratio test (where the constant  $z_0$  is such that  $\Pr(\hat{A}_q^2(t) > z_0) = \alpha$ , and  $I\{\cdot\}$  is the indicator function); it then follows that the local power  $\eta_W$  of  $T_{z_0}$  is given by:  $\eta_W = \Pr(T_{z_0} = 1 | H_n^a : \lambda = \lambda_n)$ . In the next corollary we give a second order asymptotic expansion for  $\eta_W$ .

**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:

$$\eta_W = 1 - G_{q;\lambda}(z_0) + C_1(z_0)/n^{1/2} + C_2(z_0)/n + o(1/n) \quad (13)$$

with the constants  $C_1(z_0)$  and  $C_2(z_0)$

$$\begin{aligned} C_1(z_0) &= \int_{z_0}^{\infty} g_{q;\lambda}(x) dx + \int_{z_0}^{\infty} g_{q+2;\lambda}(x) dx + \int_{z_0}^{\infty} g_{q+4;\lambda}(x) dx; \\ C_2(z_0) &= \int_{z_0}^{\infty} g_{q;\lambda}(x) dx + \int_{z_0}^{\infty} g_{q+2;\lambda}(x) dx + \int_{z_0}^{\infty} g_{q+4;\lambda}(x) dx + \\ &\quad \int_{z_0}^{\infty} g_{q+6;\lambda}(x) dx + \int_{z_0}^{\infty} g_{q+8;\lambda}(x) dx \end{aligned}$$

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

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

$$\int_{W^r W^r, z_0}^{\infty} \hat{A}_q(W^r; \theta^r) dW = 1 - G_{q;\lambda}(z_0):$$

■

It is interesting to note that the power function depends (to second order) also on the constant  $B^{r;s}, r;s$  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 than 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  $H_0^\pi : \mu^r = 0$ . First note from (9) that the third and fourth order cumulants of the signed square root  $W_r$  of  $W$  are  $O(1/n^{3/2})$  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  $\mathbb{P} = \mathbb{P}_1$ ; assume that the vector  $f(x_1; \mu_0)$  defined in D2 satisfies the following Cramér's condition:

$$\limsup_{k \rightarrow \infty} \frac{1}{k!} E \exp \left( i t^T f(x_1; \mu_0) \right) < 1$$

Then there exists constants  $a_{jk}^0$  (not depending on  $n$ ) such that the following holds:

$$\sup_{u \in \mathbb{R}_+} \left| P_0(W \cdot u) - \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{1}{j!k!} a_{jk}^0 \frac{Z^u}{g_{q+2j}(x)} dx \right| = o(1/n) \quad (14)$$

where

$$a_{00}^0 = 1; \quad a_{11}^0 = \frac{1}{2} B^{r;s, r;s=2}; \quad a_{12}^0 = B^{r;s, r;s=2};$$

**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  $O(1/n^{1/2})$  term vanishes, as well as the integral  $\int_{\mathbb{R}^q} h^{rs}(w) \hat{A}_q(w) dw$  (for  $r \neq s$ ). The validity of

expansion (13) follows by using Chandra's (1985) Theorem which holds for all Borel subset  $C$  of  $\mathbb{C}$  satisfying:

$$\sup_{C \subset \mathbb{C}} \int_{(C)^2} g_{q; \lambda}(x) dx = O(n^{-2}); \quad \lambda \neq 0:$$

■

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  $N(0; \pm r^s) + O(n^{-3/2})$  (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  $(1 + B^0/n)$  where  $B^0 = B^{r;s}$ ,  $r;s=q$  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:

$$\sup_{u \in \mathbb{R}_+} \left| P_0(W = (1 + B^0/n) \cdot u) - \int_{\mathbb{R}_+} g_q(x) dx \right| = o(n^{-1}) \quad (15)$$

with  $B^0$  (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  $o(n^{-1})$  is obtained by considering a valid Edgeworth expansion for  $W$ . In terms of a formal Edgeworth expansion, the error can be replaced by  $O(n^{-2})$  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  $o(n^{-3/2})$ ).

