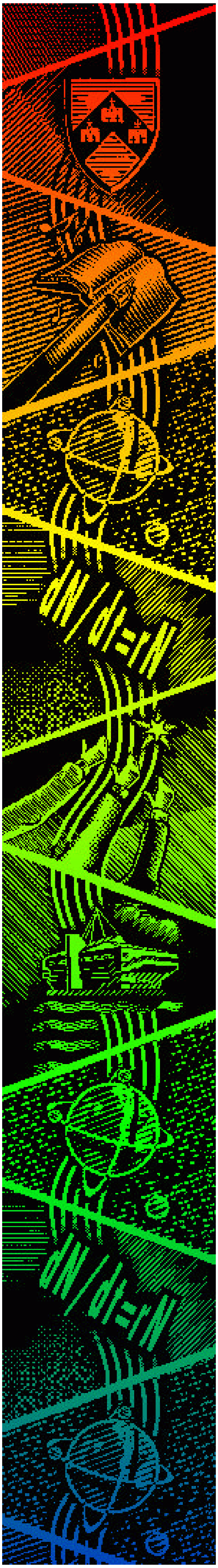

THE UNIVERSITY *of York*

Discussion Papers in Economics

No. 2000/28

Empirical Likelihood Specification Testing in Linear Regression Models

by

Francesco Bravo

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

Empirical Likelihood Specification Testing in Linear Regression Models*

Francesco Bravo†

University of York

Abstract

This paper analyses the higher order asymptotic behaviour of a profiled empirical likelihood ratio which can be used as a specification test in linear regression models. Despite the presence of nuisance parameters, a simple Bartlett correction factor is obtained and used to improve to third order the accuracy of commonly used tests such as the inclusion of irrelevant variables without any distributional assumptions about the error process.

* This paper is based on a revised version of parts of Chapter 3 of my Ph. D. dissertation at the University of Southampton. I would like to thank Grant Hillier, Andrew Chesher and Jan Podivinsky for useful suggestions and comments, and especially Song Xi Chen for clarifying an important point in the paper. Partial financial support under ESRC Grant R00429634019 is gratefully acknowledged. 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

1 Introduction

There has been growing interest in the last decade to developing nonparametric (i.e. distribution free) inferential techniques to analyse semiparametric econometric models. One such technique is based on estimating an unknown multinomial likelihood supported on the observations, subject to some constraints which are assumed to hold and represent the only information available in the sample. The resulting constrained multinomial probabilities can be used to construct a broad class of nonparametric, asymptotically χ^2 test statistics which can be interpreted as empirical goodness of fit type of tests (Baggerly, 1998). Examples of tests included in this broad class are the maximum entropy statistic (Efron, 1981), the Euclidean likelihood statistic (Owen, 1990), and the empirical likelihood ratio (Owen, 1988). All these tests are accurate[1] to an error of order $O_p\left(n^{-1}\right)$ by an Edgeworth expansion argument, however, as recently shown by Baggerly (1998), the empirical likelihood ratio is the only test admitting a Bartlett correction. This remarkable property implies the possibility of obtaining highly accurate inference in semiparametric models without resorting to other computationally more intensive competing methods such as the bootstrap, and can be explained by means of the dual likelihood theory developed for martingales by Mykland (1995). Specifically, in the case of a simple null hypothesis, the empirical likelihood ratio can be considered as an artificial likelihood[2] in the dual parameter, i.e. in the Lagrange multiplier associated with the constraints in the original maximisation problem. The existence of Bartlett type identities for the dual parameter combined with an Edgeworth expansion argument can then be advocated to justify the Bartlett correctability of the empirical/dual likelihood ratio test (for more details

[1] By accuracy we refer to how close nominal and actual coverage probabilities (and type I errors) are to each others.

[2] By artificial likelihood we mean 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 measures.

about the relationship between empirical and dual likelihood inference see Bravo (1999a)). Unfortunately when dealing with composite hypothesis the dual likelihood argument breaks down and the Bartlett correctability property is in general lost (see for example Bravo (1999b) or Lazar and Mykland (1999)).

In this paper we address the issue of the Bartlett correctability of the empirical likelihood ratio for a subset of the regression parameters in a linear regression model. By generalising the argument originally proposed by Chen (1994) for a simple regression model, we establish a general Bartlett correctability result for a profiled empirical likelihood which holds for any subset of the regression parameters to be tested. This interesting result is based on the same approach used by DiCiccio, Hall and Romano (1991) in their seminal paper about the Bartlett correctability of the empirical likelihood ratio test in the so called "smooth" function model (see Hall (1992) for a definition). By exploiting the well known property that in a (partitioned) linear regression model the least squares estimator for a subset of the regression parameters (in our case the nuisance parameters) can be expressed in terms of the least squares estimator of the remaining regression parameters (the parameters of interest), we can obtain a third order asymptotic expansion for the profiled empirical likelihood ratio which is a function only of the parameters of interest. This is the key point in the paper because we can then follow DiCiccio et al.'s (1991) approach, to show that the $O\left(n^{-1}\right)$ term appearing in the Edgeworth expansion for the signed square root of the empirical likelihood ratio is a linear function of a χ^2 random variate and hence adjusting the empirical likelihood ratio through multiplication or division by a constant of the form $1+B/n$ will eliminate the $O\left(n^{-1}\right)$ term in the adjusted statistic. The resulting Bartlett corrected empirical likelihood ratio test can be used to carry out standard specification analysis tests, which are third order accurate without any distributional assumption about the error process.

The remaining part of the paper is structured as follows: in the next section, after recalling briefly the main feature of empirical likelihood based inference, we develop a stochastic expansion for the profiled empirical likelihood ratio test which can be used as a specification test. Section 3 contains the main result of the paper, Section 4 some Monte Carlo evidence about the finite sample behaviour of the proposed test statistic, and Section 5 some concluding remarks. All the proofs are contained in the Appendix.

As it is customary in the literature on higher order asymptotics, we use (unless otherwise stated) tensor notation and the summation convention (i.e. for any two repeated indices, their sum over that index is understood) and adopt the following conventions in order to distinguish between parameters of interest and nuisance parameters: letters $r, s, \ldots$ etc. denote the original dimension of the parameter vector β, Greek letters index the $p \times 1$ vector $(p < k)$ vector parameters of interest, while the first four Roman letters $a, \ldots, d$ index the nuisance parameter.

2 A stochastic expansion for the profiled empirical likelihood ratio

Consider the following linear regression model in tensor notation:

$$y_i = x_i^r \beta_r + \varepsilon_i, \quad \varepsilon_i \sim i.i.d. \left(0, \sigma^2\right) \tag{1}$$

where x^r is a $1 \times k$ vector of weakly exogenous regressors -alternatively we can consider stochastic regressors, in which case we make the stronger assumption that x_i^r is independent of ε_i for all i. In the rest of the paper, we focus on the non random regressor case, as the random regressor case can be handled in the same manner.

Suppose we are interested to test the (simple) null hypothesis $H_0 : \beta_r = \beta_r^0$; clearly, under correct specification of (1), the following set of k orthogonality conditions

$$Ex_i^r \left(y_i - x_i^s \beta_s^0\right) = EZ^r = 0 \tag{2}$$

hold (with probability 1).

Let p_i denote the ith element of the unit simplex S in R^n and let $\widehat{p}_i = 1/n$ be the nonparametric maximum likelihood estimator for p_i.

An empirical likelihood ratio test for H_0 is obtained by solving the following constrained maximisation problem:

$$\begin{aligned} W_{\beta_r^0} &= \max_{p \in S} 2 \sum_{i=1}^{n} \log np_i \\ \text{s.t.} \quad p_i &\geq 0, \quad p_i Z_i^r = 0, \end{aligned} \tag{3}$$

where $S = \{p_i : \sum_{i=1}^n p_i = 1\}$. A Lagrange multiplier argument shows that the optimal probabilities $\{p_i\}_{i=1}^n$ are given by

$$p_i = \left(1 + \lambda_r Z_i^r\right)^{-1} / n$$

where λ_r is a vector of Lagrange multipliers. It follows that the solution for (3) is found by minimising

$$W_{\beta_r^0} = -2 \sum_{i=1}^{n} \log\left(1 + \lambda_r Z_i^r\right) \tag{4}$$

with respect to λ_r, which becomes the (dual) parameter of interest for a fixed value of β_r and suggest a dual likelihood interpretation for the empirical likelihood ratio test: the original null hypothesis $H_0 : \beta_r = \beta_r^0$ can be expressed in terms of its dual formulation $H_0 : \widehat{\lambda}_r = 0$, where $\widehat{\lambda}_r = \partial W_{\beta_r^0} / \partial \lambda_r = 0$ is the (unique) minimiser of (4).

Consider now the following partitioned regression model

$$y_i = x_i^a \beta_a + x_i^\alpha \beta_\alpha + \varepsilon_i, \tag{5}$$

and suppose that we are interested to test the subset β_α of the whole parameter vector β_r, so that the null hypothesis becomes $H_0^c : \beta_\alpha = \beta_\alpha^0$. An empirical likelihood ratio test for H_0^c can be obtained by profiling (i.e. maximising out) the nuisance parameter β_a, leading to the profiled empirical likelihood ratio

$$W_{\beta_\alpha^0} = \max_{\beta_a \in B} W_{\beta_r^0}, \quad B = \left\{\beta_a \mid \beta_\alpha = \beta_\alpha^0\right\} \tag{6}$$

which produces tests and confidence regions for β_α with asymptotic $\chi^2(p)$ calibration and coverage error $O_p(n^{-1})$, but, as opposed to the case of a simple null hypothesis, this coverage error cannot be improved to the order $O_p(n^{-2})$ by exploiting the duality between λ_r and β_r (see Lazar and Mykland (1999) for a simple example). The Bartlett type identities (Mykland, 1994) for the dual parameter λ_r cannot be advocated to eliminate to an order $O_p\left(n^{-3/2}\right)$ the terms involving the profiled parameter β_a appearing in the asymptotic expansions of the third and fourth order cumulants of (6). On the other hand, in their seminal paper about the Bartlett correctability property of the empirical likelihood ratio for a smooth function of a mean, DiCiccio et al. (1991) did not use these Bartlett type identities and obtained the Bartlett factor by a careful (and painstaking) examination of the algebraic structure of the third and fourth cumulant of the signed square root of the empirical likelihood ratio test.

As mentioned in the Introduction, the approach we follow in this paper is based on this latter approach: given a stochastic expansion for the whole parameter vector β_r, we first find a stochastic expansion for the nuisance parameter vector β_a, which, by the so-called Frish-Waugh theorem, is a function of the vector of parameters of interest β_α, The resulting expansion is then inserted in the original stochastic expansion for β_r, yielding (after some simplifications) to a stochastic expansion for the profiled empirical likelihood ratio which is simply a function of the parameter of interest β_α. This latter fact is the key factor to proving the Bartlett correctability of the statistic under investigation, as it allows us to use DiCiccio et al. (1991) technique to establish the asymptotic order of the relevant cumulants.

Let us introduce some additional notation and quantities. Let

$$V^{rs} = EZ_i^r Z_i^s/n = \sigma^2 x_i^r x_i^s/n, \quad V_{rs} = \left(V^{rs}\right)^{-1} \tag{7}$$

denote the $k \times k$ (average) covariance matrix associated with the orthogonality conditions (2), and its inverse. Partition V_{rs} as

$$V_{rs} = n \begin{bmatrix} s_{ab} & s_{a\alpha} \\ s_{\alpha a} & s_{\alpha\beta} \end{bmatrix}, \tag{8}$$

where

$$\begin{aligned} s_{ab} &= \left(x_i^a x_i^c\right)^{-1}\left(\delta^{cb} + \left(x_j^c x_j^\alpha\right)^{-1} s_{\alpha\beta}\left(x_j^\beta x_j^d\right)^{-1}\left(x_k^d x_k^b\right)^{-1}\right), \\ s_{\alpha\beta} &= \left(x_i^\alpha m_{ij} x_j^\beta\right)^{-1}, \quad s_{a\alpha} = -\left(x_i^a x_i^c\right)^{-1}\left(x_j^c x_j^\beta\right) s_{\alpha\beta}, \end{aligned}$$

define

$$U^{rs} = \left(V^{rs}\right)^{1/2} = \begin{bmatrix} u^{ab} & u^{a\alpha} \\ u^{\alpha a} & u^{\alpha\beta} \end{bmatrix}, \tag{9}$$

and

$$m_{ij} = \delta^{ij} - p_{ij}, \quad p_{ij} = x_i^a \left(x_k^a x_k^b\right)^{-1} x_j^b;$$

(i.e. the usual projection matrices in linear regression models). Finally define the following (averaged) arrays:

$$\begin{aligned} \kappa^{r_1 r_2 \ldots r_v} &= U^{r_1 s_1} U^{r_2 r_2} \ldots U^{r_v r_v} E\left(Z_i^{r_1} Z_i^{r_2} \ldots Z_i^{r_v}\right)/n \\ K^{r_1 r_2 \ldots r_v} &= U^{r_1 s_1} U^{r_2 r_2} \ldots U^{r_v r_v} Z_i^{r_1} Z_i^{r_2} \ldots Z_i^{r_v} - \kappa^{r_1 r_2 \ldots r_v} \end{aligned} \tag{10}$$

so that K^r is a $k \times 1$ averaged standardised vector of orthogonality conditions, K^{rs} is a $k \times k$ averaged matrix whose expectation is δ^{rs} (i.e. the Kronecker delta) and so on. Notice that $\kappa^{r_1 r_2 \ldots r_v}$ is $O_p(1)$ and $K^{r_1, r_2 \ldots r_k}$ is $O_p\left(n^{-1/2}\right)$.

