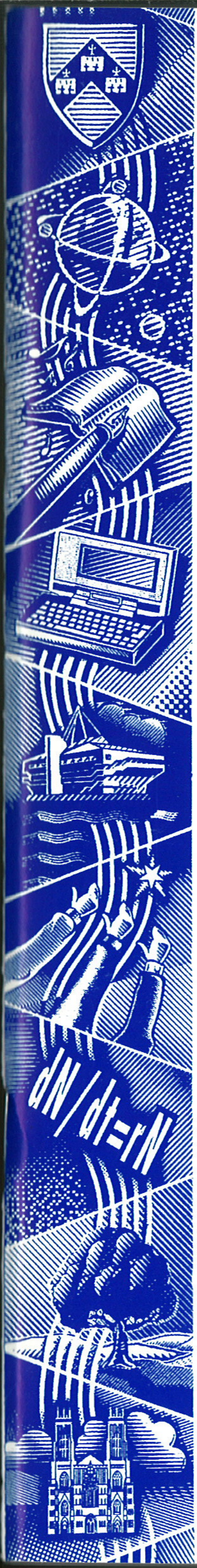

THE UNIVERSITY *of York*

Discussion Papers in Economics

No. 2000/30

Higher Order Asymptotics and the Bootstrap for Empirical Likelihood J Tests

by

Francesco Bravo

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

The purpose of this series is to circulate current research in economics by members of the Department of Economics and Related Studies at the University of York. All enquiries and contributions should be sent to the Editor, Discussion Paper Series, Department of Economics and Related Studies, University of York, Heslington, York, YO1 5DD. Telephone: +44 (1904) 433776.

Discussion paper abstracts and general information on the Department can be accessed via our Internet address www.york.ac.uk/depts/econ/

A list of the most recent discussion papers in the series can be found on the inside back cover and at www.york.ac.uk/depts/econ/dp

---

Related Discussion Paper and Research Paper series at the University of York include the following:

- CENTRE FOR DEFENCE ECONOMICS Research Monographs
  (Contact: Margaret Cafferky, Centre for Defence Economics)

- CENTRE FOR EXPERIMENTAL ECONOMICS Discussion Papers
  (Contact: Professor John Hey, EXEC)

- CENTRE FOR HEALTH ECONOMICS Discussion Paper Series
  (Contact: The Publications Secretary, CHE)

- CENTRE FOR HOUSING POLICY Discussion Paper Series
  (Contact: Discussion Paper Series Editor, Centre for Housing Policy)

- ENVIRONMENTAL ECONOMICS AND ENVIRONMENTAL MANAGEMENT Discussion Paper Series
  (Contact: Dr. Jack Pezzey, EEEM)

- SPRU Papers
  (Contact: The Publications Office, Social Policy Research Unit)

- YORK RESEARCH PARTNERSHIP Discussion Paper Series
  (Contact: Dr. Bernard Stafford, York Research Partnership)

- YORK HEALTH ECONOMICS CONSORTIUM Occasional Papers
  (Contact: Julie Holmes, YHEC)

Higher order asymptotics and the bootstrap for empirical likelihood J tests*

BY FRANCESCO BRAVO

Abstract

In this paper we obtain a second order Edgeworth approximation to the density of a likelihood ratio type J test for overidentifying restrictions by embedding the moment conditions into the empirical likelihood framework. The resulting asymptotic expansion can be used to correct to an order $o\left(n^{-1}\right)$ the critical values of the empirical likelihood ratio J test and to justify the second order correctness of an "hybrid" bootstrap procedure which we propose to bypass the difficult calculation of the cumulants appearing in the Edgeworth density of the empirical likelihood ratio J test. The resulting bootstrap calibrated empirical likelihood ratio test seems to perform well, as shown in a small Monte Carlo study, and suggest that the combination of the empirical likelihood method together with a suitable bootstrap procedure is an extremely useful method for estimation/inference in moment based econometric models.

*This paper is based on Chapter 4 of my Ph. D. dissertation at the University of Southampton. I would like to thank Grant Hillier for valuable conversations. Many thanks also to Andrew Chesher, Peter Phillips, Jan Podivinsky and partecipants at ESEM99, Santiago de Compostela, Spain for useful comments. Partial financial support under ESRC grant R00429634019 is gratefully acknowledged. A code in S is available upon request. Any remaining errors are my own responsibility.

1 Introduction

THERE HAS BEEN GROWING INTEREST in the so called information theoretic approach to inference for semiparametric econometric models. This approach embeds constraints (such as moment conditions) assumed to contain all the information in the data in a fully nonparametric (i.e. distribution free) set-up, by adding some additional parameters associated with these constraints. The resulting estimators and test statistics have (to first order) sampling properties similar to their bootstrap analogues, but whereas bootstrap uses resampling, they are all based on reweighting the data. These weights are the solution of a constrained optimisation problem, which takes different forms according to the criterion used to measure the closeness of the empirical to the estimated distribution supported on the data, and therefore can be interpreted as constrained (or implied, using Back and Brown's (1993) terminology) probabilities. Interestingly, all these information theoretic methods are based, at least asymptotically, on Stein's (1956) concept of least favourable families: for each fixed value of the parameter of interest, by maximising a multinomial distribution assumed to have probability atoms at the data points , one reduces the original nonparametric problem to a least favourable parametric subfamily[1]. The resulting subfamily defines implied probabilities as implicit functions of the Lagrange multiplier arising from the constrained optimisation. A dual likelihood argument (Mykland, 1995) can then be advocated to show that inference for the original parameter vector can be based on the vector of Lagrange multipliers -the dual parameter in dual likelihood's terminology- associated with the original constrained optimisation.

In this paper we focus on the empirical likelihood method originally developed by Owen (1988), which can be thought of as resulting from minimising the forward Kullback-Liebler (i.e. the likelihood) distance between the empirical measure and a constrained probability measure. Our main interest is to investigate the higher order asymptotic behaviour of an empirical likelihood ratio test for overidentifying restrictions (often called the J test) in moments-based econometric models.

We make several contributions: firstly, we derive a stochastic expansion for the empirical likelihood ratio J (ELJ henceforth) test. The resulting expansion extends to higher order asymptotics some of the results obtained by Qin and Lawless (1994). Using Hayakawa's (1977) technique, this latter expansion is then "transformed" into a valid (in Sargan's (1980) sense) Edgeworth expansion for the density of ELJ. The resulting Edgeworth density can be used to obtain asymptotic refinements to the critical values of the test and it is of its own interest as it highlights how the introduction of nuisance parameters (which in our set-up are represented by the overidentifying restrictions themselves) complicates dramatically the higher order asymptotic analysis for empirical likelihood based tests.

Secondly, we introduce a "hybrid" bootstrap procedure which bypasses the notoriously difficult problem of calculation the empirical cumulants appearing in the Edgeworth density of ELJ. This bootstrap approach is based on Owen's (1988) original idea of using the bootstrap to calibrate empirical likelihood ratios. It can be considered as a variant of the (intentionally) biased class of bootstrap procedures recently introduced by Hall and Presnell (1999). In the biased bootstrap approach, the observations are resampled according to some weights (probabilities) chosen so that they satisfy a set of constraints. Indeed, the empirical likelihood method, and more generally other empirical discrepancy based methods (Corcoran, 1998) are examples of biased bootstrap procedures. In our approach, the observations are still resampled uniformly with replacement (as in standard bootstrap theory) but the resulting resample is used in the bootstrap estimation procedure only if it satisfies the fundamental condition for existence and positiveness of empirical likelihood ratios as given in assumption A0 in Section 2 below.

Using fairly standard arguments, we show that the resulting bootstrap calibrated empirical likelihood ratio J test provides higher order asymptotic refinements up to the $o(n^{-1})$ order and seems to work extremely well in finite samples (at least in the small Monte Carlo experiment carried out in Section 4).