**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  $B^0$  itself is more complicated as we need to estimate these product of derivatives. Under the additional assumption:

D4 The vector  $f(x_i; \mu_0)$  defined in D2 satisfies the following moment condition

$$E \|f(x_i; \mu_0)\|^8 < 1,$$

we can consistently estimate the Bartlett factor  $B^0$  (i.e. the consistency follows by a straightforward application of Chebyshev's inequality) by introducing the array  $\mathfrak{g}^{(i)}$  for any set of indices  $i; j; \dots$  in  $I_\circ$  and  $1 \leq i; j; \dots \leq n$  such that its coefficients satisfy the criterion of unbiasedness and are given by the general formula  $(i-1)^{v_i-1} (\circ_i-1)! = (n-i-1)^{(\circ_i-1)}$  with  $\circ \leq n$  and  $(n-i-1)^{(\circ_i-1)} = (n-i-1)(n-i-2) \dots (n-i-\circ+1)$  (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

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

$$\begin{aligned} \mathbb{R}_{r;s} &= \sum_{ij} \mathbb{R}_{r;s}^{ij} f_r^i f_s^j = n; & \mathbb{R}_{rst} &= \sum_i f_r^i f_s^i f_t^i = n; & \mathbb{R}_{r;s;t} &= \sum_{ijk} \mathbb{R}_{r;s;t}^{ijk} f_r^i f_s^j f_t^k = n \\ \mathbb{R}_{rt;su} &= \sum_{ij} \mathbb{R}_{rt;su}^{ij} f_r^i f_t^j f_s^i f_u^j = n; & \mathbb{R}_{r;s;tu} &= \sum_{ijk} \mathbb{R}_{r;s;tu}^{ijk} f_r^i f_s^j f_t^k f_u^i = n; \\ \mathbb{R}_{r;s;t;u} &= \sum_{ijkl} \mathbb{R}_{r;s;t;u}^{ijkl} f_r^i f_s^j f_t^k f_u^l = n \end{aligned}$$

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

$$\begin{aligned} \mathbb{B}^0 &= \frac{1}{3} \left( \mathbb{R}_{r;s;t;u} = 6 + \mathbb{R}_{rt;su} = 2 + 4\mathbb{R}_{r;s;tu} = 3 + 20\mathbb{R}_{r;t;v} \mathbb{R}_{suw}^{v;w} = 3 + \right. \\ &\quad \left. 8\mathbb{R}_{r;t;v} \mathbb{R}_{suw}^{k;l;w} = 3 + 10\mathbb{R}_{r;s;t} \mathbb{R}_{uvw}^{k;l;w} = 3 + \mathbb{R}_{r;s} \mathbb{R}_{t;u} = 2 \right) \end{aligned} \quad (16)$$

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  $\mu_0$  shows in fact that:

$$\mathbb{B}^0 = B^0 + C^r(\mu) U^r = n^{1/2} + O_p(1/n)$$

(where  $C^r(\mu) = (\partial B^0 / \partial \mu^r) |_{\mu=\mu_0}$ ) which implies that the difference between  $W = (1 + B^0/nq)$  and  $W = 1 + \mathbb{B}^0/nq$  is given by the following integral:

$$\int_0^1 \dot{A}_q(w^r; \pm r^s) w^r w^r C^s(\mu) w^s dw = n^{3/2}$$

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  $B^0$  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  $\mathbf{f}^n_{i=1}$  denote a bootstrap sample obtained by the original sample  $\mathbf{f}_i g_{i=1}^n$ . Let also  $f_r^n(\mathbf{x}_i; \mu)$  denote a centered bootstrap GS; recentering here is essential as it makes the bootstrapped GS unbiased conditionally on the original sample. Let  $\mathbf{f}^n(z_1; \mu_0) = (U_r^n, U_{rs}^n, U_{rst}^n, U_{rstu}^n)$  be the bootstrap vector analogous to the one described in Section 2. Assume that with bootstrap probability  $P^n \rightarrow 1$  (w.b.p.1) the following holds

$$\text{BE1 } O(2) \text{ ch } f_r^n(\mathbf{x}_1; \mu_0) \approx f_r^n(\mathbf{x}_n; \mu_0) \text{ for } n \text{ sufficiently large}$$

which justifies the existence and positiveness of a bootstrapped empirical likelihood ratio for the parameter  $\mu$ .