Let $ch\{A\}$ denote the convex hull of the set $A \subseteq R^k$, λ_{1n} denote the smallest eigenvalue of $x_i^r x_i^s$, $m_{1n} = \max\{\|x_i^r\| : i = 1, 2, ..., n\}$, and $\|\cdot\|$ denote the Euclidean norm. Assume that the following regularity conditions hold uniformly in n (with probability 1)

A1 $0 \in ch\{Z_1^r, Z_2^r, ..., Z_n^r\}$ *for* n *sufficiently large,*

A2 x_i^r *is of full rank* k*,*

A3 $\underline{\lim}_{n\to\infty} \lambda_{1n}/n > 0, \quad m_{1n} = O\left(n^{\delta}\right)$ *for some* $\delta \in [0, 1/2)$,

A4 ε's have a non-zero absolutely continuous component which has a positive Lebesgue density on an open subset of R,

A5 ε's have finite 15th absolute moment and $\overline{\lim}_{n\to\infty} (1/n) \sum_{i=1}^{n} \left\|x_i^r\right\|^{15} < \infty$.

Remark I. Assumption A1 is standard in empirical likelihood theory, as it implies the existence and positiveness of the empirical likelihood ratio (4); as a consequence of this assumption, we emphasise that all the results presented in the paper are to be intended conditional on A1. Assumptions A2-A4 are standard in asymptotic theory for least squares (see for example Amemiya (1985, Ch. 3)): they imply the consistency and normality of the least squares estimator of β which is asymptotically equivalent, by a standard dual likelihood theory argument, to the maximum empirical likelihood estimator solution of $\partial W/\partial\beta = 0$. The remaining two assumptions are used to justify the existence of a valid Edgeworth expansion for the empirical likelihood ratio test. In particular A5 implies that there exists a set $\mathcal{K}_n$ such that

$$\Pr\left\{\widetilde{W}_{\beta_r^0}/n \in \mathcal{K}_n\right\} = 1 + O\left(n^{-5/2}\right)$$

(see for example Chandra and Ghosh (1979)), where $\widetilde{W}_{\beta_r^0}/n$ is given by

$$\begin{aligned}\widetilde{W}_{\beta_r^0}/n &= K^r K^r - K^{rs} K^r K^s + 2\kappa^{rst} K^r K^s K^t/3 + K^{rs} K^{ts} K^r K^t + \\ &\quad 2K^{rst} K^r K^s K^t/3 - 2\kappa^{rsu} K^{tu} K^r K^s K^t + \left(\kappa^{rsv}\kappa^{tuv} - \kappa^{rstu}/2\right) K^r K^s K^t K^u,\end{aligned} \tag{11}$$

i.e. a stochastic expansion for the empirical likelihood ratio test (4) (see (Chen, 1993)).

Consider now the profiled empirical likelihood ratio $W_{\beta_\alpha^0}$ defined in (6); to derive the required stochastic expansion we need to develop a stochastic expansion for the nuisance parameter β_a ; let

$$\widehat{\beta}_a = \left(x_i^a x_i^b\right)^{-1} x_j^b \left(y_j - x_j^\alpha \widehat{\beta}_\alpha\right), \quad \widehat{\beta}_\alpha = \left(x_i^\alpha m_{ij} x_j^\beta\right)^{-1} \left(x_i^\beta m_{ij} y_j\right) \tag{12}$$

be the usual partitioned least squares estimators for β_a and β_α. Consider the following arrays

$$\begin{aligned}
\widetilde{\varepsilon}_i &= y_i - x_i^a \widehat{\beta}_a - x_i^\alpha \beta_{\alpha 0}, \quad \gamma_i^r = U^{rs} x_i^s / n, \\
\zeta^{ra} &= \gamma_i^r x_i^a, \quad \zeta^{ab} = n\, \zeta^{ra}\, \zeta^{rb}, \\
\widetilde{\tau}_{rs} &= n \gamma_i^r \zeta^{sa} \widetilde{\varepsilon}_i \beta_a^1, \quad \tau_{rs}^{11} = n \zeta^{ra} \zeta^{sb} \beta_a^1 \beta_b^1 \\
\widetilde{\tau}_{rst} &= n^2 \gamma_i^r \gamma_j^s \zeta^{ta} \widetilde{\varepsilon}_i \widetilde{\varepsilon}_j \beta_a^1,
\end{aligned} \tag{13}$$

where β_a^1 is defined in (14) below. We can then prove the following proposition.

Proposition 1 *Suppose that β_a has a stochastic expansion of the following form:*

$$\beta_a = \widehat{\beta}_a + \beta_a^1 + \beta_a^2 + \beta_a^3 + O_p\left(n^{-2}\right), \tag{14}$$

where $\beta_a^j = O_p\left(1/n^{j/2}\right)$. Then

$$\begin{aligned}
\beta_a &= \widehat{\beta}_a + \left(x_i^a x_i^b\right)^{-1} x_j^b x_j^\alpha \left(\widehat{\beta}_\alpha - \beta_\alpha^0\right) - \zeta_{ab} \zeta^{rb} \times \\
&\quad \left[K_{\widehat{\beta}_a}^{rs} - \kappa^{rst}\left(K_{\widehat{\beta}_a}^{t} - \zeta^{tc}\beta_c^1\right)\right]\left(K_{\widehat{\beta}_a}^{s} - \zeta^{sc}\beta_c^1\right) + O_p\left(n^{-2}\right),
\end{aligned} \tag{15}$$

where $K_{\widehat{\beta}_a}^r$ is the set of $k \times 1$ orthogonality conditions (2) *evaluated at $\widehat{\beta}_a$, the least squares estimator of β_a defined in* (12).

Proof. See the Appendix. ■

Using (15), we can now write the stochastic expansion for the profiled empirical ratio test for the composite hypothesis $H_0^C : \beta_\alpha = \beta_{\alpha 0}$ as follows:

$$\begin{aligned}
\widetilde{W}_{\beta_a^0}/n &= \left(K_{\widehat{\beta}_a}^r - \zeta^{ra}\beta_a^1\right)\left(K_{\widehat{\beta}_a}^r - \zeta^{rb}\beta_b^1\right) - \left(K_{\widehat{\beta}_a}^r - \zeta^{ra}\beta_a^1\right) \times \\
&\quad \left(K_{\widehat{\beta}_a}^s - \zeta^{sb}\beta_b^1\right)\left(K_{\widehat{\beta}_a}^{rs} - 2\widetilde{\tau}_{rs} + \tau_{rs}^{11}\right) + 2\kappa^{rst}\left(K_{\widehat{\beta}_a}^r - \zeta^{ra}\beta_a^1\right) \times \\
&\quad \left(K_{\widehat{\beta}_a}^s - \zeta^{sb}\beta_b^1\right)\left(K_{\widehat{\beta}_a}^t - \zeta^{tc}\beta_c^1\right)/3 + K_{\widehat{\beta}_a}^{rt} K_{\widehat{\beta}_a}^{st} \times \\
&\quad \left(K_{\widehat{\beta}_a}^s - \zeta^{sb}\beta_b^1\right)\left(K_{\widehat{\beta}_a}^t - \zeta^{tc}\beta_c^1\right) + 2\left(K_{\widehat{\beta}_a}^r - \zeta^{ra}\beta_a^1\right) \times \\
&\quad \left(K_{\widehat{\beta}_a}^s - \zeta^{sb}\beta_b^1\right)\left(K_{\widehat{\beta}_a}^t - \zeta^{tc}\beta_c^1\right)\left(K_{\widehat{\beta}_a}^{rst} - 3\widetilde{\tau}_{rst}\right)/3 -
\end{aligned} \tag{16}$$

$$2\kappa^{rst}K^{tu}_{\widehat{\beta}_a}\left(K^{r}_{\widehat{\beta}_a}-\zeta^{ra}\beta^1_a\right)\left(K^{s}_{\widehat{\beta}_a}-\zeta^{sb}\beta^1_b\right)\times$$
$$\left(K^{u}_{\widehat{\beta}_a}-\zeta^{uc}\beta^1_c\right)-\zeta^{ab}\beta^2_a\beta^2_b+\left(\kappa^{rsv}\kappa^{tuv}-\kappa^{rstu}/2\right)$$
$$\left(K^{r}_{\widehat{\beta}_a}-\zeta^{ra}\beta^1_a\right)\left(K^{s}_{\widehat{\beta}_a}-\zeta^{sb}\beta^1_b\right)\left(K^{t}_{\widehat{\beta}_a}-\zeta^{tc}\beta^1_c\right)\times$$
$$\left(K^{u}_{\widehat{\beta}_a}-\zeta^{ud}\beta^1_d\right)+O_p\left(n^{-5/2}\right).$$

Noting that the $k\times 1$ vector $\left(K^{r}_{\widehat{\beta}_a}-\zeta^{ra}\beta^1_a\right)$ is equal to $\eta^{r\alpha}\left(\widehat{\beta}_\alpha-\beta^0_\alpha\right)$, where the $k\times p$ matrix $\eta^{r\alpha}$ is

$$\eta^{r\alpha}=u^{r\beta}x^{\beta}_i m_{ij}x^{\alpha}_j=u^{r\beta}s^{\alpha\beta}$$

and the matrix $u^{r\beta}$ is given by the upper and lower right corners of the matrix U^{rs} defined in (9), we obtain (by contracting over the index r)

$$\eta^{r\alpha}\eta^{r\beta}\left(\widehat{\beta}_\alpha-\beta^0_\alpha\right)\left(\widehat{\beta}_\alpha-\beta^0_\beta\right)=\sigma^2 s^{\alpha\beta}\left(\widehat{\beta}_\alpha-\beta^0_\alpha\right)\left(\widehat{\beta}_\alpha-\beta^0_\beta\right),$$

with $s^{\alpha\beta}=x^{\alpha}_i m_{ij}x^{\beta}_j$ which gives the leading term in the asymptotic expansion of the distribution of the profiled empirical likelihood ratio $W_{\beta^0_\alpha}$[3].

We can now express the stochastic expansion for the profiled empirical likelihood ratio given in (16) as a function of $\left(\widehat{\beta}_\alpha-\beta^0_\alpha\right)$; for notational convenience let $\left(\widehat{\beta}_\alpha-\beta^0_\alpha\right)$ $=\widehat{\beta}_\alpha$ and define the following arrays:

$$\begin{aligned}\xi^{rs}&=n^2U_{ra}U_{sb}p_{ij}\gamma^a_i\gamma^b_j,\ \varsigma^{rs}=n^2\gamma^r_i\gamma^s_j p_{ik}p_{jl}\varepsilon_k\varepsilon_l\\ \varpi^{rst}&=n^2U_{ru}U_{sv}\gamma^t_i Z^u Z^v\varepsilon_i.\end{aligned}\qquad(17)$$

Some algebra shows that the required expansion is

$$\widetilde{W}_{\beta^0_a}/n=\left(s^{\alpha\beta}/\sigma^2\right)\widehat{\beta}_\alpha\widehat{\beta}_\beta+\eta^{r\alpha}\eta^{s\beta}\left(-K^{rs}+2\xi^{rs}-\varsigma^{rs}+K^{rt}K^{st}\right)\widehat{\beta}_\alpha\widehat{\beta}_\beta+\qquad(18)$$

[3] As $VAR(\beta_\alpha)=\sigma^2 s_{\alpha\beta}/n$, a straightforward application of the central limit theorem yields

$$n^{1/2}\left(s^{\alpha\beta}/\sigma^2\right)^{1/2}\left(\widehat{\beta}_\beta-\beta^0_\beta\right)\xrightarrow{d}N\left(0,\delta^{\alpha\beta}\right),$$

whence the quadratic form

$$n\left(s^{\alpha\beta}/\sigma^2\right)\left(\widehat{\beta}_\alpha-\beta^0_\alpha\right)\left(\widehat{\beta}_\beta-\beta^0_\beta\right)\xrightarrow{d}\chi^2(p).$$

$$2\kappa^{rst}\eta^{r\alpha}\eta^{s\beta}\eta^{t\gamma}\widehat{\beta}_\alpha\widehat{\beta}_\beta\widehat{\beta}_\gamma/3 - \zeta_{ab}\zeta^{ra}\zeta^{sb}\eta^{t\alpha}\eta^{u\beta} \times$$
$$\left(K^{rs}K^{tu} - 2\kappa^{rst}\eta^{v\gamma}K^{uv}\widehat{\beta}_\gamma + \kappa^{rst}\kappa^{uvw}\eta^{v\gamma}\eta^{w\delta}\widehat{\beta}_\gamma\widehat{\beta}_\delta\right)\widehat{\beta}_\alpha\widehat{\beta}_\beta +$$
$$2\eta^{r\alpha}\eta^{s\beta}\eta^{t\gamma}\left(K^{rst} - 3\varpi^{rst} - 3\kappa^{rsu}K^{tu}\right)\widehat{\beta}_\alpha\widehat{\beta}_\beta\widehat{\beta}_\gamma/3 +$$
$$\left(\kappa^{rsv}\kappa^{tuv} - \kappa^{rstu}/2\right)\eta^{r\alpha}\eta^{s\beta}\eta^{t\gamma}\eta^{u\delta}\widehat{\beta}_\alpha\widehat{\beta}_\beta\widehat{\beta}_\gamma\widehat{\beta}_\delta + O_p\left(n^{-5/2}\right).$$

which is a function of the parameters of interest only.