The paper is organised as follows: in Section 2, after briefly reviewing the empirical likelihood theory, we introduce the notion of quasi-dual likelihood, which is a simple extension of the dual likelihood concept and it is necessary to deal with the fact that the ELJ test is not only a function of the dual parameter but also of the underlying parameters of interest. Since the estimation of these latter parameters can be seen as a saddlepoint type of problem, we introduce a saddlepoint estimator which shares the same first order asymptotic properties of the efficient Generalised Method of Moment (GMM) estimator originally developed by Hansen (1982) and can be used to obtain a simple asymptotically χ^2 test for overidentifying restrictions. In Section 3 we first develop a stochastic expansion for the ELJ test, and then obtain a valid Edgeworth approximation for its density. In Section 4 we introduce the bootstrap calibration approach, discuss some related resampling procedure and justify its second order correctness. We also present a small Monte Carlo experiment which is carried out to assess the finite sample properties of the bootstrap calibrated ELJ . Finally, Section 5 contains some concluding remarks. All the proofs and derivations of the asymptotic expansions presented in the paper are reported in Appendices A-E. Unless otherwise stated the sum $\sum$ is always intended as $\sum_{i=1}^{n}$.

2 Empirical likelihood J test and quasi-dual likelihood

Consider the following moment condition model, as in Hansen (1982)

$$E\psi\left(z_i,\theta_0\right)=0 \tag{1}$$

where $\{z_i\}_{i=1}^{n}$ is a sequence of independent $m\times1$ random vectors distributed according an unknown distribution F_0, θ_0 is a $p\times1$ vector of unknown parameters and $\psi\left(z_i,\theta_0\right)$

is a known Borel measurable function with r components ($r > p$). Notice that we are assuming that the dimension of the vector of estimating equations $\psi(z_i, \theta_0)$ is strictly bigger than the dimension of the parameter vector θ. The case $r = p$ (i.e. just identified econometric models) is fully analysed in Bravo (1999). Let ^{T} denote transpose.

Following Stein's (1956) seminal idea (see also Chamberlain (1987)) the unknown distribution F_0 can be (arbitrarily well) approximated by a multinomial distribution, with probability atoms $p_i = dF(z_i) = \Pr(Z = z_i)$. By profiling the empirical likelihood function $\prod_{i=1}^{n} p_i$ over the simplex $\sum p_i = 1$ and the empirical counterpart of (1) (i.e. $\sum p_i \psi(z_i, \theta_0) = 0$), we have by a Lagrange multiplier argument that the resulting implied probabilities are given by:

$$p_i = \left(1 + \lambda(\theta_0)^T \psi(z_i, \theta_0)\right)^{-1} / n, \tag{2}$$

with the $r \times 1$ vector of Lagrange multipliers $\lambda(\theta_0)$ depending implicitly on the original parameter θ_0, because of the restriction $\sum p_i \psi(z_i, \theta_0) = 1$. As shown in Appendix A, when the overidentifying restrictions (2.1) hold, the probability limit of the estimator $\widetilde{\lambda}(\theta_0)$ of $\lambda(\theta_0)$ is 0, implying that the implied probabilities (2) correspond to the empirical measure $\widehat{p}_i = n^{-1}$, i.e. the nonparametric (unrestricted) maximum likelihood estimate of F_0. Therefore, we can think of testing the validity of the overidentified model (1), as testing for the parametric restriction $H_0 : \lambda(\theta_0) = 0$. As in a fully parametric context, we can then define a classical likelihood ratio test, which amounts to comparing the unrestricted probabilities $\widehat{p}_i = n^{-1}$ with the constrained probabilities p_i given in (2). Simple algebra shows that (twice) the resulting (convex dual) empirical likelihood ratio function is:

$$W^-(\theta, \lambda(\theta)) = -\sum \log\left(1 + \lambda^T(\theta)\psi(z_i, \theta)\right), \tag{3}$$

which depends on the $r \times 1$ vector of Lagrange multipliers and on the $p \times 1$ vector of unknown parameters θ. We can interpret the criterion function (3) as a quasi-dual likelihood, in the sense that we are still associating the vector of Lagrange multipliers $\lambda(\theta)$ with the sample version of the moment conditions (2.1), as implicitly assumed by the "standard" Mykland's (1995) dual likelihood based test $H_0 : \lambda(\theta_0) = 0$ (see Bravo (1999) for more details), but there is an essential difference when working with overidentified models such as (1). To see this latter point, let

$$\widetilde{\lambda}(\theta_0) = \arg\min_{\lambda} W^{-}(\theta_0, \lambda(\theta_0)) \tag{4}$$

be the unique (given the convex dual formulation of the criterion function (3) in $\lambda(\theta)$) minimiser of the quasi-dual likelihood (3). It is easy to see that when the moment model (2.1) is correct,

$$E_{F_0} \exp\left\{-W^{-}\left(\theta_0, \widetilde{\lambda}(\theta_0)\right)\right\} \to 1$$

(because $\widetilde{\lambda}(\theta_0) \xrightarrow{p} 0$ under the null hypothesis, see Proposition 2.1 below), yet, because the moment restrictions are not satisfied in the sample for overidentified models, we need also an estimate $\widetilde{\theta}$ of the parameter θ. Such an estimator can be obtained by maximising the profiled quasi-dual likelihood $W^{-}\left(\theta, \widetilde{\lambda}(\theta)\right)$ with respect to θ, which implies that we need to solve the following (well-posed, given the convex dual formulation in (3)) saddlepoint problem:

$$\widetilde{\theta} = \arg\max_{\theta\in\Theta}\min_{\lambda} W^{-}(\theta, \lambda(\theta)). \tag{5}$$

This latter equation defines what Qin and Lawless (1994) call the maximum empirical likelihood estimator for θ. As the two estimators $\widetilde{\lambda}(\theta)$ and $\widetilde{\theta}$ (solution of (5)) are consistent for 0 and θ_0 (i.e. the optimal value for $\lambda(\theta)$ and the true value of θ), this implies that the expectation of the quasi-dual likelihood function $W^{-}(\theta_0, \lambda(\theta_0))$ is characterised by a unique saddlepoint at $\lambda(\theta) = 0$ and $\theta = \theta_0$. This justifies asymptotically the quasi-dual likelihood approach in overidentified models.

This notion of quasi-dual likelihood is a straightforward extension of Mykland's (1995) dual likelihood which is used in Bravo (1999) to characterise the higher order asymptotic behaviour of empirical likelihood (i.e. dual likelihood) ratios for exactly identified econometric models, by making heavy use of Bartlett type identities as described in Mykland (1994). These identities characterise uniquely a dual likelihood as an artificial likelihood but do not hold in full generality in the present context, because of the presence of the maximum empirical likelihood estimator. Hence, since the criterion function (3) does not satisfy them, we prefer to use the term quasi-dual likelihood (in analogy with White's (1982) quasi-maximum likelihood approach, where the standard Bartlett identities do not hold in misspecified models).

Let $\psi^{\alpha}(z_i,\theta)$ denote the αth component of the $r\times 1$ vector $\psi(z_i,\theta)$ ($\alpha = 1,2,...,r > p$), $\|\cdot\|$ be the Euclidean (or matrix) norm, and $\Gamma(\theta_0,\tau)$ be an open sphere with center θ_0 and radius τ, *int* $\{S\}$ be the interior of a set $S \subseteq R^p$, and *ch* $\{S\}$ the convex hull for the set $S \subseteq R^p$.