Assume also that w.b.p.1.

$$\text{BD1 } \sup_{\mathbf{z}_1 \in \mathcal{Z}} E |f^n(\mathbf{z}_1; \mu_0)|^{4+\epsilon} < \infty$$

$$\text{BD2 } \sup_{\mathbf{z}_1 \in \mathcal{Z}} \sup_{\mathbf{k} \in \mathcal{K}} |j^{(0)} W_{\mu_0}(\mathbf{z}_1) = @_{\mathbf{z}_1} r_1 \dots @_{\mathbf{z}_1} r_v j^{5+\epsilon} < \infty; \quad |j r_1 + \dots r_v j^{(0)} = 5;$$

$$\text{BD3 } E |U_{R_{v_1}}^n, U_{R_{v_2}}^n, \dots, U_{R_{v_k}}^n| < \infty \text{ for any partition } R_{v_j} \text{ } j = 1, \dots, k \text{ of the set } R_v \text{ (see (4)).}$$

We can then prove the following theorem:

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

$$\limsup_{k \rightarrow \infty} E \exp \left( \frac{1}{k} \sum_{i=1}^k f^n(\mathbf{x}_i; \mu_0) \right) < \infty$$

holds, then conditional on the original sample  $\hat{\mathbf{A}}$ , there exist constants  $a_{jk}^n$  ( $k = 1, 2$ ) (not depending on  $n$ ) such that the following holds:

$$\sup_{\mathbf{u} \in \mathcal{R}_+} |P_0^n(W^n \cdot \mathbf{u})| \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \frac{1}{j! k!} a_{jk}^n \int_{\mathcal{Z}} g_{q+2k}(\mathbf{x}) d\mathbf{x} = o(1/n); \quad (17)$$

Proof. See Appendix C. ■

Let  $z_{\alpha}^n$  be now a constant such that  $P_0^n(W \leq z_{\alpha}^n) = \alpha$  (i.e. an  $\alpha$  level for the bootstrap test  $W^n$ ). 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  $W^n$ )** The level of the empirical likelihood ratio test with bootstrap corrected critical value  $z_{\alpha}^n$  is given by:

$$P_0(W \leq z_{\alpha}^n) = \alpha + O(1/n^2); \quad (18)$$

Proof. By direct comparison of expansion (13) with its bootstrap analog, as the difference  $a_{jki} - a_{jk}^n = o_p(1)$ , it follows that

$$\sup_{z \in \mathbb{R}_+} |P_0(W \leq z) - P_0^n(W \leq z)| = O(1/n^2);$$

and hence the results follows immediately replacing  $z$  with  $z_{\alpha}^n$ . ■

As originally suggested by Hall and LaScala (1991), the bootstrap distribution  $W^n$  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:

$$\sum_{B=1}^{n_b} W_B^n = q(1 + B_b/n) \quad (19)$$

for  $B_b$ . It should be noted that  $B_b \rightarrow B^0$  for  $n_b \rightarrow \infty$ , 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

$$f_r^a(x_i^a; \mu) = f(x_i^a; \mu) - E^a f(x_i^a; \hat{\mu})$$

where  $\hat{\mu}$  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:

$$E f(x_i; \mu_0)^T = E \begin{bmatrix} x_{1i}^2 - \mu_{10} & x_{2i}^2 - \mu_{20} & \cdots & x_{qi}^2 - \mu_{q0} \end{bmatrix} \quad (20)$$

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  $H_0: \mu_0^T = \begin{bmatrix} 1 & \cdots & 1 \end{bmatrix}$  and  $\mu_0^T = \begin{bmatrix} 1_2 & \cdots & 1_2 \end{bmatrix}$  ( $1_2 = 5((1=2) - (5=2))^{1/2}((3=2))^2 = 5=3$ ), 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 than 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

$$\varepsilon(y_i; x_i^r; \tau_r) = y_i - g(x_i^r; \tau_r)$$

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

$$E(\varepsilon(y_i; x_i^r; \tau_r) | x) = 0$$

which implies an unconditional moment restriction of the form

$$E A(x_i^r) \varepsilon(y_i; x_i^r; \tau_r) = 0$$

for some  $q \times n$  matrix of instruments  $A(x_i^r)$ . The matrix of optimal instruments is  $A(x_i^r) = \frac{1}{\sigma_{ij}} \frac{\partial \varepsilon(y_i; x_i^r; \tau_r)}{\partial x_j^r}$  (where  $\frac{\partial \varepsilon(y_i; x_i^r; \tau_r)}{\partial x_j^r} = \frac{\partial}{\partial x_j^r} (y_i - g(x_i^r; \tau_r)) = -\frac{\partial g(x_i^r; \tau_r)}{\partial x_j^r}$  and  $\sigma_{ij} = E(\varepsilon(y_i; x_i^r; \tau_r) \varepsilon(y_j; x_j^r; \tau_r) | x)$ ; assuming further that the conditional variance takes the following functional specification  $\sigma_{ii} = h(x_i^r)$ , for some measurable function  $h(\cdot) : \mathbb{R}^q \rightarrow \mathbb{R}^+$  ( $q^0 \leq q$ ) we obtain the optimally weighted GS