In the next section we analyse the higher order asymptotic behaviour of (18): by finding its signed square root W^α (i.e. a $p \times 1$ random vector such that $\widetilde{W}_{\beta_a^0}/n = nW^\alpha W^\alpha + O_p\left(n^{-5/2}\right)$), we will show that W^α is sufficiently close to a multivariate normal vector with identity matrix, in the sense that its third and fourth order cumulants are $O\left(n^{-3/2}\right)$ and $O\left(n^{-2}\right)$, respectively, so that a simple scale adjustment to the mean of $W_{\beta_\alpha^0}$ will improve the coverage error to $O\left(n^{-2}\right)$.

3 Main result

Let $W^\alpha = W_1^\alpha + W_2^\alpha + W_3^\alpha$ be the signed square root of $\widetilde{W}_{\beta_a^0}/n$, where each subcomponent $W_j^\alpha = O_p\left(n^{-j/2}\right)$ is given by:

$$\begin{aligned}
W_1^\alpha &= \left(s^{\alpha\beta}\right)^{1/2}\widehat{\beta}_\beta/\sigma, \qquad (19)\\
\left(s^{\alpha\beta}/\sigma^2\right)^{1/2} W_2^\beta &= -\eta^{r\alpha}\eta^{s\beta}K^{rs}\widehat{\beta}_\beta/2 + \kappa^{rst}\eta^{r\alpha}\eta^{s\beta}\eta^{t\gamma}\widehat{\beta}_\beta\widehat{\beta}_\gamma/3,\\
\left(s^{\alpha\beta}/\sigma^2\right)^{1/2} W_3^\beta &= \eta^{r\alpha}\eta^{s\beta}\left(\xi^{rs} - \varsigma^{rs}/2 + K^{rt}K^{st}\right)\widehat{\beta}_\beta + \eta^{r\alpha}\eta^{s\beta}\eta^{t\gamma}\left(K^{rst} - 3\varpi^{rst}\right)\widehat{\beta}_\beta\widehat{\beta}_\gamma/3 -\\
&\sigma^2\kappa^{rst}\kappa^{uvw}\eta^{t\alpha}\eta^{u\beta}\eta^{v\gamma}\eta^{w\delta}\left(\zeta_{ab}\zeta^{ra}\zeta^{sb} + s_{\varepsilon\zeta}\eta^{r\varepsilon}\eta^{s\zeta}/9\right)\widehat{\beta}_\beta\widehat{\beta}_\gamma\widehat{\beta}_\delta/2 -\\
&\sigma^2\left(\zeta_{ab}\zeta^{ra}\zeta^{sb}\eta^{t\alpha}\eta^{u\beta}/2 + s_{\gamma\delta}\eta^{r\alpha}\eta^{s\beta}\eta^{t\gamma}\eta^{u\delta}/8\right)K^{rs}K^{tu}\widehat{\beta}_\beta +\\
&+\left(\sigma^2\kappa^{rst}\eta^{r\alpha}\eta^{s\beta}\eta^{t\gamma}s_{\varepsilon\zeta}\eta^{u\varepsilon}\eta^{v\zeta}/6 - \kappa^{rst}\eta^{r\alpha}\eta^{s\beta}\eta^{t\gamma}\right)K^{tu}\widehat{\beta}_\beta\widehat{\beta}_\gamma +\\
&\eta^{r\alpha}\eta^{s\beta}\eta^{t\gamma}\eta^{u\delta}\left(\kappa^{rsv}\kappa^{tuv}/2 - \kappa^{rstu}/4\right)\widehat{\beta}_\beta\widehat{\beta}_\gamma\widehat{\beta}_\delta.
\end{aligned}$$

In the next proposition, we evaluate the asymptotic order of the first four cumulants of $n^{1/2}W^\alpha$.

Proposition 2 *Let $k^{\alpha_1,\ldots,\alpha_v}$ denote the vth cumulant of $n^{1/2}W^\alpha$, let $X_i^\alpha = m_{ij}x_j^\alpha$ and $\rho_i^k = E\left(\varepsilon_i^k/\sigma^k\right)$ be the standardised kth moment; it then follows that:*

$$
\begin{aligned}
k^\alpha &= -\left(s_{\alpha\beta}\right)^{1/2} s_{\gamma\delta} X_i^\beta X_i^\gamma X_i^\delta \rho_i^3/6n^{1/2} + O\left(n^{-3/2}\right), \qquad (20)\\
k^{\alpha,\beta} &= \delta^{\alpha\beta} + \Big(\delta^{\alpha\beta} + \left(s_{\alpha\varepsilon}\right)^{1/2}\left(s_{\beta\zeta}\right)^{1/2} s_{\gamma\delta} X_i^\varepsilon X_i^\zeta X_i^\gamma X_i^\delta \rho_i^4/2n - \\
&\quad \left(s_{\alpha\varepsilon}\right)^{1/2}\left(s_{\beta\zeta}\right)^{1/2} s_{\gamma\eta} s_{\delta\iota} X_i^\varepsilon X_i^\gamma X_i^\delta X_j^\zeta X_j^\eta X_j^\iota \rho_i^3 \rho_j^3/3n - \\
&\quad \left(s_{\alpha\varepsilon}\right)^{1/2}\left(s_{\beta\zeta}\right)^{1/2} s_{\gamma\delta} s_{\eta\iota} X_i^\varepsilon X_i^\gamma X_i^\delta X_j^\zeta X_j^\eta X_j^\iota \rho_i^3 \rho_j^3/36n\Big) + O\left(1/n^2\right),\\
k^{\alpha_1,\ldots\alpha_v} &= O\left(1/n^{3/2}\right) \text{ or less for } v \geq 3.
\end{aligned}
$$

Proof. See the Appendix. ■

This asymptotic order for the cumulants of the signed square root of the profiled empirical likelihood ratio is a sufficient condition for proving the existence of a Bartlett correction via standard Edgeworth expansion theory, as shown in the next theorem.

Assume also that A5-A7 reported in the appendix hold with probability 1. Then we can prove the following theorem.

Theorem 3 *Under assumptions A2-A7, then*
(A) there exists a valid (in Bhattacharya and Rao (1976, Theorem 20.1) sense) Edgeworth expansion for the signed square root of the profiled empirical likelihood ratio test for the $p \times 1$ vector β_α^0.
Moreover, by using the transformation from R^p to R_+, $T : W^\alpha \rightarrow \widetilde{W}_{\beta_a^0}$
(B) the Edgeworth expansion for the profiled empirical likelihood ratio can be expressed as

$$\Pr\left\{\widetilde{W}_{\beta_\alpha^0} \leq c_\alpha\right\} = \alpha - Bc_\alpha g_p\left(c_\alpha\right)/n + O\left(n^{-3/2}\right), \qquad (21)$$

where the constant c_α is such that: $\Pr\left\{\widetilde{W}_{\beta_\alpha^0} \leq c_\alpha\right\} = \alpha$, *$g_p(\cdot)$ is the density function of a $\chi^2(p)$ random variate and B is the Bartlett factor:*

$$B = p + \left(X_i^\alpha s_{\alpha\beta} X_i^\beta\right)^2 \rho_i^4/2n - \left(X_i^\alpha s_{\alpha\beta} X_j^\beta\right)^3 \rho_i^3 \rho_j^3/3n^2. \qquad (22)$$

.

Proof. See the Appendix. ■

Remark II Notice that the resulting Bartlett correction differs from the "standard" adjustment for a multivariate mean in that there is an additional term which takes into consideration the fact that we are considering a profiled empirical likelihood ratio; this term would disappear in the case of no nuisance parameters.

Theorem 3 shows that an empirical likelihood confidence interval $I_\alpha = \left\{\beta_\alpha^{\mathbf{0}} \mid \widetilde{W}_{\beta_\alpha^0} \leq c_\alpha\right\}$ has coverage error $O\left(n^{-1}\right)$. Since:

$$\begin{aligned} E\left(\widetilde{W}_{\beta_\alpha^0}\right) &= nE\left(W_1^\alpha W_1^\beta + 2W_1^\alpha W_2^\beta + 2W_1^\alpha W_3^\beta + W_2^\alpha W_2^\beta\right) + O\left(n^{-2}\right) \\ &= p + B/n + O\left(n^{-2}\right), \end{aligned}$$

we can prove the following corollary to Theorem 3.

Corollary 4 *Under the conditions set forth in Theorem 3, then,*

$$\Pr\left\{\widetilde{W}_{\beta_\alpha^0}/E\left(\widetilde{W}_{\beta_\alpha^0}\right) \leq c_\alpha\right\} = \alpha + O\left(n^{-2}\right). \tag{23}$$

Proof. See the Appendix. ■

In practice the Bartlett factor B is not known, but it can be consistently estimated by replacing the expectation operator with its sample analogue, and the innovations with the least squares residuals $\widehat{\varepsilon}_i$, yielding the following empirical version of the Bartlett correction:

$$\widehat{B} = p + \left(X_i^\alpha s_{\alpha\beta} X_i^\beta\right)^2 \widehat{\rho}_i^4/2n - \left(X_i^\alpha s_{\alpha\beta} X_j^\beta\right)^3 \widehat{\rho}_i^3 \widehat{\rho}_j^3/3n^2. \tag{24}$$

Remark III It is worth noting that in the case of the empirical version of the Bartlett correction $\widehat{B}$ we may still obtain the same level of accuracy of Corollary 4, by noting that

$$\widehat{B} = B + R_n,$$

where the $O_p\left(n^{-1/2}\right)$ remainder R_n is actually of order $O_p\left(n^{-1}\right)$ by the even-odd properties of the Hermite tensors appearing in the Edgeworth expansion of the joint distribution of the components of W^α and $\widehat{B}$.

In the next section we illustrate how the present theory can be applied to commonly used specification tests which would correspond to standard F tests under the assumption of normality of the errors.

4 Empirical likelihood specification testing: some Monte Carlo evidence

Before illustrating the theory with some examples, we briefly discuss some computational aspects related to solving (numerically) the mathematical programs (4) and (6).

Firstly, notice that the two programs can be formulated as a saddlepoint problem:

$$W_{\beta_\alpha^0} = \max_{\beta_a \in B} \min_{\lambda_r} W_{\beta_r^0}; \tag{25}$$

given the convexity of the objective function in the dual parameter λ_r, the minimisation problem can be easily handled by any optimisation routines. Profiling the vector β_a out can be handled as in Owen (1990) by using a nested algorithm in which an optimisation routine at the outer level calls a function at the inner level that minimises λ_r. Alternatively, we can apply directly the multivariate Newton's algorithm to $W_{\beta_r^0}$ as a function of both λ_r and β_a; this amounts to Newton's method for solving the nonlinear system of $k + (k - p)$ first order conditions:

$$\begin{cases} \partial W_{\beta_r^0} / \partial \lambda_r = 0 \\ \partial W_{\beta_r^0} / \partial \beta_a = 0 \end{cases}$$

with starting point in the iterative process set to $(\lambda_r^{\mathbf{0}}, \beta_\alpha^{\mathbf{0}}) = \left(0, \widehat{\beta}_a\right)$ where $\widehat{\beta}_a$ is the least squares estimator for the nuisance parameter vector β_a. It is worth noting that the convergence of the Newton's method to a saddlepoint is known as a pitfall of optimisation routines; in the present case this is exactly the desired behaviour of the solution of (25). For the Monte Carlo study reported below we have used Newton's method as implemented in the NLSYS routine for the GAUSS programming language, with convergence to the required saddlepoint achieved in general after 4 or 5 iterations.