Assume that with probability 1 (w.p.1 henceforth):

A0 $0 \in ch\{\psi(z_1,\theta), \psi(z_2,\theta), ..., \psi(z_n,\theta)\}$ as $n \to \infty$,

A1 *The parameter space* $\Theta \subseteq R^p$ *is compact,* $E\psi(z_i,\theta) = 0$ *for a unique* $\theta_0 \in$ *int* $\{\Theta\}$, *and the distribution of* $\psi(z_1,\theta_0)$ *is non-lattice,*

A2 *For sufficiently small* $\tau > 0$, *and* $\eta > 0$, $\sup_{\theta\in\Gamma(\theta,\tau)} \|\psi(z_i,\theta)\|^{2(1+\eta)}$ *is bounded by some integrable function* $\Psi_1(z_i)$ $\forall\theta$,

A3 $\psi^{\alpha}(z_i,\theta)$ *is continuous in* θ *for almost every* z_i,

A4 *(i)* $E\psi(z_i,\theta_0)\psi(z_i,\theta_0)^T$ *is positive definite, (ii)* $E\|\psi(z_i,\theta_0)\|^{2\delta} < \infty$, *for some* $\delta \geq 1$,

A5 $\partial\psi(z_i,\theta)/\partial\theta$ *and* $\partial^2\psi^{\alpha}(z_i,\theta)/\partial\theta\partial\theta^T$ *are continuous at* θ_0,

A6 $E\partial\psi(z_i,\theta_0)/\partial\theta^T$ *is of full column rank,* A7 *For sufficiently small* $\tau > 0$, $\sup_{\theta^* \in \Gamma(\theta_0, \tau)} \|\partial \psi(z_i, \theta^*)/\partial \theta\|$, $\sup_{\theta^* \in \Gamma(\theta_0, \tau)} \left\|\partial^2 \psi^\alpha(z_i, \theta^*)/\partial \theta \partial \theta^T\right\|$ *are bounded by some integrable functions* $\Psi_2(z_i)$, $\Psi_3(z_i)$ *respectively.*

REMARK I Assumption A0 is standard (although crucial) in empirical likelihood theory, as it implies that $\Pr\left\{\theta \left| 1 + \lambda(\theta)^T \psi(z_i, \theta) \geq 1/n \right.\right\} \to 1$ as $n \to \infty$ for some fixed θ (i.e. the empirical likelihood ratio exists and it is positive); Assumptions A1- A7 are standard in GMM literature: in particular, A1-A4(i) are sufficient to establish the consistency of the estimators, $\widetilde{\lambda}\left(\widetilde{\theta}\right)$ and $\widetilde{\theta}$ say, of $\lambda(\theta_0)$ and θ respectively, while the remaining assumptions are used to obtain their joint asymptotic normality (see (7) below). Assumption A4 (ii) is used to characterise the rate of convergence of the empirical likelihood ratio test (see Theorem 4 in Appendix A); $\delta = 1$ is enough to obtain weak convergence of the test statistic under the null hypothesis; to obtain the rate $O\left(n^{-1/2}\right)$, we need to strengthen A4(ii) to $\delta = 2$.

For notational convenience, we drop the functional dependence of the Lagrange multiplier $\lambda(\theta)$ on the underlying parameter characterising the expectation of model (1); hence let $\lambda := \lambda(\theta)$; also let $W^- := W^-(\theta, \lambda(\theta))$ and ∂_τ denote the partial derivative operator with respect to the vector parameter τ.

The saddlepoint problem (5) suggests that a natural estimator for λ and θ is based on the following $(r+p) \times 1$ first order conditions:

$$
(6) \qquad \begin{aligned}
\partial_\lambda W^- &= -n^{-1} \sum \left(1 + \widetilde{\lambda}^T \psi\left(z_i, \widetilde{\theta}\right)\right)^{-1} \psi\left(z_i, \widetilde{\theta}\right) = 0, \\
\partial_\theta W^- &= -n^{-1} \sum \left(1 + \widetilde{\lambda}^T \psi\left(z_i, \widetilde{\theta}\right)\right)^{-1} \left(\partial \psi\left(z_i, \widetilde{\theta}\right)/\partial \theta\right)^T \widetilde{\lambda} = 0
\end{aligned}
$$

which can be solved by the multivariate Newton's algorithm, and whose solution gives the saddlepoint estimate associated with the saddlepoint estimator $\widetilde{\zeta} = \left[\begin{array}{cc} \widetilde{\lambda} & \widetilde{\theta} \end{array}\right]^T$ for the parameters λ and θ. Notice also that from a computational point of view, the original overidentified moment model becomes simply a just identified moment model which is easier to estimate by any numerical optimisation routines.

In Appendix A we prove consistency and asymptotic normality for $n^{1/2}\widetilde{\zeta}$. In particular we show that

$$(7) \qquad n^{1/2}\begin{bmatrix} \widetilde{\lambda} \\ \widetilde{\theta}-\theta_0 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma^{-1}-\Sigma^{-1}DV^{-1}D^T\Sigma^{-1} & O \\ O & V^{-1} \end{bmatrix}\right),$$

where O is a matrix of 0 of appropriate dimension, $D = E\partial\psi(z_i,\theta)/\partial\theta^T$, $\Sigma = E\psi(z_i,\theta)\psi(z_i,\theta)^T$, and $V = D^T\Sigma^{-1}D$.

From (7), it is evident the first order equivalence between the maximum empirical likelihood estimator, $n^{1/2}\left(\widetilde{\theta}-\theta_0\right)$, and the (efficient) GMM estimator (i.e. they have the same asymptotic covariance matrix V^{-1}). Notice also that the two estimators are asymptotically independent (Qin and Lawless, 1994). This latter property will be exploited to derive a valid Edgeworth type expansion of Section 3 for ELJ.

As in standard dual likelihood theory we can now build a likelihood ratio (a quasi-dual likelihood ratio in the present context) which can be used directly to test whether the overidentifying restrictions in model (1) hold; specifically, in Appendix A (see Theorem 6) we show that

$$(8) \qquad W_J = \sum \log\left(1+\widetilde{\lambda}^T\psi\left(z_i,\widetilde{\theta}\right)\right) \xrightarrow{d} \chi^2(r-p) \quad \text{as } n\to\infty.$$

One of the most interesting feature of the empirical likelihood (and more generally other empirical discrepancy based test statistic) is that the information implied in the moment condition $E\psi(z_i,\theta_0)=0$ can be used to provide a more efficient estimator of the distribution p_i than the nonparametric maximum likelihood estimator $\widehat{p}_i = 1/n$.

Recall that the constrained probabilities $\widetilde{p}_i$ are given by: $\left(1+\widetilde{\lambda}^T\psi\left(z_i,\widetilde{\theta}\right)\right)^{-1}/n$ (with $\begin{bmatrix} \widetilde{\lambda}^T & \widetilde{\theta}^T \end{bmatrix}^T$ solution of the saddlepoint problem (5)) and hence an estimator of the unknown distribution function F_0 can be based on $\widetilde{F_0} = \sum\widetilde{p}_i I\{z_i \le z\}$ ($I\{\cdot\}$ is the indicator function). In particular Theorem 7 in Appendix A shows that estimator $\widetilde{p}_i$ has limiting distribution given by:

$$(9) \qquad n^{1/2}\sum(\widetilde{p}_i-p_i)I(z_i\le z) \xrightarrow{d} N\left(0,\eta^2\right)$$

where $\eta^2 = \sigma^2 - B^T \Sigma^{-1} \left(I - D V^{-1} D^T \Sigma^{-1} \right) B$, $\sigma^2 = p_i (1 - p_i) I \{ z_i \leq z \}$, and $B = E (\psi (z_i, \theta_0) I \{ z_i \leq z \})$ for some fixed z. It is well-known that the maximum non-parametric likelihood estimator $\widehat{p}_i$ has limiting distribution:

$$n^{1/2} \sum (\widehat{p}_i - p_i) I (z_i \leq z) \xrightarrow{d} N \left(0, \sigma^2\right),$$

hence, it is evident that the probability estimator $\widetilde{p}_i$ based on the empirical likelihood approach is more efficient than the empirical probabilities $\widehat{p}_i = 1/n$.