$$E \frac{1}{\sigma_{ii}} \varepsilon(y_i; x_i^r; \tau_r) = 0;$$

The corresponding bootstrapped GS is then based on

$$E \frac{1}{\sigma_{ii}} \varepsilon(y_i^s; x_i^s; \tau_s) = \frac{1}{\sigma_{ii}} (y_i^s - g(x_i^s; \tau_s)) - \frac{1}{\sigma_{ii}} \varepsilon(y_i; x_i^r; \tau_r) = 0;$$

where  $y_i^s = x_i^{r,s} \beta_r + \varepsilon_i^s$  is the  $i$ th bootstrap pseudo-observation,  $\beta_r$  is a heteroskedasticity corrected non linear least square estimator the unknown  $q \times 1$  vector  $\tau_r$ ,  $\varepsilon_i^s$  is the bootstrap sample drawn from  $\varepsilon_i = y_i - x_i^r \beta_r$ ,  $x_i^{r,s}$  are sampled (independently from  $\varepsilon_i^s$ ) from the empirical distribution of  $x^r$ , and  $\sigma_{ii}^s = E \varepsilon_i^{s,s}$ . In the Monte Carlo study, we consider the following specification for  $g(\cdot)$ :  $y_i = \exp(\tau_0 + \tau_1 x_i^1) + \varepsilon_i$  with innovations  $\varepsilon_i \gg N(0, 1)$  or  $\gg t(4)$  and heteroskedasticity function  $\sigma_{ii} = x_{i1}^2$ . 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

$$E x_i^r \tilde{A}(y_i - x_i^r \beta_r) = 0$$

for the psi-function  $\tilde{A} : \mathbb{R} \rightarrow \mathbb{R}$  satisfying  $E \tilde{A}(\epsilon_i) = 0$ . The bootstrapped GS is

$$E^n x_i^r (\tilde{A}(y_i^* - x_i^r \beta_r) - \epsilon_i^*) = 0$$

where  $y_i^* = x_i^r \hat{\beta}_r + \epsilon_i^*$  is the bootstrap pseudo-observation,  $\hat{\beta}_r$  is an M-estimator for the unknown  $q \in 1$  vector  $\beta_r$ ,  $\epsilon_i^*$  is the bootstrap sample drawn from  $\epsilon_i = y_i - x_i^r \hat{\beta}_r$  and  $\epsilon_i^* = E^n \epsilon_i^*$ . Following Huber (1973), we specify the psi-function  $\tilde{A}(\epsilon)$  as

$$(y_i - x_i^r \beta_r) \{ f_j |y_i - x_i^r \beta_r| \cdot kg + k \epsilon \operatorname{sgn}(y_i - x_i^r \beta_r) \{ f_j |y_i - x_i^r \beta_r| \} > kg$$

with the constant  $k = 1.4$ , the scale parameter  $\frac{1}{4} = 1$ ,  $\operatorname{sgn}(\epsilon)$  and  $\{ f \epsilon \}$  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  $\beta_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$ .

Table 5 and 6 here

#### EXAMPLE 4. Quasi and Pseudo-likelihood models

We consider quasi and pseudo-likelihood models together because the analysis is 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  $q \in 1$  quasi-score

$$E @^r (x_i^r; \eta_r) = @^{-s} o_{ij}(\eta_r) y_{ji} - x_j^r; \eta_r = 0;$$

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

$$E^a @^r (x_i^r; \eta_r) = @^{-s} o_{ij}(\eta_r) j_{-r} = b_r^a j^a = 0$$

where  $b_i^a$  is a bootstrap sample drawn from the centered residuals  $b_i = b_i - \bar{b}$ ,  $b_i = n$  with  $b_i = o_{ii}^{-1=2} b_r y_{ji} - x_j^r; b_r$  and  $b_r$  is the quasi-maximum likelihood estimator for  $\eta_r$ .

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

$$E @^r (x_i^r; \eta_r) = @^{-s} o_{ij}(\eta_r) @^r x_j^s; \eta_r = @^{-t} @^r (x_k^r; \eta_r) = @^{-t} v^{kl}(\eta_r) (y_{li} - \eta(x_i^r; \eta_r)) = 0$$

for a given specification of the matrix  $v_{ij}(\eta_r) = \text{COV}(y_{li} - \eta(x_i^r; \eta_r)) y_{ji} - x_j^r; \eta_r$ .