We now give 3 different examples which illustrate how we can use profiled empirical likelihood ratios for specification testing in linear models; given that we are working with a linear regression model, it seems useful to express the Bartlett correction factor (24) using the more familiar matrix notation. In order to do so, let

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

be the partitioned regression model described in (5) with β_2 the $p \times 1$ vector of parameters of interest, and

$$m^{ij} = M_{X_1} = I - X_1\left(X_1^T X_1\right)^{-1} X_1^T, \quad X_{M2} = M_{X_1}X_2, \quad s_{\alpha\beta} = S = \left(X_2^T M_{X_1} X_2\right)^{-1}.$$

The matrix version of the empirical Bartlett correction is then given by

$$\widehat{B} = p + \sum_{i=1}^{n}\left(X_{M2i}SX_{M2i}^T\right)^2 \widehat{\rho}_i^4/2n - \sum_{i=1}^{n}\sum_{j=1}^{n}\left(X_{M2i}SX_{M2j}^T\right)^3 \widehat{\rho}_i^3\widehat{\rho}_j^3/3n^2, \tag{26}$$

where X_{M2i} is the ith row of the $n \times p$ matrix X_{M2}.

EXAMPLE 1. **Test for the (overall) significance of the regressors**

The model under investigation is

$$y = \beta_0 + X\beta_1 + \varepsilon \tag{27}$$

and we want to test the hypothesis that $H_0^C : \beta_1 = 0$. In this case the dimension of the nuisance parameter (the intercept β_0) is 1, while $\dim(\beta_1) = p = k - 1$. The matrix $M_\iota = I - n^{-1}\iota\iota^T$ (where ι is a $n \times 1$ vector of ones) transforms the data to deviations from the mean, hence $X_{M2} = \left(X - \overline{X}\right)$ and $S = \left(\left(X - \overline{X}\right)^T\left(X - \overline{X}\right)\right)^{-1}$, respectively. The empirical Bartlett factor (26) is:

$$\widehat{B} = k-1+\sum_{i=1}^{n}\left[\left(X - \overline{X}\right)_i S\left(X - \overline{X}\right)_i^T\right]^2 \widehat{\rho}_i^4/2-n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left[\left(X - \overline{X}\right)_i S\left(X - \overline{X}\right)_j^T\right]\widehat{\rho}_i^3\widehat{\rho}_j^3/3. \tag{28}$$

EXAMPLE 2. **Test for the inclusion of irrelevant regressors**

The model under investigation is given by:

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon \tag{29}$$

and we want to test the hypothesis $H_0^C : \beta_2 = 0$. Then the empirical Bartlett factor is as in (26) with the appropriate degrees of freedom corresponding to $\dim(\beta_2)$.

EXAMPLE 3. **Test for a single regressor**

The last example deals with testing a specific value of one parameter and corresponds to a traditional (squared) t test under normality. The model considered is:

$$y = X_1\beta_1 + x_2\beta_2 + \varepsilon \tag{30}$$

with $\dim(\beta_1) = k - 1$ and $\dim(\beta_2) = 1$; the hypothesis under investigation is $H_0^C : \beta_2 = \beta_{20}$ for some specified value β_{20}. The empirical Bartlett correction is then as in (26) with 1 degree of freedom and the vector x_2 replacing the matrix X_2.

We now present some Monte Carlo evidence of the finite sample performance of the proposed empirical likelihood based specification tests; the model analysed is

$$y_i = \beta_0 + x_{1i}\beta_1 + x_{2i}\beta_2 + x_{3i}\beta_3 + x_{4i}\beta_4 + \varepsilon_i$$

and we consider both the cases of stochastic and fixed regressors.

The regressors are generated as $x_{ij} \sim N(0, I)$ for the stochastic case, while for the fixed regressors we generate x_1 as an equally spaced grid of numbers between -1 and

1, $x_2 = x_1^2$ and x_3 as the expected normal order statistic in sample of size n, so that there is no substantial leverage effect on the design; the last regressor is generated as $N(0,1)$ as we limit ourselves to testing for the inclusion of irrelevant regressor(s). The errors ε's are drawn independently from the x's and are specified as $N(0,1)$, $\chi^2(4)$ and $t(5)$. The first specification is useful to compare our approximation with the (exact) F statistic; the other two error specifications examines the effect of a skewed and of a symmetric but with thick tails distribution, respectively. All the following results are based on 5000 Monte Carlo replications for sample sizes $n = 20$, $n = 40$ and $n = 80$; the case $n = 100$ is not reported as the figures obtained as pretty much the same as the case $n = 80$. We have also simulated bigger sizes such as $n = 200$; in this case the first order approximation for the profiled empirical likelihood was already satisfactory (maximum size distortion obtained was 0.07 for the $\chi^2(4)$ case and the correction did improve it to 0.069).

We consider first the stochastic regressors case: the parameter vector β is specified as $[1, 0.8, -0.5, 1.5, 0]^T$ [4]so that we are testing the inclusion of irrelevant regressor as in (30). Table 1 below report the Monte Carlo results. For each specification of the innovation process, the first row reports the size of the profiled empirical likelihood ratio $W_{\widehat{\beta}_a}(\beta_{\alpha 0})$, the second one the size of the Bartlett corrected, while the third is based on the critical value of a $F(1, n-5)$ for normal innovations and $F(1, \infty)$ for the other two distributions.

TABLE 1 approximately here

[4]It should be noted that the choice of these values (and all the others in the reported simulation studies) for the regression parameters is completely arbitrary.

In the next simulation, we partition the vector β as $\beta_1 = [1, 0.8, -0.5]^T$ and $\beta_{20} = [1.5, 0.5]^T$ and the null hypothesis is $H_0 : \beta_2 = \beta_{20}$. Table 2 reports the results.

TABLE 2 approximately here

In Figure 1, we show the Q-Q plots for the six tests in the stochastic regressors case. It is evident that the correction is always effective. Interestingly, the Q–Q plots are relatively straight, indicating that the asymptotic approximations are rather good. It is also evident that there is a tendency of the sampling distribution of the (profiled) empirical likelihood ratio to generate outliers, as shown also in Figure 2 for the fixed regressors case.

FIGURE 1 approximately here

We next consider fixed regressors. The parameter vector β is partitioned as $\beta_1 = [1, -0.5, 2, 2.5]^T$ and $\beta_2 = 0$, so that we are testing the inclusion of an irrelevant regressor as in (30). Table 3 reports the results of the Monte Carlo experiment.

TABLE 3 approximately here

Finally we consider the following parameterisation $\beta_1 = [1, -0.3, 2]^T$ and $\beta_{20} = [0, 0]^T$ and test $H_0 : \beta_2 = \beta_{20}$ as in (29). Table 4 reports the results for such a test.

TABLE 4 approximately here

In Figure 2 we present the Q-Q plots for the fixed regressors case; their shape is similar to the stochastic regressors case, with the Bartlett correction noticeably improving the accuracy of the approximation.

FUGURE 2 approximately here

From this simple Monte Carlo analysis, it seems that the Bartlett corrected profiled empirical likelihood ratio test outperforms standard asymptotic (or exact) specification tests for subsets of the regressors' coefficients in a linear regression model and that the dimension of the nuisance parameters does not affect the performance of the tests. It is also clear that the correction is quite effective to reduce size distortion even in small sample, although there is a tendency to overcorrect the original statistic for a sample size of 20. The Q-Q plots reported in Figure 1 and 2 confirm this fact.

5 Conclusions

We have introduced a class of empirical likelihood based specification tests which can be used in linear regression models. By generalising a result of Chen (1994), we were able to obtain a Bartlett correction factor for a profiled version of the empirical likelihood ratio, by exploiting the linearity of the least square estimator for the regression parameters and the Frish-Waugh theorem. This latter result is in itself quite interesting because the introduction of nuisance parameters leads generally to the non Bartlett correctability of the resulting profiled empirical likelihood ratio (as opposed to standard parametric likelihood ratios). A small Monte Carlo study is used to assess the finite sample performance of the profiled empirical likelihood ratio as well as of the Bartlett corrected version of the test. Overall, the Bartlett corrected test seem to perform reasonably well as shown also by the Q-Q plots, though size distortions are still present. It is also worth pointing out that we can achieve the same level of accuracy by bootstrapping directly the Bartlett correction (details for more general situations can be found in Bravo (1999a)), but given the extreme simplicity of the correction itself, this alternative procedure seems unnecessary.

Despite these encouraging results, our approach cannot be easily extended to time series regressions. Although it is possible to obtain a Bartlett corrected empirical likelihood ratio test for the regression parameters of a time series regression by modifying a result of Kitamura (1997), this is not a straightforward extension for the situation analysed in this paper. In a weak dependent setup, we need to introduce some form of blocking the observations to take the serial correlation in account. The length of the blocks depends on an additional parameter $M = o\left(n^{1/2-1/2\delta}\right)$ for some $\delta > 0$ which slows the rate of consistency of the Lagrange multiplier λ to an order $O_p\left(Mn^{-1/2}\right)$, and the overall coverage error for confidence intervals to the order $O\left(n^{-5/6}\right)$. On the other side it is not difficult to show that the rate of consistency of the profiled parameter $\widehat{\beta}_a$ is still the standard $O_p\left(n^{-1/2}\right)$. This fact implies that we are working with two quantities characterised with different orders in probability and it is not clear how the overall order of probability of the asymptotic expansion for the profiled empirical likelihood ratio is affected. Moreover the arrays of observations described in (17) which were essential to express the asymptotic expansion for the profiled empirical likelihood ratio (16) as a function of the parameters of interest as given in the expansion (18) do not generalise to weak dependent observations. It seems therefore difficult to extend our analysis to the more general setting of time series regressions. Possibly in this case it is more useful to use some form of bootstrap to increase the accuracy of the test statistic.

A second limitation of the present approach is that we cannot test more general restrictions; as mentioned before one of the key feature of the linear regression model is that we can express the least squares estimator of the nuisance parameters as a function of the least square estimator for the parameter of interest. This fact allowed us to modify the original approach of DiCiccio et al. (1991) to Bartlett correcting empirical likelihood ratios and obtain the desired correction factor. In the case of general nonlinear restrictions this is clearly not the case therefore it seems plausible to deduce that it is not possible to obtain Bartlett corrections with the approach used in this paper.

Recently Bravo (2000) has obtained Bartlett type adjustments for the empirical goodness of fit class of tests, using arguments similar to those used by Cordeiro and Ferrari (1991) for correcting (up to an order $o\left(n^{-1}\right)$) Rao's score test. This fact opens the possibility of obtaining asymptotic refinements without relying on the asymptotic order of magnitude of the cumulants of the empirical likelihood ratio (or of quantities related to it) and hence might be possibly exploited in the context of testing nonlinear restrictions in linear regression models. This certainly deserve attention for future research.

References

Amemiya, T.: 1985, *Advanced Econometrics*, New York: Basil Blackwell.

Baggerly, K.: 1998, Empirical likelihood as a goodness of fit measure, *Biometrika* **85**, 535–547.

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.: 1977, Rifinements of the multidimensional central limit theorem and applications, *Annals of Probability* **1**, 1–27.

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

Bravo, F.: 1999a, Empirical likelihood based inference with applications to some econometric models, Submitted for publication.

Bravo, F.: 1999b, Higher order asymptotics for the empirical likelihood ratio J tests, Submitted for publication.

Bravo, F.: 2000, On Bartlett type adjustments for empirical discrepancy test statistics, Mimeo, University of York.

Chandra, T. and Ghosh, J.: 1979, Valid asymptotic expansions for the likelihood ratio and other perturbed chi-square variables, *Sankhyā A* **41**, 22–47.

Chen, S.: 1993, On the accuracy of empirical likelihood confidence regions for linear regression model, *Annals of the Institute of Statistical Mathematics,* **45**, 621–637.

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

Cordeiro, G. and Ferrari, S.: 1991, A modified score statistic having chi-squared distribution to order n^{-1}, *Biometrika* **78**, 573–582.

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.

Hall, P.: 1992, *The Bootstrap and Edgeworth Expansion*, Lecture Notes in Statistics, New York: Springer & Verlag.

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.

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.

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

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

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.