Notwithstanding all these nice first order asymptotic properties characterising the empirical likelihood approach to inference for overidentified econometric models, a simple simulation study shows that the limiting χ^2 distribution, despite improving upon both recently developed (Hansen, Heaton, and Yaron, 1996) GMM tests and standard bootstrap based tests (see Section 4 below), gives still a relatively poor approximation for small sample sizes (see in particular Table II). To illustrate this, we consider the overidentified model:

$$E \left[\begin{array}{cc} z - \theta_0 & z^2 - s(\theta) \end{array} \right]^T = \left[\begin{array}{cc} 0 & 0 \end{array} \right]^T, \tag{10}$$

with the following two specifications: $z \sim N(\theta, \theta^2 + 1)$ with $s(\theta) = 2\theta^2 + 1$ (Qin and Lawless, 1994) and $z \sim \chi^2(1)$ with $s(\theta) = \theta^2 + 2\theta$ (Imbens, Spady, and Johnson, 1998). Tables I and II report the simulations result for sample sizes $n = 50$ and $n = 100$. newline

TABLES I and II approxamately here

This simple Monte Carlo experiment suggests that higher order expansions might be helpful in order to obtain asymptotic refinements to the limiting χ^2 distribution. Interestingly, the least favourable construction allows us to use the higher order asymptotics machinery developed for standard parametric statistical theory, despite the fact that the class of models under investigation in the present paper are genuinely semiparametric.

3 An Edgeworth approximation to the distribution of the empirical likelihood ratio J Test

In this section we first develop a stochastic expansions for the ELJ; we then use Hayakawa (1977) technique to obtain an Edgeworth approximation to the distribution of the ELJ test obtained in the previous section.

In addition to Assumptions A0 to A7, we assume also the following holds with probability 1,

A8 *The function* $\psi(z_i, \theta)$ *is five times continuously differentiable in a neighbourhood of* θ_0,

A9 *For some* $\epsilon > 0$, $\sup_{\theta^* \in \Gamma(\theta_0,\tau)} \|\psi(z_i, \theta^*)\|^{4+\epsilon} < \Psi_4(z_i)$ *with* $\Psi_4(z_i)$ *a bounded integrable function,*

A10 *For sufficiently small* τ, *and some* $\epsilon^{'} > 0$ *any* $j_1 + j_2 + ... + j_p = j$ $(j = 1, 2, 3, 4)$, $\sup_{\theta^* \in \Gamma(\theta_0,\tau)} \left|\partial^j \psi^\alpha(z_i, \theta^*) / \partial^{j_1}\theta_1 \partial^{j_j}\theta_2 \cdots \partial^{j_p}\theta_p\right|^{v+\epsilon^{'}} < \Psi_5(z_i)$ *with* $\Psi_5(z_i)$ *a bounded integrable function for* $v = 1, ..., 4$.

A11 *For sufficiently small* τ, *and some* $\epsilon^{''} > 0$,

$$\sup_{\theta^* \in \Gamma(\theta_0,\tau)} \left|\partial^5 \psi^\alpha(z_i, \theta^*) / \partial^{j_1}\theta_1 \partial^{j_j}\theta_2 \cdots \partial^{j_{p'}}\theta_p\right|^{v+\epsilon^{''}} < \Psi_6(z_i)$$

with $\Psi_6(z_i)$ *a bounded integrable function for any* $j_1 + j_2 + ... + j_{p'} = 5$.

A12 *For sufficiently small* τ', *some* $\varepsilon > 0$ *and* θ_0,

$$\sup_{\lambda^* \in \Gamma(0,\tau')} \left|\partial^5 W_J / \partial^{j_1}\lambda_1 \partial^{j_j}\lambda_2 \cdots \partial^{j_{p'}}\lambda_r\right|^{v+\varepsilon} < \Psi_7(z_i)$$

with $\Psi_7(z_i)$ *a bounded integrable function for any* $j_1 + j_2 + ... + j_{p'} = 5$.

REMARK II Assumptions A8-A11 ensure that the derivatives arrays involving the maximum empirical likelihood estimator in the stochastic expansion $(C1)$ in Appendix C are "well behaved"; in particular it follows that $\Pr\left\{\widetilde{\theta} \in \Gamma(\theta_0, \tau)\right\} > 1 - o(n^{-1})$ by applying a fixed point argument as in Bhattacharya and Ghosh (1978) to the extremum estimator $\widetilde{\theta} := \partial W^- / \partial\theta \mid_{\lambda=0} = 0$, and that the remainder 5th order array $\partial^5 \psi^\alpha \left(z_i, \theta_0 + \varsigma\left(\widetilde{\theta} - \theta_0\right)\right) / \partial^{j_1}\theta_1 \partial^{j_j}\theta_2 \cdots \partial^{j_{p'}}\theta_p$ for any $\varsigma \in [0,1]$ is bounded. Assumption A12 is standard in dual likelihood theory, implying the existence of a stochastic expansion for the quasi-maximum dual likelihood estimator $\widetilde{\lambda} := \partial W^- / \partial\lambda \mid_{\theta=\theta_0} = 0$, such that $\Pr\left\{\widetilde{\lambda} \in \Gamma(0, \tau')\right\} > 1 - o(n^{-1})$ by von Bahr's inequality. Hence by the asymptotic independence of we can deduce that the following: $\Pr\left\{\widetilde{\lambda}, \widetilde{\theta} \in \Gamma(\theta_0, \tau) \bigcap \Gamma(0, \tau')\right\} > 1 - o(n^{-1})$.

Throughout the rest of the paper, we use the tensor summation convention (i.e. for any two repeated indices, their sum is understood) and the notation $[\cdot]$ denotes initial summation over repeated indices within the parenthesis -see Hayakawa (1977) for more details. We are also using the following alphabetic conventions: Greek alphabet indices range from 1 to r (i.e. the number of restrictions), the letters $a, b, ..., h$ go from 1 to $r+p$ (corresponding therefore to the dimension of the saddlepoint estimator $\zeta = \left[\begin{array}{cc} \widetilde{\lambda} & \widetilde{\theta} - \theta_0 \end{array}\right]^T$) and letters starting from s onwards run from 1 to p (i.e. the dimension of the parameter θ).

Recalling from Section 2 that $\psi^\alpha = \psi^\alpha(z_i, \theta)$ is the αth component of the vector $\psi(z_i, \theta)$, let $\psi^\alpha_s = \partial\psi^\alpha / \partial\theta^s, \psi^\alpha_{st} = \partial\psi^\alpha / \partial\theta^s \partial\theta^t$ etc. denote the partial derivatives with respect to the sth component of the parameter vector θ^s; let $\zeta^a = \left[\begin{array}{cc} \widetilde{\lambda}^\alpha & \left(\widetilde{\theta} - \theta_0\right)^s \end{array}\right]$ be the componentwise expression for the saddlepoint estimator ζ defined in Section 2. Finally, we consider the arrays of derivatives of W_J with respect to λ^α and θ^s scaled by powers of $n^{-k/2}$ as in Hayakawa (1977); let the scaled arrays of derivatives be:

(11) $$U_{a_1 a_2 \ldots a_k} = n^{-k/2} \partial^k W / \partial\zeta^{a_1} \partial\zeta^{a_2} \ldots \partial\zeta^{a_k} \quad k = 1, ..., 4,$$

so that

$$U_a = n^{-1/2}\left[\begin{array}{cc} \partial_{\lambda^\alpha} W & \partial_{\theta^s} W \end{array}\right]$$

is the vector defining the saddlepoint equations (6), and so on for higher order arrays (see Appendix B for details). Any array of derivatives defined in (11) evaluated at the saddlepoint is distinguished by the addition of a tilde.