The bootstrap analog is (as in the quasi-likelihood case)

$$E^a @^r (x_i^r; \eta_r) = @^{-s} o_{ij}(\eta_r) @^r x_j^s; \eta_r = @^{-t} @^r (x_k^r; \eta_r) = @^{-t} v^{kl,a}(\eta_r) = 0;$$

the only (important) difference being that the estimated residuals are obtained by using the matrix  $v_{ij}^{-1=2}$ . 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  $y_i \gg \text{Pois}(\lambda_i)$  with parameter  $\lambda_i = \exp(x_i^r \beta_r + \beta_i)$ , where  $x_i^r$  is a  $q \in 1$  vector of exogenous variables and  $\beta_i$  is a specification error. We assume that:  $E(\exp \beta_i) = 1$  and  $\text{VAR}(\exp \beta_i) = 1$ , and use  $N(\beta_i; 1)$  and  $\text{Pois}(\lambda_i)$  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  $\beta_i \gg N(0.35; 0.7056)$  (so that  $E(y_i) = 1$ ); Tables 7 and 8 report the results, for the null hypothesis  $H_0: \beta_r = \beta_0 \quad \beta_1 = 0$ .

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  $p \in 1$  parameter vector  $\beta_a$  (which may include  $\beta_r$  as well), so that residual specification. The resulting GS is  $E(\epsilon(y_i; x_i^r; \mu_{r^0}) | x) = 0$  with the  $2 \in 1$  vector

$$\epsilon(y_i; x_i^r; \mu_{r^0}) = y_i - g(x_i^r; \beta_r) - (y_i - g(x_i^r; \beta_r))^2 \beta_i h(x_i^r; \beta_a)$$

depending on the  $r^0 \in 1$  ( $r^0 = q + s$ ) vector of parameters, for some measurable function  $h(\cdot): \mathbb{R}^s \rightarrow \mathbb{R}^+$ . A straightforward calculation shows that the optimal instrument matrix is given in this case by

$$i^{r^0} = \begin{bmatrix} E(\epsilon(y_i; x_i^r; \mu_{r^0}) | x) = 0 & 0 \\ E(h(x_i^r; \beta_a) | x) = \beta_r & E(h(x_i^r; \beta_a) | x) = \beta_a \end{bmatrix}$$

covariance matrix of  $\epsilon_i$  and  $\epsilon_i^2$ . 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  $\sigma^2$ .

## 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  $\alpha$  mixing processes, Kitamura (1997) proves that it is still possible to obtain higher order accuracy (specifically up to  $O(n^{-5/6})$ ) 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  $C$  on the Borel field  $(\mathbb{R}^q; \mathcal{B}^q)$  supporting the unknown measure  $P_\mu$  to some pointwise separable, (to ensure measurability) Vapnik-Cervonenkis 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

$$\sup_{e \in E} |(P_n - P)e^T f(z; \mu_0)| \xrightarrow{a.s.} 0$$

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

$$\Pr \left( \inf_{e \in E} P_n e^T f(z; \mu_0) > \epsilon \right) \leq \frac{1}{2} e^{-\frac{3}{4} n \epsilon^2} \quad \text{all but finitely often w.p.1} \quad (21)$$

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-Cervonenkis inequality (see for example Gaenssler (1983, Lemma 10)). This fact in turns implies that  $0 \in \text{int} \{f(z_1; \mu_0) > 0, \dots, f(z_2; \mu_0) > 0\}$  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.
- Barndorff-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 specification 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 confidence 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, Efficient instrumental variables estimation of nonlinear models, *Econometrica* 58, 809{837.
- Owen, A.: 1988, Empirical likelihood ratio confidence intervals for a single functional, *Biometrika* 36, 237{249.
- Owen, A.: 1990, Empirical likelihood ratio confidence 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, Efficient 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.
- Gibsov, 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:

$$\begin{aligned} h^r &= w^r; \quad h^{rs} = h^r h^s; \quad h^{rst} = h^r h^s h^t; \quad [3] h^{\pm st}; \\ h^{rstu} &= h^r h^s h^t h^u; \quad [6] h^r h^s \pm^{tu} + [3] \pm^{rs} \pm^{tu}; \end{aligned} \quad (22)$$