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

Appendix

Proof of Proposition 1

Recall that $\beta_a = \widehat{\beta}_a + \beta_a^1 + \beta_a^2 + \beta_a^3 + O_p\left(n^{-2}\right)$; generalising Chen (1994) approach, to obtain β_a^1 we solve the following quadratic

$$\max_{\beta_a^1} K_1^r K_1^r$$

where K_1^r is the set of averaged $k \times 1$ estimating equations, evaluated at the point $\widehat{\beta_a} + \beta_a^1$; simple differentiation shows that the required $(k-p) \times 1$ maximiser is given by:

$$\beta_a^1 = \left(x_i^a x_i^b\right)^{-1} x_j^b x_j^\alpha \left(\widehat{\beta_\alpha} - \beta_{\alpha 0}\right) \tag{31}$$

which is $O_p\left(n^{-1/2}\right)$ as required. Next, we determine β_a^2, which solves:

$$\max_{\beta_a^2} \left(K_2^r K_2^r - K_2^{rs} K_2^r K_2^s + 2\kappa^{rst} K_2^r K_2^s K_2^t/3\right),$$

with K_2^r evaluated this time at $\widehat{\beta_a} + \beta_a^1 + \beta_a^2$, and corresponds to maximising with respect to β_a^2 the following expression:

$$\zeta^{ab}\beta_a^2\beta_b^2 + 2\zeta^{sa}K_{\widehat{\beta_a}}^{rs}K_{\widehat{\beta_a}}^{r}\beta_a^1 - 2\gamma_{ra}\gamma_{sb}K_{\widehat{\beta_1}}^{rs}\beta_a^1\beta_b^2 - 2\kappa^{rst}\gamma_{sa}\gamma_{tb}K_{\widehat{\beta_a}}^{r}\beta_a^1\beta_b^2 +$$
$$2\kappa^{rst}\gamma_{ta}K_{\widehat{\beta_a}}^{r}K_{\widehat{\beta_a}}^{s}\beta_a^2 + 2\kappa^{rst}\gamma_{ra}\gamma_{sb}\gamma_{tc}\beta_a^1\beta_b^1\beta_c^2,$$

with $K_{\widehat{\beta_a}}^r$ as in (**??**) evaluated at $\widehat{\beta_a}$. Differentiating this last expression, we obtain

$$\beta_a^2 = -\zeta_{ab}\zeta^{rb}\left[K_{\widehat{\beta_a}}^{rs} - \kappa^{rst}\left(K_{\widehat{\beta_a}}^{s} - \zeta^{sc}\beta_c^1\right)\right]\left(K_{\widehat{\beta_a}}^{s} - \zeta^{sc}\beta_c^1\right), \tag{32}$$

where ζ_{ab} is the matrix inverse of ζ^{ab}. Using the same technique, it can be shown that $\beta_a^3 = O_p\left(1/n^2\right)$, whence (15) follows immediately. ■

Proof of Proposition 2

Suppose that X is a random vector having the same distribution of X_i $(i = 1, 2, ..., n)$, and suppose that $h^1, h^2, ...$ are real valued functions such that $E\left\{h^j(X)\right\} = 0$, $j =$

$1, 2, \ldots$. Let $H^j = h^j(X)$ and let $L^j = \sum_{i=1}^n h^j(X_i)/n$. Then

$$\begin{aligned} E\left(L^jL^k\right) &= E\left(H^jH^k\right)/n, \quad E\left(L^jL^kL^l\right) = E\left(H^jH^kH^l\right)/n^2, \qquad (33)\\ E\left(L^jL^kL^lL^m\right) &= [3](n-1)E\left(H^jH^k\right)E\left(H^lH^m\right)/n^3 + E\left(H^jH^kH^lH^m\right)/n^3,\\ E\left(L^jL^kL^lL^mL^n\right) &= [10]E\left(H^jH^k\right)E\left(H^lH^mH^n\right)/n^3 + O\left(n^{-4}\right),\\ E\left(L^jL^kL^lL^mL^nL^o\right) &= [15]E\left(H^jH^k\right)E\left(H^lH^m\right)E\left(H^nH^o\right)/n^3 + O\left(n^{-4}\right), \end{aligned}$$

where the symbol $[k]$ indicates the sum over k similar terms obtained by suitable permutation of indices.

Let

$$\begin{aligned} \Gamma_1 &= \left(s^{\alpha\varepsilon}\right)^{1/2}\left(s^{\beta\zeta}\right)^{1/2}\left(s^{\gamma\eta}\right)^{1/2}\left(s^{\delta\vartheta}\right)^{1/2} X_i^\varepsilon X_i^\zeta X_i^\eta X_i^\vartheta \rho_i^4, \quad \Gamma_2 = [3]\,\delta^{\alpha\beta}\delta^{\gamma\delta}, \qquad (34)\\ \Gamma_3 &= [4]\left(s^{\alpha\alpha'}\right)^{1/2}\left(s^{\beta\beta'}\right)^{1/2}\left(s^{\gamma\gamma'}\right)^{1/2}\left(s^{\delta\delta'}\right)^{1/2} s^{\varepsilon\zeta} X_i^{\alpha'} X_i^{\beta'} X_i^{\gamma'} X_j^{\delta'} X_j^\varepsilon X_j^\zeta \rho_i^3\rho_j^3,\\ \Gamma_4 &= [3]\left(s^{\alpha\alpha'}\right)^{1/2}\left(s^{\beta\beta'}\right)^{1/2}\left(s^{\gamma\gamma'}\right)^{1/2}\left(s^{\delta\delta'}\right)^{1/2} s^{\varepsilon\zeta} X_i^{\alpha'} X_i^{\beta'} X_i^\varepsilon X_j^{\gamma'} X_j^{\delta'} X_j^\zeta \rho_i^3\rho_j^3. \end{aligned}$$

Using (33) and (34), it follows that:

$$\begin{aligned} E\left(W_1^\alpha\right) &= 0, \quad E\left(W_2^\alpha\right) = -\left(s_{\beta\gamma}\right)^{1/2} s_{\gamma\delta} X_i^\alpha X_i^\beta X_i^\delta \rho_i^3/6n, \qquad (35)\\ E\left(W_3^\alpha\right) &= O\left(n^{-2}\right), \quad E\left(W_1^\alpha W_1^\beta\right) = \delta^{\alpha\beta}/n,\\ E\left(W_1^\alpha W_2^\beta\right) &= \delta^{\alpha\beta}/2n^2 - \left(s_{\alpha\varepsilon}\right)^{1/2}\left(s_{\beta\zeta}\right)^{1/2} s_{\gamma\delta} X_i^\varepsilon X_i^\gamma X_i^\delta X_i^\zeta \rho_i^4/2n^2 +\\ &\quad \left(s_{\alpha\varepsilon}\right)^{1/2}\left(s_{\beta\zeta}\right)^{1/2} s_{\gamma\delta}s_{\eta\iota} X_i^\varepsilon X_i^\gamma X_i^\delta X_j^\zeta X_j^\eta X_j^\iota \rho_i^3\rho_j^3/3n^2,\\ E\left(W_1^\alpha W_3^\beta\right) &= 5\left(s_{\alpha\delta}\right)^{1/2}\left(s_{\beta\varepsilon}\right)^{1/2} s_{\gamma\zeta} X_i^\delta X_i^\varepsilon X_i^\gamma X_i^\varepsilon \rho_i^4/8n^2 -\\ &\quad 29\left(s_{\alpha\varepsilon}\right)^{1/2}\left(s_{\beta\zeta}\right)^{1/2} s_{\gamma\delta}s_{\eta\iota} X_i^\varepsilon X_i^\gamma X_i^\delta X_j^\zeta X_j^\eta X_j^\iota \rho_i^3\rho_j^3/72n^2 -\\ &\quad \left(s_{\alpha\varepsilon}\right)^{1/2}\left(s_{\beta\zeta}\right)^{1/2} s_{\gamma\eta}s_{\delta\iota} X_i^\varepsilon X_i^\gamma X_i^\delta X_j^\zeta X_j^\eta X_j^\iota \rho_i^3\rho_j^3/72n^2,\\ E\left(W_2^\alpha W_2^\beta\right) &= \left(s_{\alpha\delta}\right)^{1/2}\left(s_{\beta\varepsilon}\right)^{1/2} s_{\gamma\zeta} X_i^\delta X_i^\varepsilon X_i^\gamma X_i^\varepsilon \rho_i^4/4n^2 -\\ &\quad 7\left(s_{\alpha\delta}\right)^{1/2}\left(s_{\beta\varepsilon}\right)^{1/2} s_{\gamma\zeta}s_{\eta\iota} X_i^\delta X_i^\gamma X_i^\zeta X_j^\varepsilon X_j^\eta X_j^\iota \rho_i^3\rho_j^3/36n^2 +\\ &\quad \left(s_{\alpha\delta}\right)^{1/2}\left(s_{\beta\varepsilon}\right)^{1/2} s_{\gamma\eta}s_{\zeta\iota} X_i^\delta X_i^\gamma X_i^\zeta X_j^\varepsilon X_j^\eta X_j^\iota \rho_i^3\rho_j^3/36n^2,\\ E\left(W_1^\alpha W_1^\beta W_1^\gamma\right) &= \left(s_{\alpha\delta}\right)^{1/2}\left(s_{\beta\varepsilon}\right)^{1/2}\left(s_{\gamma\zeta}\right)^{1/2} X_i^\delta X_i^\varepsilon X_i^\zeta \rho_i^3/n^2, \end{aligned}$$

$$
\begin{aligned}
E\left(W_1^{\alpha}W_1^{\beta}W_2^{\gamma}\right) &= -\left(s_{\alpha\delta}\right)^{1/2}\left(s_{\beta\varepsilon}\right)^{1/2}\left(s_{\gamma\zeta}\right)^{1/2}X_i^{\delta}X_i^{\varepsilon}X_i^{\zeta}\rho_i^3/3n^2 - \\
&\quad \left(s_{\beta'\gamma'}\right)^{1/2}s^{\gamma'\delta}X_i^{\alpha}X_i^{\beta'}X_i^{\delta}\rho_i^3\delta_{\beta\gamma}/6n^2, \\
E\left(W_1^{\alpha}W_1^{\beta}W_1^{\gamma}W_1^{\delta}\right) &= \left(\Gamma_1+(n-1)\Gamma_2\right)/n^3, \quad E\left(W_1^{\alpha}W_1^{\beta}\right)E\left(W_1^{\gamma}W_1^{\delta}\right)=\Gamma_2/n^2, \\
E\left(W_1^{\alpha}W_1^{\beta}W_1^{\gamma}W_2^{\delta}\right) &= \left(-6\Gamma_1+4\Gamma_2-\Gamma_3/6+2\Gamma_4/3+[12]\,E\left(W_1^{\alpha}W_1^{\beta}\right)E\left(W_1^{\gamma}W_2^{\delta}\right)\right)/n^3, \\
E\left(W_1^{\alpha}W_1^{\beta}W_1^{\gamma}W_3^{\delta}\right) &= \left(2\Gamma_1-19\Gamma_4/9\right)/n^3, \\
E\left(W_1^{\alpha}W_1^{\beta}W_2^{\gamma}W_2^{\delta}\right) &= \left(3\Gamma_1-\Gamma_2+\Gamma_3/6+5\Gamma_4/9+E\left(W_1^{\alpha}W_1^{\beta}\right)E\left(W_2^{\gamma}W_2^{\delta}\right)\right)/n^3,
\end{aligned}
$$

so that

$$
E\left(W_1^{\alpha}W_1^{\beta}W_2^{\gamma}\right)=E\left(W_1^{\alpha}W_1^{\beta}\right)E\left(W_2^{\gamma}\right)-E\left(W_1^{\alpha}W_1^{\beta}W_1^{\gamma}\right)/3+O\left(n^{-3}\right) \tag{36}
$$

$$
\begin{aligned}
E\left(W_1^{\alpha}W_1^{\beta}W_1^{\gamma}W_1^{\delta}\right)-[3]\,E\left(W_1^{\alpha}W_1^{\beta}\right)E\left(W_1^{\gamma}W_1^{\delta}\right) &= \left(\Gamma_1-\Gamma_2\right)/n^3, \qquad (37)\\
[4]\,E\left(W_1^{\alpha}W_1^{\beta}W_1^{\gamma}W_2^{\delta}\right)-[12]\,E\left(W_1^{\alpha}W_1^{\beta}\right)E\left(W_1^{\gamma}W_2^{\delta}\right) &= \left(-6\Gamma_1+2\Gamma_2-\Gamma_3/6+2\Gamma_4/3\right)/n^3, \\
[6]\,E\left(W_1^{\alpha}W_1^{\beta}W_2^{\gamma}W_2^{\delta}\right)-[6]\,E\left(W_1^{\alpha}W_1^{\beta}\right)E\left(W_2^{\gamma}W_2^{\delta}\right) &= \left(3\Gamma_1-\Gamma_2+\Gamma_3/6-5\Gamma_4/9\right)/n^3, \\
[4]\,E\left(W_1^{\alpha}W_1^{\beta}W_1^{\gamma}W_3^{\delta}\right)-[12]\,E\left(W_1^{\alpha}W_1^{\beta}\right)E\left(W_1^{\gamma}W_3^{\delta}\right) &= \left(2\Gamma_1-\Gamma_4/9\right)/n^3,
\end{aligned}
$$

and finally

$$
-\left([4]\,E\left(W_1^{\alpha}W_1^{\beta}W_1^{\gamma}\right)+[12]\,E\left(W_1^{\alpha}W_1^{\beta}W_2^{\gamma}\right)-2\,[6]\,E\left(W_1^{\alpha}W_1^{\beta}\right)E\left(W_2^{\gamma}\right)\right)E\left(W_2^{\delta}\right)=O\left(n^{-4}\right) \tag{38}
$$