Using Chandra and Ghosh (1979) general result, it can be shown that under the additional assumptions A8-A12, on a compact set $\mathcal{K}_n$ such that

$$\Pr\{\mathcal{K}_n\} = 1 + o\left(n^{-1}\right),$$

there exists a stochastic expansion for W_J under the null hypothesis whose form is

$$(12) \qquad \begin{aligned} W_J = {} & U^{\alpha\beta}U_\alpha U_\beta - U_{abc}\left[U^{a\alpha}U_\alpha\right]\left[U^{b\alpha}U_\alpha\right]\left[U^{c\alpha}U_\alpha\right]/3 - \\ & U_{abe}\left[U^{a\alpha}U_\alpha\right]\left[U^{b\beta}U_\beta\right]U^{ef}\left[U^{c\gamma}U_\gamma\right]\left[U^{d\delta}U_\delta\right]U_{cdf}/4 + \\ & U_{abcd}\left[U^{a\alpha}U_\alpha\right]\left[U^{b\alpha}U_\alpha\right]\left[U^{c\alpha}U_\alpha\right]\left[U^{d\alpha}U_\alpha\right]/12 + o_p\left(n^{-1}\right). \end{aligned}$$

where the matrix inverse $U^{\alpha\beta}$ in the leading term of the expansion is the sample version of $k^{\alpha\beta} = \Sigma^{-1}\left(I - DV^{-1}D^T\Sigma^{-1}\right)$ (see (7) above). (see Appendix C for the derivation).

Let:

$$(13) \qquad \begin{aligned} k_{ab} &= E\left(U_{ab}\right), k_{a,b} = E\left(U_a U_b\right), k_{abc} = n^{1/2}E\left(U_{abc}\right), \\ k_{a,bc} &= n^{1/2}E\left(U_a U_{bc}\right), k_{a,b,c} = n^{1/2}E\left(U_a U_b U_c\right), k_{abcd} = nE\left(U_{abcd}\right), \\ k_{ab,cd} &= nE\left(U_{ab}U_{cd}\right), k_{a,b,cd} = nE\left(U_a U_b U_{cd}\right), k_{a,b,c,d} = nE\left(U_a U_b U_c U_d\right) \end{aligned}$$

be the (multivariate) joint moments and cumulants of the arrays defined in (11) (see for example (McCullagh, 1987, Ch. 3)).

In order to obtain an asymptotic expansion for the density of W_J, we follow an approach similar to the one developed by Peers (1971) and Hayakawa (1977). Using a multivariate Edgeworth Type A expansion to an order $o\left(n^{-1}\right)$ for the three arrays

U_a, U_{ab}, U_{abc}, we are able to derive an asymptotic expansion for the characteristic function (CF) of W_J. By a formal Fourier inversion of this latter expansion, we get the required Edgeworth expansion.

Notice that, under the null hypothesis $H_0 : \lambda = 0$, the vector U_a has dimension $r \times 1$, while the arrays U_{ab} and U_{abc} are $(r+p)^2$ and $(r+p)^3$, respectively. This is not a serious problem in terms of obtaining a valid Edgeworth expansion for W_J in decreasing powers of n^{-j} $(j = 0, 2, 4, ...)$, given the asymptotic independence of the two estimators $\widetilde{\lambda}$ and $\widetilde{\theta}$ (see Remark III below, for more details). It does imply, though, that the $O(n^{-1})$ term in the expansion is not just a finite linear combination of $\chi^2(r-p+j)$ densities (as it is typically the case in the asymptotic expansion of a likelihood ratio in the parametric case, see for example Chandra (1985)). Indeed, as shown below (cf. (16)), the resulting expansion involves also the first derivative of a $\chi^2(r-p)$ density, with coefficients given by mixed cumulants related to the $p \times 1$ maximum empirical likelihood estimator $\widetilde{\theta}$.

Let $k(\cdot,\cdot,\cdot)$ denote the various cumulants associated to the U's as described in (13).

In general, the requisite Edgeworth series $h(U_a, U_{ab}, U_{abc}) = h(\cdot)$ can be informally expressed using the following compact operator representation:

$$(14) \quad h(\cdot) = \exp\left(\sum{}' (-1)^{x+y+z} \frac{(\partial_a)^x}{x!}\frac{(\partial_{ab})^y}{y!}\frac{(\partial_{abc})^z}{z!} k(x,y,z)\right) g(U_a, U_{ab}, U_{abc})$$

where the sum $\sum'$ is over $x + 2y + 3z \geq 3$, $(\partial.)^{x'}$'s are partial derivative operators applied $x' = x, y, z$ times (i.e. for example $\partial_a^2 \cdot = (\partial_a \cdot)(\partial_b \cdot))$ to the function $g(\cdot)$, which is a degenerate multivariate normal distribution representing the $O(1)$ (joint) distribution of the components of U_a, U_{ab} and U_{abc}

$$(15) \qquad g(U_a, U_{ab}, U_{abc}) = (2\pi)^{-r/2} \left|k_{\alpha,\beta}^{-1/2}\right| \exp\left(-k^{\alpha,\beta} U_\alpha U_\beta / 2\right) \times \prod_{a,b} \delta\{U_{ab} - k_{ab}\} \prod_{a,b,c} \delta\left\{U_{abc} - n^{-1/2} k_{abc}\right\}.$$

The functions $\delta\{\cdot\}$ in (15) are Dirac's delta functions whose properties are described in Hayakawa (1977, p. 365).

As shown in Appendix D, the resulting CF_{W_J} is, up to the order $o(n^{-1})$,

$$(16) \qquad CF_{W_J}(t) = (1-2\iota t)^{-(r-p)/2}\left(1+n^{-1}\left(P_0'\iota t+\sum_{j=0}^{3}P_j/(1-2\iota t)^j\right)\right)$$

where $\iota = (-1)^{1/2}$, and the various P's are complicated scalar functions of the cumulants as described in (13) and of some matrices defined in $(D3)$ of Appendix D, resulting from the integration of the $\delta\{\cdot\}$ functions. By a formal inversion of (16), we can prove the following theorem:

Theorem 1 *The asymptotic expansion for the density function of the empirical likelihood ratio test for overidentifying restrictions under the null hypothesis* (1) *is, up to the order* $o(n^{-1})$*:*

$$(17)\ \varphi_{W_J}(x) = f_{\chi^2}(x;r-p)+n^{-1}\left(P_0'f_{\chi^2}'(x;r-p)+\sum_{j=0}^{3}P_jf_{\chi^2}(x;r-p+2j)\right)$$

where $f_{\chi^2}(x;r-p)$ *denotes the density of a central chi-squared distribution with* $r-p$ *degrees of freedom and* $f_{\chi^2}'(x;r-p)$ *is its first derivative.*

REMARK III In order to justify expansion (17) as a valid Edgeworth expansion, we may combine the general results of Chandra and Ghosh (1979, Remark 2.7), Chandra (1985) and Sargan (1980). In particular, by using the delta method, the *odd-even* and the *modulo 2* properties of the generalised polynomials described in Chandra (1985, p. 102 and ff.) hold, so that the resulting expansion is expressed in the usual even powers of n^j $(j = 0, 1, 2...)$; this fact also implies that the expansions for the cumulants themselves are decreasing powers of n^{-1} as in ordinary parametric likelihood calculations. Despite this regular (parametric) behaviour of cumulants, the expansion features also the first derivative of a $\chi^2(r-p)$ density. Its presence is due essentially to the maximum empirical likelihood estimator which can be thought of as a nuisance parameter. Given its asymptotic independence with the quasi maximum dual likelihood estimator $\widetilde{\lambda}$, we can use Sargan's (1980) Theorem 1 to provide a formal justification to the asymptotic expansion (17). In practice, we consider the quasi-dual likelihood W^- in (3) as Sargan's (1980) function e. It is not difficult to check that his regularity conditions are met for W^- (at least to the second order approximation), including his Lemma 1 which holds for $\partial W^- (\lambda(\theta), \theta) / \partial\theta = 0^2$.

It should be noted that expansion (17) does not contain the second derivative of a $\chi^2 (r-p)$ as we are evaluating the derivative arrays (11) under the null (cf. Appendix B).