Let  $\zeta = {}^\circ r {}^\circ r$ ,  $\hat{A}_q({}^\circ r; \pm^{rs})$  be the  $q$  variate normal density with mean  ${}^\circ r$  and identity covariance matrix  $\pm^{rs}$ , and  $t_r$  be a vector of auxiliary real variables: Also, let  $w^{R^\circ} = w^{r_1} w^{r_2} \dots w^{r^\circ}$  and  $b^{R^\circ} = b^{r_1 r_2 \dots r^\circ}$ . To prove the lemma, we use the transformation  $T : w^r \rightarrow (x; v^r)$  (with  $x = w^r w^r$ ,  $v^r = w^r = (w^s w^s)^{1/2}$  and Jacobian  $J = x^{q=2i-1=2}$ ) and the following identity:

$$w^{R^\circ} b^{R^\circ} \hat{A}_q({}^\circ r; \pm^{rs}) = \sum_{R^\circ} b^{R^\circ} \hat{A}_q({}^\circ r; \pm^{rs}) \exp(w^r \pm^{rs} t_s) \Big|_{t_r=0}$$

where  $\sum_{R^\circ}$  indicates summation over the partition  $\{r\} = \{r_1, \dots, r_p\}$  of  ${}^\circ$  indices into  $p$  non-empty blocks such that the resulting homogeneous polynomial in  ${}^\circ R^\circ$  is even or odd according to the number of indices in the set  $R^\circ$ . (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:

$$(2^{1/4})^{i^{q=2}} \sum_{R^\circ} b^{R^\circ} \exp f_i(x + \zeta) = 2g J \int \exp \left( x^{1=2} v^r \pm^{rs} t_s \right) \Big|_{t_r=0} (v^r dv^r). \quad (23)$$

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:

$$C(x; \zeta) = \sum_{R^\circ} b^{R^\circ} {}_0F_1(; q=2; x(\zeta + t^r t^r + 2 {}^\circ r t^r) = 4) \Big|_{t_r=0}$$

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

$$\sum_{r=1}^{\infty} b^r {}_0F_1(; \zeta; \zeta) = \sum_{t=0}^{\infty} b^{\circ r} x {}_0F_1(q=2+1; x\zeta=4) = 2(q=2);$$

$$\sum_{r,s=1}^{\infty} b^{rs} {}_0F_1(; \zeta; \zeta) = \sum_{t=0}^{\infty} b^{rr} x {}_0F_1(q=2+1; x\zeta=4) = 2(q=2) + b^{rs \circ r \circ s} x^2 {}_0F_1(q=2+2; x\zeta=4) = 2(q=2)_2;$$

$$\sum_{r,s,t=1}^{\infty} b^{rst} {}_0F_1(; \zeta; \zeta) = \sum_{t=0}^{\infty} [3] b^{rss \circ r} x^2 {}_0F_1(q=2+2; x\zeta=4) = 2(q=2)_2 + b^{rst \circ r \circ s \circ t} x^3 {}_0F_1(q=2+3; x\zeta=4) = 8q(q=2)_3;$$

$$\begin{aligned} b^{rstu} {}_0F_1(; \zeta; \zeta) &= \sum_{t=0}^{\infty} [3] b^{rss} x^2 {}_0F_1(q=2+2; x\zeta=4) = 2(q=2)_2 + \\ &[6] b^{rstt \circ r \circ s} x^3 {}_0F_1(q=2+3; x\zeta=4) = 8q(q=2)_3 + \\ &b^{rstu \circ r \circ s \circ t \circ u} x^4 {}_0F_1(q=2+4; x\zeta=4) = 16q(q=2)_4; \end{aligned}$$

from which the following can be easily deduced

$$\begin{aligned} b^r w^r \hat{A}_q(\circ r; \pm^{rs}) &= b^{\circ r} g_{q+2;\zeta}(x); \\ b^{rs} w^r w^s \hat{A}_q(\circ r; \pm^{rs}) &= b^{rs \circ r \circ s} g_{q+4;\zeta}(x) + b^{rr} g_{q+2;\zeta}(x); \\ b^{rst} w^r w^s w^t \hat{A}_q(\circ r; \pm^{rs}) &= b^{rst \circ r \circ s \circ t} g_{q+6;\zeta}(x) + 3b^{rss \circ s} g_{q+4;\zeta}(x); \\ b^{rstu} w^r w^s w^t w^u \hat{A}_q(\circ r; \pm^{rs}) &= b^{rstu \circ r \circ s \circ t \circ u} g_{q+8;\zeta}(x) + 6b^{rstt \circ r \circ s} g_{q+6;\zeta}(x) + 3b^{rrss} g_{q+4;\zeta}(x); \end{aligned} \quad (24)$$