Combining now (35), (36), (37) and (38) with the formulae for the first four cumulants (see e.g. McCullagh (1987, p. 31)) yields the following:

$$
\begin{aligned}
k^{\alpha} &= E\left(W^{\alpha}\right)=E\left(W_1^{\alpha}+W_2^{\alpha}\right)=-\left(s_{\beta\gamma}\right)^{1/2}s^{\gamma\delta}X_i^{\alpha}X_i^{\beta}X_i^{\delta}E\left(\varepsilon_i^3/\sigma^3\right)/6n+O\left(n^{-3}\right), \\
k^{\alpha,\beta} &= E\left(W^{\alpha}W^{\beta}\right)-E\left(W^{\alpha}\right)E\left(W^{\beta}\right) \\
&= E\left(W_1^{\alpha}W_1^{\beta}+2W_1^{\alpha}W_2^{\beta}+2W_1^{\alpha}W_3^{\beta}+W_2^{\alpha}W_2^{\beta}\right)-E\left(W^{\alpha}\right)E\left(W^{\beta}\right)+O\left(n^{-4}\right) \\
&= \delta^{\alpha\beta}/n+\left(\left(s_{\alpha\varepsilon}\right)^{1/2}\left(s_{\beta\zeta}\right)^{1/2}s_{\gamma\delta}X_i^{\varepsilon}X_i^{\zeta}X_i^{\gamma}X_i^{\delta}\rho_i^4/2+\right. \\
&\quad \left(s_{\alpha\varepsilon}\right)^{1/2}\left(s_{\beta\zeta}\right)^{1/2}s_{\gamma\eta}s_{\delta\iota}X_i^{\varepsilon}X_i^{\gamma}X_i^{\delta}X_j^{\zeta}X_j^{\eta}X_j^{\iota}\rho_i^3\rho_j^3/3n- \\
&\quad \left.\left(s_{\alpha\varepsilon}\right)^{1/2}\left(s_{\beta\zeta}\right)^{1/2}s_{\gamma\delta}s_{\eta\iota}X_i^{\varepsilon}X_i^{\gamma}X_i^{\delta}X_j^{\zeta}X_j^{\eta}X_j^{\iota}\rho_i^3\rho_j^3/36n\right)/n^2+O\left(n^{-4}\right),
\end{aligned}
$$

$$
\begin{aligned}
k^{\alpha,\beta,\gamma} &= E\left(W^\alpha W^\beta W^\gamma\right) - [3]\, E\left(W^\alpha W^\beta\right) E\left(W^\gamma\right) + E\left(W^\alpha\right) E\left(W^\beta\right) E\left(W^\gamma\right) \\
&= E\left(W_1^\alpha W_1^\beta W_1^\gamma\right) + [3]\, E\left(W_1^\alpha W_1^\beta W_2^\gamma\right) - [3]\, E\left(W_1^\alpha W_1^\beta\right) E\left(W_2^\gamma\right) + O\left(n^{-3}\right) \\
&= O\left(n^{-3}\right) \\
k^{\alpha,\beta,\gamma,\delta} &= E\left(W^\alpha W^\beta W^\gamma W^\delta\right) - [3]\, E\left(W^\alpha W^\beta\right) E\left(W^\gamma W^\delta\right) - [4]\, E\left(W^\alpha W^\beta W^\gamma\right) E\left(W^\delta\right) + \\
&\quad 2\,[6]\, E\left(W^\alpha\right) E\left(W^\beta\right) E\left(W^\gamma W^\delta\right) - 6E\left(W^\alpha\right) E\left(W^\beta\right) E\left(W^\gamma\right) E\left(W^\delta\right) \\
&= E\left(W_1^\alpha W_1^\beta W_1^\gamma W_1^\delta\right) + [4]\, E\left(W_1^\alpha W_1^\beta W_1^\gamma W_2^\delta\right) + [4]\, E\left(W_1^\alpha W_1^\beta W_1^\gamma W_3^\delta\right) + \\
&\quad [6]\, E\left(W_1^\alpha W_1^\beta W_1^\gamma W_1^\delta\right) - [3]\, E\left(W_1^\alpha W_1^\beta\right) E\left(W_1^\gamma W_1^\delta\right) - \\
&\quad [12]\, E\left(W_1^\alpha W_1^\beta\right) E\left(W_1^\gamma W_2^\delta\right) - [12]\, E\left(W_1^\alpha W_1^\beta\right) E\left(W_1^\gamma W_3^\delta\right) - \\
&\quad [6]\, E\left(W_1^\alpha W_1^\beta\right) E\left(W_2^\gamma W_2^\delta\right) + [4]\, E\left(W_1^\alpha W_1^\beta W_1^\gamma\right) E\left(W_2^\delta\right) - \\
&\quad [12]\, E\left(W_1^\alpha W_1^\beta W_2^\gamma\right) E\left(W_2^\delta\right) + 2\,[6]\, E\left(W_1^\alpha W_1^\beta\right) E\left(W_2^\gamma\right) E\left(W_2^\delta\right) + O\left(n^{-4}\right) \\
&= O\left(n^{-4}\right)
\end{aligned}
$$

The order of the higher order cumulants can then be deduced by using James and Mayne's (1962) general results. ■

Proof of Theorem 3

In order to show part (A), we first prove that there exists a valid Edgeworth expansion for the signed squared root of the empirical likelihood ratio orthogonality condition $EZ^r = E\left(X^T\varepsilon\right) = 0$ (i.e. for the whole parameter vector β). Part (b) then follows by a simple integration argument.

As in Chen (1994) we switch now to matrix notation; let U_j be the jth ($j = 1, 2, ..., k$) row of the matrix $U = U_{rs}$ defined in (9), and let U_{-r} define the $(k - r + 1)\times k$ matrix obtained after deleting top r ($r = 0, 1, ..., k-1$) rows from U. Define the following $j_1 \times k^2$ and $j_2 \times k^3$ (where $j_1 = k\,(k+1)/2$ and $j_2 = j_1 + k\,(k+1)\,(2k+1)/12$) matrices:

$$
\Upsilon_1 = \left[\begin{array}{cccc} U_1 \otimes U_{-0} & U_2 \otimes U_{-1} & ... & U_k \otimes U_{-k-1} \end{array}\right]^T,
$$

$$\Upsilon_2 = \Big[U_1 \otimes U_1 \otimes U_{-0}... \quad U_1 \otimes U_k \otimes U_{-k-1} \quad U_2 \otimes U_2 \otimes U_{-1}... \\ U_2 \otimes U_k \otimes U_{-k-1} \quad ... \quad U_k \otimes U_k \otimes U_{-k-1} \Big]^T,$$

where $\otimes$ denote the Kronecker product and let

$$\Xi_1 = X_i \otimes X_i \Upsilon_1^T \left(\varepsilon_i^2 - \sigma^2\right), \quad \Xi_2 = X_i \otimes X_i \otimes X_i \Upsilon_2^T \left(\varepsilon_i^3 - \mu_3\right)$$

where X_i is a $1 \times k$ vector.

Let $\overline{\omega}_n = n^{1/2}\overline{\omega}$, where $\overline{\omega} = \sum_{i=1}^n \omega_i/n$, and

$$\omega_i = n^{1/2} \left[\begin{array}{ccc} X_i^T U \varepsilon_i & \Xi_1 & \Xi_2 \end{array} \right]^T, \tag{39}$$

be the $(k + j_1 + j_2) \times 1$ vector containing all distinct first three multivariate central moments of UZ_i.

In order to prove the validity of the Edgeworth expansion for the random vector ω_i, we need some preliminary lemmae. Assume that

A6 $\underset{n\to\infty}{\underline{\lim}} \lambda_{in}/n > 0, \quad m_{in} = O\left(n^{\delta}\right)$ *for some* $\delta \in [0, 1/2)$ *for* $i = 2, 3$

where λ_{in} and m_{in} are the smallest eigenvalues and the $\max_i$ for the $n^2 \times p^2$ and $n^3 \times p^3$ matrices $X \otimes X$ and $X \otimes X \otimes X$, respectively.

Firstly, notice that under A2 the ω_i's are independent 0 mean random vectors.

Lemma 5 *Under A3 and A5 and A6 for $n \geq N$, then for $v \leq 5$,*

$$\overline{\lim_{n\to\infty}} (1/n) \sum_{i=1}^n E \left\|\omega_i\right\|^k < \infty. \tag{40}$$

Proof. First notice that A3 and A6 imply that there exist an n' large enough such that such that $n/\lambda_{in} \leq c_i$ $(i = 1, 2, 3)$ for some constants c_i. The result then follows by noting that

$$\begin{aligned} E\left\|\omega_i\right\|^v &\leq 3^v \left[E\left(\left\| n^{1/2} U X_i^T \right\|^v + E\left(\left\| n^{1/2} \Xi_1 \right\|^v \right) + E\left(\left\| n^{1/2} \Xi_2 \right\|^v \right) \right) \right] \\ &\leq 3^v \left[c_1^{v/2} \left\| X_i^T \right\|^v |\mu|^v + c_2^v \left\| X_i^T \right\|^{2v} |\mu|^{2v} + c_3^{3v/2} \left\| X_i^T \right\|^{3v} |\mu|^{3v} \right]. \end{aligned} \tag{41}$$

by the Loève and Chauchy-Schwarz inequality, where μ^v is the vth moment of ε_i.By A5 the right hand side of the second line of (41) is uniformly bounded, whence the result. ■

Let

$$\begin{aligned}
\Lambda_{12} &= U \sum_{i=1}^{n} X_i^T \left(X_i \otimes X_i\right) E\left(\varepsilon_i^3\right) \Upsilon_1^T / n, \\
\Lambda_{13} &= U \sum_{i=1}^{n} X_i^T \left(X_i \otimes X_i \otimes X_i\right) E\left(\varepsilon_i^4\right) \Upsilon_2^T / n, \\
\Lambda_{22} &= \Upsilon_1 \sum_{i=1}^{n} \left(X_i^T \otimes X_i^T\right) \left(X_i \otimes X_i\right) \left[E\left(\varepsilon_i^4\right) - E^2\left(\varepsilon_i^2\right)\right] \Upsilon_1^T / n, \\
\Lambda_{23} &= \Upsilon_1 \sum_{i=1}^{n} \left(X_i^T \otimes X_i^T\right) \left(X_i \otimes X_i \otimes X_i\right) E\left(\varepsilon_i^5\right) \Upsilon_2^T / n, \\
\Lambda_{33} &= \Upsilon_2 \sum_{i=1}^{n} \left(X_i^T \otimes X_i^T \otimes X_i^T\right) \left(X_i \otimes X_i \otimes X_i\right) \left[E\left(\varepsilon_i^6\right) - E^2\left(\varepsilon_i^3\right)\right] \Upsilon_2^T / n,
\end{aligned}$$

and let η_2 and η_3 denote the largest eigenvalues of Λ_{22} and Λ_{33}, respectively. Assume that

A7 $\eta_2 - \lambda_1^{-1} > 0$ and $\eta_3 - \lambda_1^{-1} - \left(\eta_2 - \lambda_1^{-1}\right)^{-1} > 0$, where λ_1 is the smallest eigenvalue of $X^T X$.

we can then prove the following lemma.

Lemma 6 *Under A7,* $\Sigma = (1/n) \sum_{i=1}^{n} COV\left(\omega_i\right)$ *is positive definite, and*

$$\mu_\omega = (1/n) \sum_{i=1}^{n} \left\|\Omega \omega_i\right\|^v < \infty, \tag{42}$$

for $0 \leq v \leq 5$, *where* $\Omega^2 = \Sigma^{-1}$.