Hence, we may deduce that:

$$\sup_{\xi \in R} \left| \Pr\{W_J \leq \xi\} - \int_{-\infty}^{\xi} \varphi_{W_J}(x) \right| = o\left(n^{-1}\right)$$

under the additional regularity conditions:

A13 *The vector $U^i = U^i_{a_1 a_2 \ldots a_k}$ $(a \geq 1, 1 \leq k \leq 4)$, containing all the different elements of the first four U arrays for the ith observation has finite absolute fourth moments bounded uniformly in ξ :*

$$E\left(\left\| U^i_{a_1 a_2 \ldots a_k} \right\|^4 < \infty\right)$$

A14 *The following Cramér condition holds:*

$$\limsup_{\|t\| \to \infty} \left| E \exp\left(\iota t^T U^i\right) \right| < 1.$$

Upon integration of equation (17), we obtain the cumulative distribution function (cdf) of W_J up to $o(n^{-1})$, which is given by:

$$\text{(18)} \quad cdf_{W_J}(x) = F_{\chi^2}(x, r-p) + n^{-1} \left(\sum_{j=0}^{3} P_j F_{\chi^2}(x, r-p+2j) + P_0' x f_{\chi^2}(x; r-p) \right)$$

where $F_{\chi^2}(x; r-p)$ is the distribution function of a central chi-squared variate. The correct (up to order $o(n^{-1})$) critical value cv_α of $cdf_{W_J}(x)$, at the level α, can therefore be obtained as a solution of:

$$cdf_{W_J}(x_{cv_\alpha}) = \alpha. \tag{19}$$

In order to compute the asymptotic refinements in (18), we need to estimate the various P's appearing in (18). This implies that we are effectively approximating the distribution of ELJ with an empirical Edgeworth distribution. By strengthening the moments assumption A13 to:

A13' $E\left(\left\|U^i_{a_1 a_2 \ldots a_k}\right\|^8 < \infty\right)$,

we can get $n^{1/2}$ consistent estimates of the relevant cumulants evaluated at the maximum empirical likelihood estimator estimator, and hence the resulting empirical Edgeworth density approximates the density of ELJ through the order $O(n^{-1})$.

The computation of the empirical cumulants is based on the arrays ϕ^{I_ν} for any set of indices $i_1, i_2, \ldots$ in I_ν and $1 \leq i_1, i_2, \ldots \leq n$ such that its coefficients satisfy the criterion of unbiasedness (McCullagh, 1987, p.91) and are given by the general formula:

$$(-1)^{\nu-1}(\nu-1)!/(n-1)^{(\nu-1)}$$

with $\nu \leq n$ and $(n-1)^{(\nu-1)} = (n-1)(n-2)\ldots(n-\nu+1)$ (more computational details can be found in Bravo (1999)). This calculation is not only time consuming but very difficult, see for example McCullagh (1987, Ch. 4)

An alternative way to obtain asymptotic refinements to the empirical likelihood J test is to use the bootstrap. In the next section we explore such an alternative. We also provide some numerical evidence of the magnitude of the finite sample errors in the sizes of the empirical likelihood based J test with asymptotic, Edgeworth based and bootstrapped critical values.

4 The bootstrap calibrated empirical likelihood ratio J test

As originally pointed out by Hall and Horowitz (1996), bootstrapping overidentified moment based econometric models does not provide automatic asymptotic refinements for asymptotically pivotal statistics because the population moment condition does not hold in the sample. As a consequence the standard bootstrap estimator for the J statistic for overidentifying restrictions is inconsistent (Brown and Newey, 1995). Hall and Horowitz (1996) suggest to base the bootstrap estimation on the recentered moment conditions (with respect to the empirical moment condition evaluated at the GMM estimator). Alternatively, Brown and Newey (1995) suggest using the empirical likelihood estimator in place of the usual uniform probabilities and to implement the bootstrap estimation by sampling the observations with weights given by the resulting empirical likelihood based probabilities. This method is an example of a class of weighted bootstrap techniques, the intentionally biased bootstrap (b-bootstrap henceforth) introduced by Hall and Presnell (1999). In this latter approach, conditionally on the data, the resampling probabilities are chosen so as to minimise a given distance of the weighted bootstrap distribution from the usual uniform probabilities subject to some constraint. In the context of overidentified moment models, the constraints are represented by the overidentifying restrictions themselves therefore the b-bootstrap sample carries out the recentering automatically. With either method of recentering, the coverage error is $o\left(n^{-1}\right)^{3}$.

The bootstrap procedure we propose was originally suggested by Owen (1988) in his seminal paper about empirical likelihood inference. As mentioned in the Introduction, it can be cast in the heuristic behind the b-bootstrap approach, but rather than biasing the sample so that it fulfils a constraint, we check whether the (uniform) bootstrap sample itself fulfils assumption A0. The requirement that 0 is contained in the convex hull span by the estimating equations guarantees essentially the existence and positiveness of the ELJ ratio test; it then follows that if the bootstrap sample does satisfy this assumption, then the resulting bootstrapped ELJ, say W_J^*, will have the required asymptotic $\chi^2(r-p)$ calibration, thus the distribution of W_J^* conditional on the original sample is the bootstrap estimator of the distribution of W_J. In Theorem 3 below we show that if we use the bootstrapped distribution W_J^* to obtain the critical value cv_α of $cdf_{W_J}(x)$ as in (19) above, then the coverage error will be of order $o(n^{-1})$ as the previous two methods, but as opposed to these latter two, our proposed bootstrap procedure exploits another important feature of empirical likelihood based inference. It is well known in fact that empirical likelihood carries out automatically an implicit studentisation of the statistic under investigation. This property is preserved by our bootstrap methodology, implying that we do not need to compute any pivotal statistic in order to obtain asymptotic refinements. Secondly, the estimation of the bootstrap saddlepoint estimator is relatively straightforward to compute, as the bootstrapped moment condition model is clearly just identified in the parameters λ and θ (as it is in the original model).

As usual let an asterisks denote quantities based on the bootstrap sample $\{z_i^*\}_{i=1}^n$. Let us assume now that:

BA0 $\Pr\{0 \in ch\{\psi_*(z_1^*, \theta), ..., \psi_*(z_n^*, \theta)\}\} \to 1$ as $n \to \infty$;

then under the previous assumptions, we can derive (conditional on the original sample) a bootstrap based saddlepoint estimator which is consistent and asymptotically normal, and a stochastic expansion as for the bootstrapped empirical likelihood ratio J test which is analogous to (12) in Section 3. Hence we can derive the following theorem:

Theorem 2 (Bootstrapped Likelihood Ratio J Test) *The asymptotic expansion for the density function of the b-bootstrap test for overidentifying restrictions under* *the null hypothesis* (1) *is up to the order* $o\left(n^{-1}\right)$:

$$(20)\varphi_{W_J^*}(x) = f_{\chi^2}(x; r-p) + n^{-1}\left(P_0^{\prime *} f_{\chi^2}^{\prime}(x; r-p) + \sum_{j=0}^{3} P_j^* f_{\chi^2}(x; r-p+2j)\right)$$

where $f_{\chi^2}(x; r-p)$ *denotes the density of a central chi-squared distribution with* $r-p$ *degrees of freedom and* $f_{\chi^2}^{\prime}(x; r-p)$ *is its first derivative.*

We can now use the bootstrapped distribution of the ELJ test to obtain asymptotic refinements for the original ELJ test; let $x_{cv_\alpha}^*$ be the critical value such that:

$$(21) \qquad \overset{*}{\Pr}\left\{W_J^* \geq x_{cv_\alpha}^*\right\} = \alpha.$$

We can prove the following theorem.

Theorem 3 *If the overidentifying restrictions of model* (1) *hold, then*

$$(22) \qquad \Pr\left\{W_J \geq x_{cv_\alpha}^*\right\} = \alpha + o\left(n^{-1}\right).$$

To assess the finite sample behaviour of the ELJ with asymptotic, Edgeworth and bootstrap critical values we carry out a small Monte Carlo experiment. The model used in the simulation has been analysed by Hall and Horowitz (1996), and more recently by Imbens, Spady, and Johnson (1998). Consider the following simplified version of asset pricing model, defined as:

$$(23) \qquad \psi(z, \theta) = \begin{bmatrix} \exp\left(\mu - \theta(x+y) + 3y\right) - 1 \\ y\left(\exp\left(\mu - \theta(x+y) + 3y\right) - 1\right) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

where μ is a known normalisation constant set to -0.72, $\theta_0 = 3$, and the vector $z^T = \begin{bmatrix} x & y \end{bmatrix}$ has a bivariate $N(0, 0.16)$ distribution with 0 correlation coefficient.