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

## B Proof of Theorem 2

Let  $B$  be the class of Borel sets satisfying :

$$\sup_{B \in \mathcal{B}} \int_{(\partial B)^2} \hat{A}_{q; \circ; \cdot} (w) dw = O(n^{-2}) ; \quad 2 \neq 0$$

where  $(\partial B)$ ,  $(\partial B)^2$  are the boundary of  $B$  and  $2$ -neighbourhood of  $(\partial B)$ , respectively a  $2 \in \mathbb{R}_+$  and  $\hat{A}_{q; \circ; \cdot} (\cdot)$  is the  $q$  dimensional multivariate normal distribution with mean  $\circ^r$  and covariance matrix  $\cdot_{r;s}$ . 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:

$$\sup_{B \in \mathcal{B}} |P_{\mu_0}(W_r \in B) - \int_B H(w) \hat{A}_{q; \circ; \cdot} (w) dw| = o(1/n) \quad (25)$$

where  $H(w) = (i-1)^v \circ_{r_1} \dots \circ_{r_v} \hat{A}_{q; \circ; \cdot} (\cdot) = \circ w^{r_1} \circ w^{r_2} \dots \circ w^{r_v}$  (see e.g. (22)) is the (fourth order) Edgeworth polynomial:

$$H(w) = 1 + k_r^2 h^r(w) + k_{r;s}^2 h^{rs}(w) + \frac{1}{2} n^{1/2} + k_r^3 h^r(w) + k_{r;s}^3 + \frac{1}{2} k_r^2 k_s^2 h^{rs}(w) + k_{r;s;t}^2 + \frac{1}{6} k_r^2 k_{s;t}^2 h^{rst}(w) + k_{r;s}^2 k_{t;u}^2 h^{rstu}(w) + \dots = 1 + \dots \quad (26)$$

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:

$$P_{\mu_0}(W_r \in u) = \int_{u-1}^{u+1} H(w) \hat{A}_{q; \circ; \cdot} (w) dw + o(1/n) : \quad (27)$$

which can be shown to be valid by the standard argument of Bhattacharya and Ghosh (1978) as the set  $A_{\mu_0}$  is such that  $Pr_{\mu_0} A_{\mu_0}^C = o(1/n)$  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  $\circ$  with  $\cdot_{r;s}^{1/2} f_s$ . After some algebra it follows that:

$$\sup_{C \in \mathcal{C}} |P_{\mu_0}(W_r W_{s \pm}^{rs} \in C) - \int_C g_{q+2k; \cdot} (x) dx| \leq \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} 1/n^{j/2} a_{jk} \quad (28)$$

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

$$\sup_{C \subset \mathbb{C}} \int_C \frac{1}{(z - \bar{z})^2} g_{q, \bar{z}}(x) dx = O(n^{-2}); \quad n \neq 0$$

where  $a_{jk}$  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  $\hat{\mu}_{\text{BMDL}}$  satisfies the bootstrapped dual likelihood equations  $\partial W_{\mu_0}^n(\cdot) = \partial \hat{\mu}^n = 0$  with bootstrap probability  $\Pr_{\hat{\mu}^n} 1 + o(1/n)$  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.  $\hat{\mu}^n = 0$ ); 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  $B$  be the class of Borel sets satisfying :

$$\sup_{B \subset \mathbb{B}} \int_B \frac{1}{(w - \bar{w})^2} \hat{A}_{q, \bar{w}}(w) dw = O(n^{-2a}); \quad a > 0$$

for some  $a > 0$ . We can then use Bhattacharya's (1987, Theorem 3.3) to deduce that

$$\sup_{B \subset \mathbb{B}} \left[ \Pr(W_r^n \in B) - \int_B H^n(w) \hat{A}_{q, \bar{w}}(w) dw \right] = o(1/n) \quad (29)$$

where  $H^n(w)$  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<sup>y</sup>. Moment condition model , N (0; 1) observations