Proof. The proof is based on Lemma 2.1 of Chen (1993). There exists a nonsingular matrix S such that

$$S^T \Sigma S = \begin{bmatrix} I_k & 0 & 0 \\ 0 & \Lambda_{22} - \Lambda_{12}^T \Lambda_{12} & 0 \\ 0 & 0 & \Lambda_{33} - \Lambda_{13}^T \Lambda_{13} - Q \end{bmatrix}$$

where $Q=\left(\Lambda_{23}^{T}-\Lambda_{13}^{T}\Lambda_{12}\right)\left(\Lambda_{22}-\Lambda_{12}^{T}\Lambda_{12}\right)^{-1}\left(\Lambda_{23}-\Lambda_{12}^{T}\Lambda_{13}\right)$. Noting that the smallest eigenvalue of $\Lambda_{22}-\Lambda_{12}^{T}\Lambda_{12}>\eta_2-\lambda_1^{-1}$ and the smallest eigenvalue of $\Lambda_{33}-\Lambda_{13}^{T}\Lambda_{13}-Q>\eta_3-\lambda_1^{-1}-\left(\eta_2-\lambda_1^{-1}\right)^{-1}$ are both positive by A7, it follows that the matrix Σ is positive definite. Let η_Σ denote the smallest eigenvalue of Σ. Clearly we have

$$\|\Omega\|=1/\eta_\Sigma^{1/2}\leq c_4=c_4\left(k,\mu_2,...,\mu_6\right),\quad \left\|\Omega^{-1}\right\|\leq c_5=c_5\left(k,\mu_2,...,\mu_6\right), \tag{43}$$

where c_4 and c_5 are finite constants depending on the first 6 moments of the ε_i's. Combining now (43) and (40) yields that $\mu_\omega=(1/n)\sum_{i=1}^{n}\|\Omega\omega_i\|^{v}\leq c_5=c_5\left(k,|\mu|^{3v}\right)$, where c_5 is a finite constant which depends on the $3v$th absolute moment of ε_i's which is finite by A5, whence (42). ■

Let $\widehat{F}_i$ be the Fourier transform of the distribution function F_i of ω_i and let $\widehat{P}_j\left(it,\{\chi_{5n}\}\right)$ be the Fourier transform of the function $P_j\left(-\phi_{0,\Sigma},\{\chi_{5n}\}\right)$ obtained by formally substituting $(-1)^{|q|}D^q\phi_{0,\Sigma}$ for $(it)^q$ $(i=\sqrt{-1})$ in the polynomial $\widehat{P}_j\left(it,\{\chi_{5n}\}\right)$, where $|q|=\sum_{j=1}^{k+j_2+j_3}q_i$, $D^q=D_1^{q_1}...D_{k+j_1+j_2}^{q_{k+j_1+j_2}}$ (D_j is the jth partial derivative operator) and $\phi_{0,\Sigma}$ is the normal density in $R^{k+j_2+j_3}$, and χ_{5n} are the first 5 cumulants of $\overline{\omega}$ (see (20)).

We can prove the following lemma.

Lemma 7 *Under A2-A7, there exist positive constants* $c_6=c_6\left(k,|\mu|^{3v},v\right)$ *and* $c_7=c_7\left(k,|\mu|^{3v},v\right)$ *such that for every* $t\in R^{k+j_2+j_3}$ *and some* $\delta\in[0,1/2)$ *satisfying:*

$$\|t\|<c_6n^{(1/2-\delta)}, \tag{44}$$

one has

$$\begin{aligned}&\left|D^q\prod_{i=1}^{n}\widehat{F}_i\left(\Sigma t/n^{1/2}\right)-\exp\left\{-\|t\|^2/2\right\}\left[1+\sum_{j=1}^{v-3}n^{-j/2}P_j\left(-\phi_{0,\Sigma},\{\chi_{5n}\}\right)\right]\right| \\ \leq\;& c_7n^{-(v-2)/2}\left[\|t\|^{2-|q|}+\|t\|^{3(v-2)+|q|}\right]\exp\left\{-\|t\|^2/4\right\},\end{aligned} \tag{45}$$

for all $q\in\left(Z_+\right)^{k+j_1+j_2}$.

Proof. Let $\widehat{G}$ be the Fourier transform of $\left[\begin{array}{ccc} \varepsilon_1 & \varepsilon_1^2-\sigma^2 & \varepsilon_1^3-\mu_3 \end{array}\right]^T$; by A4 $\widehat{G}$ is continuous at 0 and hence there exists a $\epsilon > 0$ such that $\left|\widehat{G}(z)-1\right| < 1/2$ for $\|z\| < \epsilon$.. Let

$$c_6 = \min\left\{c(v,k)\,\mu_\omega^{-1/(v-2)}, \epsilon/c_4\left(c_1^{3/2}+c_2^{3/2}+c_3^{3/2}\right)\right\} \tag{46}$$

where μ_ω is defined in (42), c_4 in (43) and $c(v,k)$ is a generic positive constant depending on k and v. Combining (44), (46) and (42) one gets

$$\|t\| \leq c(v,k)\, n^{1/2}\mu_\omega^{-1/(v-2)}. \tag{47}$$

Let $d_n = \sup\left\{a > 0 : t^T\Sigma t \leq a^2 \Rightarrow \left|\widehat{F}_i\left(\Omega t/n^{1/2}\right)-1\right| \leq 1/2\right\}$.

Suppose that $\Omega t = \left[\begin{array}{ccc} t_1 & t_2 & t_3 \end{array}\right]^T$ where $t_1 \in R^k$, $t_2 \in R^{j_1}$ and $t_3 \in R^{j_3}$. Then

$$\begin{aligned} \widehat{F}_i\left(\Omega t/n^{1/2}\right) &= \widehat{G}_i\left(\left[\begin{array}{ccc} t_1^T U X_i^T & t_2^T\Upsilon_1\left(X_i^T\otimes X_i^T\right) & t_3^T\Upsilon_2\left(X_i^T\otimes X_i^T\otimes X_i^T\right) \end{array}\right]^T/n^{1/2}\right) \\ &= \widehat{G}_i(\Delta), \quad \text{say.} \end{aligned} \tag{48}$$

Clearly under A6

$$\|\Delta\| \leq \sum_{j=1}^{3} m_j\|\Omega t\|/\lambda_j^{1/2} \leq c_4 c_6 n^{1/2-\delta}\left(c_1^{3/2}+c_2^{3/2}+c_3^{3/2}\right)n^\delta/n^{1/2} \leq \epsilon,$$

so that $\left|\widehat{F}_i\left(\Sigma t/n^{1/2}\right)-1\right| \leq 1/2$ for all $i = 1,...,n$. The result (45) follow directly from Theorem 9.9 of Bhattacharya and Rao (1976) as (46) and (47) implies condition (9.37) of that theorem. ■

Lemma 8 *Let* $\xi \in R^{k+j_2+j_3}$, $E(\xi,e) := E = \left\{\left|\xi^T\Omega\omega_i\right| > c_7\|\xi\|/n^{1/2}, i = 1,2,...,n\right\}$, *and let* $\#E$ *be the cardinality of* E. *Then for* $0 < c_7 < 1$, *one has*

$$\#E/n \geq \left(1-c_7^2\right)/\left[\left(nm_\Sigma^2/\eta_\Sigma\right)-c_7^2\right]. \tag{49}$$

Proof. Let $\#E^c = \left\{\left|\xi^T\Omega\omega_i\right| \leq c_7\|\xi\|/n^{1/2}, i = 1,2,...,n\right\}$. As $\sum_{i=1}^n\left(\xi^T\Omega\omega_i\right)^2 = \sum_{i=1}^n\left(\xi^T\Omega\omega_i\omega_i^T\Omega\xi\right) = \|\xi\|^2$, we can write

$$\|\xi\|^2 = \sum_{\#E}\left(\xi^T\Omega\omega_i\right)^2 + \sum_{\#E^c}\left(\xi^T\Omega\omega_i\right)^2 \leq \#E\|\xi\|^2 m_\Sigma^2/\eta_\Sigma + (n-\#E)\|\xi\|^2 c_7/n,$$

where the first term in the inequality follows, since $\eta_\Sigma = \min_{\|\iota\|=1} \iota^T \sum_{i=1}^n \omega_i \omega_i^T \iota / n \leq m_\Sigma^2$.
■

In the following lemma we show that the Fourier transform of the distribution of $\left[\begin{array}{ccc} \varepsilon_i & \varepsilon_i^2 - \sigma^2 & \varepsilon_i^3 - \mu_3 \end{array} \right]^T$ satisfies the Cramér condition.

Lemma 9 *Under A2-A7, let c_8 denote a positive constant, for all $i \in \#E$ one has*

$$\left| \widehat{F}_i(t) \right| \leq d$$

for some $0 < d < 1$.

Proof. As in Lemma 7, let G and $\widehat{G}$ denote the distribution and the Fourier transform of $\left[\begin{array}{ccc} \varepsilon_i & \varepsilon_i^2 - \sigma^2 & \varepsilon_i^3 - \mu_3 \end{array} \right]^T$. Using Lemma 1.4 of Bhattacharya (1977), we have

$$\overline{\lim_{\|t\| \to \infty}} \left| \widehat{G}(t) \right| < 1.$$

For $c_8 > 0$ and the same constant c_7 of Lemma 8, we have

$$\sup_{\|t\| > c_7 c_8 / 3} \left| \widehat{G}(t) \right| = d < 1. \tag{50}$$

Assume now that $t \in R^{k+j_2+j_3}$ satisfies $\|t\| > c_8$, then either $\|t_1\| > c_8/3$ or $\|t_2\| > c_8/3$ or $\|t_3\| > c_8/3$. Let $\widehat{F}_i(\xi) = \widehat{G}(\Delta)$ as in Lemma 6. Notice that for all $i \in \#E$, if $\|t_1\| > c_8/3$, we have

$$\left| t_1^T U X_i^T \right| > c_7 \|\Omega t_1\| > c_7 c_8 / 3,$$

and if $\|t_2\| > c_8/3$

$$\left| t_2^T \Upsilon_1 X_i^T \otimes X_i^T \right| > c_7 \|\Omega t_2\| > c_7 c_8 / 3$$

finally if $\|t_3\| > c_8/3$, then

$$\left| t_3^T \Upsilon_2 \left(X_i^T \otimes X_i^T \otimes X_i^T \right) \right| > c_7 \|\Omega t_3\| > c_7 c_8 / 3,$$

because $0 < c_7 < 1$. These inequalities imply altogether that

$$\left\| \left[\begin{array}{ccc} t_1^T U X_i^T & t_2^T \Upsilon_1 \left(X_i^T \otimes X_i^T \right) & t_3^T \Upsilon_2 \left(X_i^T \otimes X_i^T \otimes X_i^T \right) \end{array} \right]^T \right\| > c_7 c_8 / 3,$$

so by (50) we have

$$\left|\widehat{G}\left[\begin{array}{ccc} t_1^T U X_i^T & t_2^T \Upsilon_1\left(X_i^T\otimes X_i^T\right) & t_3^T \Upsilon_2\left(X_i^T\otimes X_i^T\otimes X_i^T\right)\end{array}\right]^T/n^{1/2}\right|\le d,$$

and so $\left|\widehat{F}_i\left(t\right)\right|\le d.$ ■

Noting that (43) with Lemmae 7 and 9 imply Theorem 20.1 of Bhattacharya and Rao (1976), it follows that

$$\sup_{B\in\mathcal{B}}\left|\Pr\left\{\overline{\omega}\in\mathcal{B}\right\}-\int_B\left[1+\sum_{j=1}^{2}n^{-j/2}P_j\left(-D,\chi_{5n}\right)\right]\phi_{k+j_1+j_2}\left(x\right)dx\right|=O\left(n^{-3/2}\right),\tag{51}$$

where $\mathcal{B}$ is any class of Borel subsets of $R^{k+j_1+j_2}$ such that $\sup_{B\in\mathcal{B}}\Phi_{k+j_1+j_2}\left(\partial B\right)^{\epsilon}=O\left(\epsilon\right)$, where $\Phi_{k+j_1+j_2}\left(\cdot\right)$ is $k+j_1+j_2$ dimensional multivariate normal distribution, $\left(\partial B\right)^{\epsilon}$ is the set in $R^{k+j_1+j_2}$ within ϵ from the boundary ∂ of B.

From (19), it is then clear that there exists a smooth function $f_n\left(\cdot\right)$ such that the signed square root of the profiled empirical likelihood ratio test $\widetilde{W}_{\beta_a^0}=f_n\left(\overline{\omega}\right)$ where

$$\overline{\omega}=\left[\begin{array}{ccc}\left(\widehat{\beta}_\alpha-\beta_{\alpha 0}\right) & K^{rs} & K^{rst}\end{array}\right]$$

and $\overline{\omega}=\left(1/n\right)\sum_{i=1}^{n}\omega_i$ where

$$\omega_i=\left[\begin{array}{ccc}\left(X_2^T M_{X_1}X_2\right)^{-1}X_2 M_{X_1}\varepsilon_i & \Xi_1 & \Xi_2\end{array}\right]^T$$

i.e. it has the same structure of the vector (39), the only difference is the dimension, being in this latter case $p+j_1+j_2$. We can then use Theorem 3.2 and Remarks 3.3. and 3.4 of Skovgaard (1981) to show that the expansion (51) may be transformed by a sufficiently smooth function $f_n\left(\cdot\right)$ to yield a valid Edgeworth expansion. This proves part (A).