After some painstaking calculations (despite the model being quite simple, with two moment conditions i.e. $\alpha = 2$ and $\dim \theta = 1$), the following Edgeworth based critical values x_{cv_α} solution for (19) for the levels $\alpha = [0.10, 0.05, 0.01]$ are found: $[4.412, 8.374, 17.645]$ for $n = 50$ and $[3.9103, 7.096, 12.012]$ for $n = 100$ which are obviously closer to the "exact" values $[5.934, 10.674, 23.267]$ and $[4.524, 8.299, 15.812]$ computed from 5000 Monte Carlo replications than the reference $\chi^2(1)$ asymptotic distribution. These high values for the upper quantiles of the distribution of the ELJ test are not surprising though, given that the test is based on the maximum empirical likelihood estimator $\widetilde{\theta}$ whose sampling distribution is characterised by occasional big outliers. For example, for $n = 50$ the 0.025 and 0.0975 quantiles for $\widetilde{\theta}$ are 2.26 and 4.61, respectively. Comparing these values with those reported in Imbens, Spady, and Johnson (1998) (2.55 and 6.92 for the continuously updated GMM and 2.35 and 3.93 for the exponential tilting), it seems that the empirical likelihood based estimator is preferrable to some sophisticated GMM estimators, but it is worse than the one based on the exponential tilting. However more simulation studies are needed in order to verify this last assertion.

The bootstrap calibration can be implemented in the following steps:

1) Sample the data-generation process,

2) Verify that condition BA0 is satisfied. A simple way to do so is to check whether $\sum \widetilde{p}_i^{-1} = n$, where $\widetilde{p}_i$ are the estimated implied probabilities as in (2) and n is the sample size. In fact, if $\widetilde{p}_i^{-1} = m$ (with $m \leq n$) ,then there are m observations within the face of the convex hull containing 0, therefore a small m (possibly tending to 0) provides strong evidence against the assumption BA0,

3) Estimate the bootstrap saddlepoint and construct the bootstrap empirical likelihood J test,

4) Repeat step 1-3 B times[4].

Table 3 reports the empirical sizes for the test for overidentifying restrictions based on the asymptotic $\chi^2(1)$ critical values, on the Edgeworth corrected as in (19) and on the bootstrap critical values as in (21) for 5000 replications using the SPLUS pseudo-random number generators, and 0.01 and $\exp\{\max(z)\}$ as starting values[5] for λ and θ respectively to initialise Newton-gradient algorithm used to compute the test statistic.

TABLE III approximately here

Firstly, it is interesting to note the good performance of the ELJ test with asymptotic $\chi^2(1)$ calibration when compared with the J test with asymptotic as well as bootstrap critical values analogue proposed by Hall and Horowitz (1996, Table I). This fact is explained by recalling that the empirical likelihood method takes into account all the information available in the sample, resulting in estimated probabilities that are more efficient (see (9)) than the uniform probability used in the standard bootstrap approach.

In view of the remark that "...[Edgeworth] corrections tend to work well when the error on the crude asymptotics is small ... and are poor when that error is large..." made by Phillips and Park (1988), it is perhaps not surprising that the Edgeworth corrected ELJ test is quite good. The performance of the bootstrapped ELJ is almost stunning (especially when compared with the standard bootstrapped J statistic), with empirical sizes not statistically different form the nominal sizes at the 0.05 significance level. Figures 1 and 2 show the "exact" and the bootstrapped densities of the ELJ test and of maximum empirical likelihood estimator $\tilde{\theta}$, respectiveley.

Figure 1 and 2 approximately here

5 Conclusions

We have developed a valid Edgeworth expansion to the density of the empirical likelihood ratio test for overidentified moment conditions econometric models. The approximation, a part from being of its own interest, is used to justify a "hybrid" bootstrap approach which is in the same spirit of the biased bootstrap method. This latter method performs extremely well at least in the small Monte Carlo study reported in this paper. Although more simulations studies are needed to assess the finite sample performance of the bootstrap method we are proposing, this paper shows that the combination of two originally competing resampling methods such as the empirical likelihood and the uniform bootstrap can be an extremely useful tool for obtaining highly accurate inference in notoriously difficult models such as those based on moment restrictions. As the empirical likelihood is just one of the possible methods to obtain (efficient) implied probabilities, it seems quite plausible that the combination of standard bootstrap with other empirical discrepancy based probabilities might deliver the same kind (if not better) accurate inference. This certainly deserves future attention.

It is also worth pointing out that the empirical likelihood based bootstrap approach proposed in this paper can be applied to other test statistics for moment based econometric models. Recently, Smith, Chesher, and Peters (1999) have obtained an Edgeworth expansion for the null distribution of the outer product of gradients (OPG) form of conditional moment test statistic, providing analytic evidence on the poor finite sample performance of the OPG test. Given that the empirical likelihood approach to inference can be cast in their augmented likelihood framework, it might be interesting to see whether our bootstrap procedure can be used to improve the finite sample properties of this class of tests.

Finally, we notice that the present approach can be extended to more general data structures, such as time series, by using blocking techniques as shown in Kitamura (1997). We believe that the same kind of techniques can be applied to our bootstrap calibrated empirical likelihood approach to inference, provided we model the dependence of the data so that we can justify (in the Edgeworth sense) the resulting asymptotic expansion as in Götze and Hipp (1983) and Götze and Hipp (1994). It should be noted that in the case of weakly dependent observations the internal studentisation property of the empirical likelihood method is particularly useful because one needs not compute the complex correction factor necessary to studentise the J

test.

Francesco Bravo, Department of Economics and Related Studies, University of York, York, YO10 5DD, United Kingdom. email: fb6@york.ac.uk

APPENDICES

A First order asymptotics for the empirical likelihood ratio J test

All the following theorems are based on or adapted from Kitamura (1997) and Qin and Lawless (1994). They are included here for completeness. In the next theorem we establish the $n^{1/2}$ consistency for the saddlepoint estimator $\tilde{\zeta} = \begin{bmatrix} \tilde{\lambda}(\theta_0) & \tilde{\theta} \end{bmatrix}^T$.

Theorem 4 *Under Assumptions A0-A4(ii),* $\tilde{\lambda}(\theta_0) \xrightarrow{p} 0$ *and* $\tilde{\theta} \xrightarrow{p} \theta_0$.

The consistency of $\tilde{\lambda}(\theta_0)$ is proven by noting that under assumption A4 (*ii*), $\max_i \|\psi(z_i, \theta_0)\| = o\left(n^{1/2}\right)$, and then following Owen's (1990) argument to show that $\left\|\tilde{\lambda}(\theta_0)\right\| = O_p\left(n^{-1/2}\right)$ which implies the consistency of $\tilde{\lambda}(\theta_0)$. Next, let $I\{\cdot\}$ denote the indicator function; by considering $\psi_{R_n}(z_i, \theta) = \psi(z_i, \theta) I\{R_n\}$ with $R_n = \left\{z : \|\psi(z, \theta)\| \leq n^{1/(2+2\eta)}, \forall\, \theta \in \Theta\right\}$, and any vector $l_n \in \Gamma\left(0, n^{-1/(2+\eta)}u\right)$ with $\|u\| = 1$, it is possible to show that

$$E\left(-n^{1/(2+\eta)} \log\left(1 + \tilde{l}_n(\theta)^T \psi_{R_n}(z_i, \theta)\right)\right) = -E\|\psi(z_i, \theta)\| + o(1)$$

by the mean value theorem and $\tilde{l}_n(\theta) = \underset{l \in \Gamma\left(0, n^{-1/(2+\eta)}u\right)}{\arg\min} E\left(-\log\left(1 + l^T \psi(z_i, \theta)\right)\right)$. By continuity and assumption A3, it also follows that:

$$\lim_{n\to\infty} \lim_{\tau \downarrow 0} n^{1/(2+\eta)} E \sup_{\theta^* \in \Gamma(\theta, \tau)} -\log\left(1 + \tilde{l}_n(\theta^*)^T \psi_{R_n}(z_i, \theta^*)\right) = -E\|\psi(z_i, \theta)\|. \quad \text{(A1)}$$