Nominal size	0:100			0:050			0:010		
n = 50	0:145 <sup>a</sup>	0:127 <sup>b</sup>	0:121 <sup>c</sup>	0:091 <sup>a</sup>	0:080 <sup>b</sup>	0:073 <sup>c</sup>	0:036 <sup>a</sup>	0:029 <sup>b</sup>	0:015 <sup>c</sup>
n = 100	0:132 <sup>a</sup>	0:119 <sup>b</sup>	0:119 <sup>c</sup>	0:079 <sup>a</sup>	0:069 <sup>b</sup>	0:064 <sup>c</sup>	0:028 <sup>a</sup>	0:020 <sup>b</sup>	0:014 <sup>c</sup>
n = 200	0:124 <sup>a</sup>	0:110 <sup>b</sup>	0:109 <sup>c</sup>	0:068 <sup>a</sup>	0:061 <sup>b</sup>	0:059 <sup>c</sup>	0:024 <sup>a</sup>	0:019 <sup>b</sup>	0:013 <sup>c</sup>
n = 500	0:111 <sup>a</sup>	0:109 <sup>b</sup>	0:105 <sup>c</sup>	0:061 <sup>a</sup>	0:060 <sup>b</sup>	0:058 <sup>c</sup>	0:015 <sup>a</sup>	0:015 <sup>b</sup>	0:013 <sup>c</sup>

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

Table 2<sup>y</sup>. Moment condition model , t (5) observations

Nominal size	0:100			0:050			0:010		
n = 50	0:192 <sup>a</sup>	0:150 <sup>b</sup>	0:143 <sup>c</sup>	0:121 <sup>a</sup>	0:093 <sup>b</sup>	0:083 <sup>c</sup>	0:052 <sup>a</sup>	0:042 <sup>b</sup>	0:032 <sup>c</sup>
n = 100	0:151 <sup>a</sup>	0:130 <sup>b</sup>	0:123 <sup>c</sup>	0:093 <sup>a</sup>	0:083 <sup>b</sup>	0:075 <sup>c</sup>	0:039 <sup>a</sup>	0:030 <sup>b</sup>	0:021 <sup>c</sup>
n = 200	0:132 <sup>a</sup>	0:126 <sup>b</sup>	0:119 <sup>c</sup>	0:081 <sup>a</sup>	0:076 <sup>b</sup>	0:070 <sup>c</sup>	0:024 <sup>a</sup>	0:020 <sup>b</sup>	0:018 <sup>c</sup>
n = 500	0:129 <sup>a</sup>	0:120 <sup>b</sup>	0:110 <sup>c</sup>	0:073 <sup>a</sup>	0:069 <sup>b</sup>	0:065 <sup>c</sup>	0:021 <sup>a</sup>	0:018 <sup>b</sup>	0:015 <sup>c</sup>

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

Table 3<sup>y</sup>. Nonlinear heteroskedastic regression model  
with N (0; 1) innovations

Nominal size	0:100			0:050			0:010		
n = 50	0:131 <sup>a</sup>	0:118 <sup>b</sup>	0:113 <sup>c</sup>	0:083 <sup>a</sup>	0:73 <sup>b</sup>	0:067 <sup>c</sup>	0:035 <sup>a</sup>	0:031 <sup>c</sup>	0:026 <sup>c</sup>
n = 100	0:115 <sup>a</sup>	0:111 <sup>b</sup>	0:109 <sup>c</sup>	0:072 <sup>a</sup>	0:065 <sup>b</sup>	0:062 <sup>c</sup>	0:026 <sup>a</sup>	0:022 <sup>b</sup>	0:020 <sup>c</sup>
n = 200	0:109 <sup>a</sup>	0:109 <sup>b</sup>	0:108 <sup>c</sup>	0:065 <sup>a</sup>	0:062 <sup>b</sup>	0:059 <sup>c</sup>	0:021 <sup>a</sup>	0:017 <sup>b</sup>	0:013 <sup>c</sup>
n = 500	0:105 <sup>a</sup>	0:105 <sup>b</sup>	0:105 <sup>c</sup>	0:061 <sup>a</sup>	0:060 <sup>b</sup>	0:057 <sup>c</sup>	0:018 <sup>a</sup>	0:016 <sup>b</sup>	0:016 <sup>c</sup>

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

Table 4<sup>y</sup>. Nonlinear heteroskedastic regression model  
with  $t(4)$  innovations

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

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

Table 5<sup>y</sup>: Robust regression model ,  $N(0; 1)$  innovations

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

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

Table 6<sup>y</sup>: Robust regression model ,  $t(4)$  innovations

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

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

Table 6<sup>y</sup>: Pseudo-Likelihood model with N (1<sup>i</sup>; 1)

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

<sup>y</sup> Based on 5000 replications. a original and b bootstrapped empirical likelihood ratio test

Table 7<sup>y</sup>: Pseudo-Likelihood model with Pois (1<sup>i</sup>)

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

<sup>y</sup> Based on 5000 replications. a original and b bootstrapped empirical likelihood ratio test