The second part of Theorem 3 follows easily; switching back to the tensor notation, (51) is given by:

$$\sup_{B\in\mathcal{B}}\left|P\left\{n^{1/2}W^{\alpha}\in B\right\}-\int_B\mathcal{H}\left(v\right)\phi_p\left(v\right)dv\right|=O\left(n^{-3/2}\right)\tag{52}$$

where $\phi_p(v)$ is the p dimensional standard multivariate normal distribution and $\mathcal{H}(v)$ is the second order Edgeworth polynomial (see for example McCullagh (1987, Ch. 5))

$$\mathcal{H}(v) = 1 + k^{\alpha}h^{\alpha}/n^{1/2} + \left(k^{\alpha,\beta} + k^{\alpha}k^{\beta}\right)h^{\alpha\beta}/2n.$$

Using the symmetry of $\phi_p(v)$, the orthogonality property of the Hermite tensors $h^{\alpha_1 \ldots \alpha_k}$ and an integration by part argument yields that up to an error of order $O\left(n^{-3/2}\right)$:

$$\begin{aligned}
\Pr\{nW^{\alpha}W^{\alpha} \leq c_{\alpha}\} &= \int_{\|v\| \leq c_{\alpha}^{1/2}} \mathcal{H}(v)\,\phi_p(v)\,dv \\
&= \Pr\left\{\chi^2(p) \leq c_{\alpha}\right\} + \delta^{\alpha\beta}\left(k^{\alpha,\beta} + k^{\alpha}k^{\beta}\right)\left[\Pr\left\{\chi^2(p+2) \leq c_{\alpha}\right\} - \right. \\
&\quad \left.\left(2^{1/2}\Gamma(1/2)\right)^{-1}\sum_{\alpha=1}^{p} v^{\alpha}\left(\exp\left(-v^{\alpha}v^{\beta}\delta^{\alpha\beta}/2\right)|_0^{c_{\alpha}^{1/2}} + v^{\alpha}d\phi(v)\right)\right].
\end{aligned}$$

Integrating the last term of the second line yields (21). ■

Proof of Corollary 4

The results follows immediately by noting that

$$\Pr\left\{\chi^2(p) \leq c_{\alpha}(1 + B/n)\right\} = \alpha + Bc_{\alpha}g_p(c_{\alpha})/n + O\left(n^{-2}\right), \tag{53}$$

so combining this latter expression with (21) we get

$$\Pr\left\{\widetilde{W}_{\beta_{\alpha}^{0}}/E\left(\widetilde{W}_{\beta_{\alpha}^{0}}\right) \leq c_{\alpha}\right\} = \alpha - Bc_{\alpha}g_p(c_{\alpha})/n + Bc_{\alpha}g_p(c_{\alpha})/n + O\left(n^{-3/2}\right);$$

as the error term $O\left(n^{-3/2}\right)$ is actually $O\left(n^{-2}\right)$ by the even-odd property of the Hermite tensors (see Barndorff-Nielsen and Hall (1988)), the result follows. ■

Tables and figures

TABLE 1. EMPIRICAL SIZE OF $\widetilde{W}_{\beta^0_\alpha}$

	$n=20$		$n=40$		$n=80$	
$N(0,1)$	0.183^a	0.118^b	0.134^a	0.101^b	0.121^a	0.071^b
	0.147^a	0.083^b	0.119^a	0.058^b	0.113^a	0.054^b
	0.151^a	0.081^b	0.117^a	0.056^b	0.113^a	0.053^b
$\chi^2(4)$	0.214^a	0.159^b	0.149^a	0.101^b	0.139^a	0.095^b
	0.163^a	0.108^b	0.138^a	0.092^b	0.115^a	0.081^b
	0.231^a	0.163^b	0.141^a	0.104^b	0.132^a	0.095^b
$t(5)$	0.202^a	0.131^b	0.131^a	0.096^b	0.128^a	0.083^b
	0.152^a	0.094^b	0.119^a	0.078^b	0.119^a	0.071^b
	0.194^a	0.142^b	0.142^a	0.082^b	0.135^a	0.085^b

a,b 10% and 5% nominal size, respectively.

TABLE 2. EMPIRICAL SIZE OF $\widetilde{W}_{\beta^0_\alpha}$

	$n=20$		$n=40$		$n=80$	
$N(0,1)$	0.199^a	0.131^b	0.141^a	0.093^b	0.122^a	0.069^b
	0.151^a	0.095^b	0.123^a	0.071^b	0.118^a	0.063^b
	0.181^a	0.094^b	0.125^a	0.069^b	0.117^a	0.059^b
$\chi^2(4)$	0.231^a	0.152^b	0.172^a	0.107^b	0.150^a	0.102^b
	0.172^a	0.098^b	0.140^a	0.091^b	0.127^a	0.093^b
	0.229^a	0.128^b	0.179^a	0.109^b	0.139^a	0.103^b
$t(5)$	0.214^a	0.145^b	0.159^a	0.101^b	0.134^a	0.092^b
	0.162^a	0.079^b	0.134^a	0.074^b	0.120^a	0.069^b
	0.202^a	0.113^b	0.160^a	0.089^b	0.134^a	0.094^b

a,b 10% and 5% nominal size, respectively.

Table 3. Empirical size of $\widetilde{W}_{\beta_{\alpha}^{0}}$

	$n=20$		$n=40$		$n=80$	
$N(0,1)$	0.192^a	0.129^b	0.135^a	0.086^b	0.128^a	0.079^b
	0.140^a	0.080^b	0.130^a	0.072^b	0.119^a	0.065^b
	0.141^a	0.079^b	0.133^a	0.071^b	0.121^a	0.065^b
$\chi^2(4)$	0.243^a	0.154^b	0.170^a	0.132^b	0.151^a	0.098^b
	0.154^a	0.119^b	0.134^a	0.099^b	0.126^a	0.085^b
	0.229^a	0.135^b	0.176^a	0.119^b	0.145^a	0.106^b
$t(5)$	0.207^a	0.135^b	0.167^a	0.102^b	0.138^a	0.088^b
	0.149^a	0.105^b	0.129^a	0.078^b	0.125^a	0.073^b
	0.195^a	0.142^b	0.139^a	0.099^b	0.135^a	0.091^b

a,b 10% and 5% nominal size, respectively.

Table 4. Empirical size of $\widetilde{W}_{\beta_{\alpha}^{0}}$

	$n=20$		$n=40$		$n=80$	
$N(0,1)$	0.208^a	0.135^b	0.151^a	0.099^b	0.130^a	0.082^b
	0.146^a	0.090^b	0.135^a	0.073^b	0.124^a	0.067^b
	0.143^a	0.089^b	0.138^a	0.083^b	0.123^a	0.064^b
$\chi^2(4)$	0.239^a	0.161^b	0.181^a	0.139^b	0.139^a	0.100^b
	0.164^a	0.125^b	0.157^a	0.104^b	0.130^a	0.090^b
	0.240^a	0.169^b	0.189^a	0.147^b	0.141^a	0.108^b
$t(5)$	0.221^a	0.136^b	0.159^a	0.121^b	0.134^a	0.087^b
	0.157^a	0.104^b	0.140^a	0.096^b	0.121^a	0.072^b
	0.209^a	0.129^b	0.168^a	0.126^b	0.133^a	0.096^b

a,b 10% and 5% nominal size, respectively.

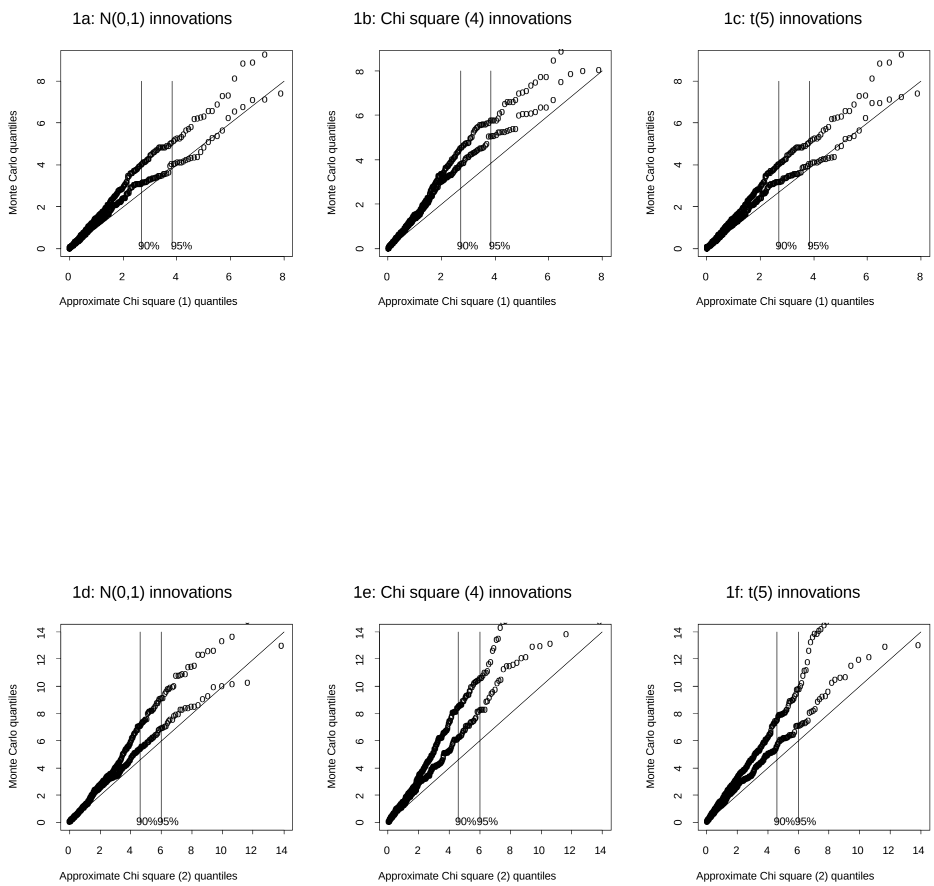

Figure 1: Q-Q plots for the stochastic regressors case. Vertical lines indicate the 90% and 95% critical values, respectively.

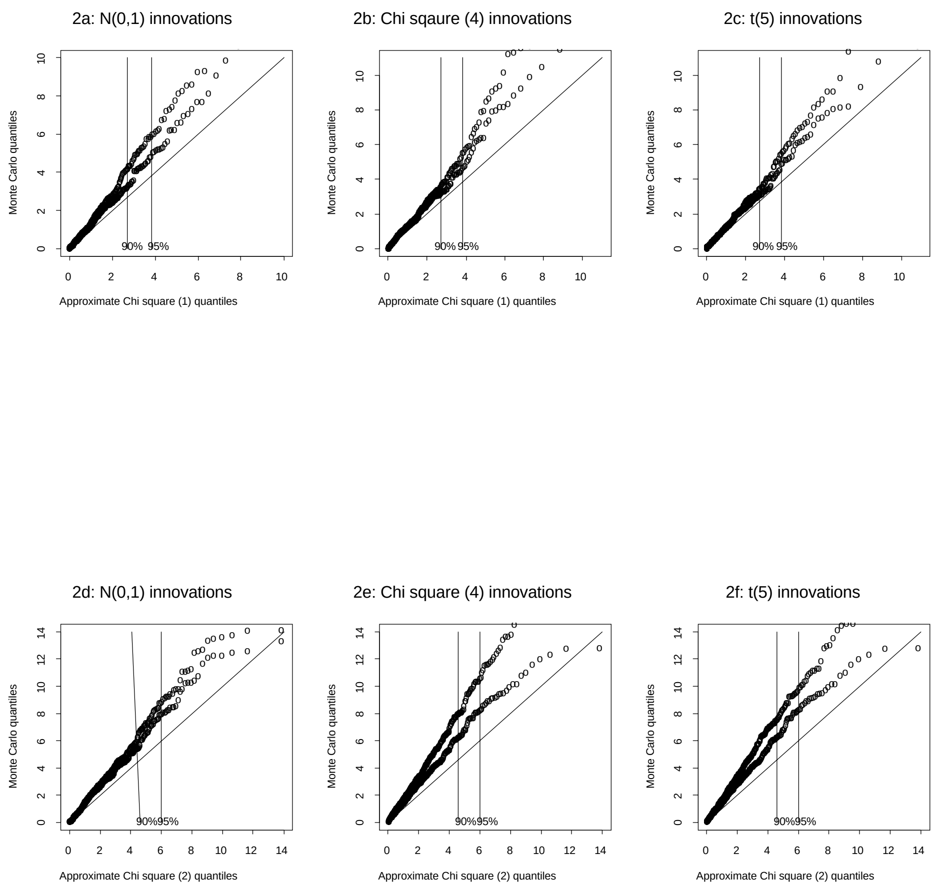

Figure 2: Q-Q plots for the fixed regressors case. Vertical lines indicate 90% and 95% critical values, respectively.