By the uniqueness of θ_0, and $(A1)$ there exist a finite number of open spheres $\Gamma(\theta_k, \tau_k)$, $k = 1, 2, ..., j$ such that $\bigcup_{k=1}^{j} \Gamma(\theta_k, \tau_k)$ covers the set $\Theta(\tau) = \Theta - \Gamma(\theta_0, \tau)$ and τ are chosen such that they satisfy:

$$n^{1/(2+\eta)} E \sup_{\theta^* \in \Gamma(\theta_k, \tau_k)} -\log\left(1 + \tilde{l}_n(\theta^*)^T \psi_{R_n}(z_i, \theta^*)\right) + o(1) = -2K_k,\; K_k > 0.$$

As $\max_i \sup_{\theta^* \in \Gamma(\theta_k,\tau)} \|\psi(z_i,\theta^*)\| = o\left(n^{1/(2+2\eta)}\right)$, the weak law of large numbers (WLLN) implies that there exists j large enough $n > n_k$:

$$\Pr\left\{-n^{-1}\sum \sup_{\theta^* \in \Gamma(\theta_k,\tau_k)} \log\left(1 + \tilde{l}_n(\theta^*)^T \psi(z_i,\theta^*)\right) > -n^{-1/(2+\eta)} K_k\right\} < \epsilon/(2j), k = 1, 2, ..., j$$

for a small $\epsilon > 0$, and hence :

$$\Pr\left\{-n^{-1}\sum \sup_{\theta^* \in \Theta(\tau)} \log\left(1 + \lambda(\theta^*)^T \psi(z_i,\theta^*)\right) > -n^{-1/(2+\eta)} K\right\} < \epsilon/2, \; K = \min_k K_k$$

for all $n > \max_k n_k$. By the minimisation property of the Lagrange multiplier vector $\tilde{\lambda}(\theta)$ there exists a sufficiently large n' such that $\forall n > n'$:

$$\Pr\left\{-n^{-1}\sum \sup_{\theta^* \in \Theta(\tau)} \log\left(1 + \tilde{\lambda}(\theta^*)^T \psi(z_i,\theta)\right) > -2n^{-1/(2+\eta)} K\right\} < \epsilon/2. \quad \text{(A2)}$$

By the convexity of the criterion function (3):

$$-n^{-1}\tilde{\lambda}(\theta)^T \sum \psi(z_i,\theta_0) \le -\sum \log\left(1 + \tilde{\lambda}(\theta)^T \psi(z_i,\theta_0)\right) \le 0,$$

as $n^{-1}\sum \psi(z_i,\theta_0) \xrightarrow{p} 0$, and $\tilde{\lambda}(\theta_0) = O_p\left(n^{1/2}\right)$, it follows:

$$n^{-1}\sum \log\left(1 + \tilde{\lambda}(\theta)^T \psi(z_i,\theta_0)\right) \xrightarrow{p} 0. \quad \text{(A3)}$$

for a sufficiently large $n > n''$. Combining $(A2)$ and $(A3)$ yields

$$\Pr\left\{n^{-1}\sum \log\left(1 + \tilde{\lambda}(\theta)^T \psi(z_i,\theta_0)\right) < -K/2\right\} < \epsilon/2$$

leading to:

$$\Pr\left\{\tilde{\theta} \in \Gamma(\theta_0,\tau)\right\} > 1 - \epsilon$$

for $n > \max\left(n', n''\right)$ and hence $\tilde{\theta} \xrightarrow{p} \theta_0$. *QED*

Having established that the unique solution of the saddlepoint problem is consistent at the standard (parametric) $n^{1/2}$ rate, we now prove its asymptotic normality.

Theorem 5 *Under A0-A7 the saddlepoint estimator $n^{1/2}\tilde{\zeta}$ has the following asymptotic distribution:*

$$n^{1/2}\begin{bmatrix} \tilde{\lambda} \\ \tilde{\theta} - \theta_0 \end{bmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma^{-1} - \Sigma^{-1}DV^{-1}D^T\Sigma^{-1} & O \\ O & V^{-1} \end{bmatrix}\right)$$

Taylor expanding (6) around the point $\begin{bmatrix} 0 & \theta_0 \end{bmatrix}^T$ (0 is a $r \times 1$ vector of zeros) yields

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = n^{1/2} \begin{bmatrix} \partial_\lambda W^- \\ \partial_\theta W^- \end{bmatrix}_{(0,\theta_0)} + n^{1/2} \begin{bmatrix} \partial^2_{\lambda\lambda^T} W^- & \partial^2_{\lambda\theta^T} W^- \\ \partial^2_{\theta\lambda^T} W^- & \partial^2_{\theta\theta^T} W^- \end{bmatrix}_{(0,\theta_0)} \begin{bmatrix} \widetilde{\lambda} \\ \widetilde{\theta} - \theta_0 \end{bmatrix} + o_p(1), \tag{A4}$$

where

$$\begin{aligned} \partial_\theta W^-_{(0,\theta_0)} &= 0, \quad \partial_\lambda W^-_{(0,\theta_0)} = -\sum \psi(z_i, \theta)/n, \quad \partial_{\theta\theta^T} W^-_{(0,\theta_0)} = O_{(p\times p)}, \\ \partial^2_{\lambda\lambda^T} W^-_{(0,\theta_0)} &= -\sum \psi(z_i, \theta)\, \psi(z_i, \theta)^T / n, \quad \partial^2_{\lambda\theta^T} W^-_{(0,\theta_0)} = -\sum \left(\partial \psi(z_i, \theta)/\partial \theta^T\right)/n. \end{aligned}$$

As $n \to \infty$, by a straightforward application of WLLN, we get

$$\partial^2_{\lambda\theta^T} W^-_{(0,\theta_0)} \xrightarrow{p} D, \quad \partial^2_{\lambda\lambda^T} W^-_{(0,\theta_0)} = -\sum \psi(z_i, \theta)\, \psi(z_i, \theta)^T / n \xrightarrow{p} \Sigma,$$

where $D = E\partial\psi(z_i, \theta)/\partial\theta^T$, $\Sigma = E\psi(z_i, \theta)\,\psi(z_i, \theta)^T$. Inverting $(A4)$ yields

$$n^{1/2} \begin{bmatrix} \widetilde{\lambda} \\ \widetilde{\theta} - \theta_0 \end{bmatrix} = \begin{bmatrix} \left(\Sigma^{-1} - \Sigma^{-1} D V^{-1} D^T \Sigma^{-1}\right) \\ -V^{-1} D^T \Sigma^{-1} \end{bmatrix} n^{-1/2} \sum \psi(z_i, \theta_0) + o(1) \tag{A5}$$

with $V = D^T \Sigma^{-1} D$, from which the (joint) asymptotic normality follows as $n^{-1/2} \sum \psi(z_i, \theta_0) \xrightarrow{d} N(0, \Sigma)$ by central limit theorem and Slutsky theorem. *QED*

Theorem 6 (Corollary 2, Qin and Lawless (1994)) *Under assumption A0 to A7, the empirical likelihood ratio J test*

$$W_J = \sum \log\left(1 + \widetilde{\lambda}^T \psi\left(z_i, \widetilde{\theta}\right)\right) \xrightarrow{d} \chi^2(r - p) \quad \text{as } n \to \infty$$

if the overidentifying restrictions (1) *hold.*

As in standard likelihood theory, Taylor expanding W_J under the null hypothesis $\lambda = 0$, (see the first term of the expansion $(C3)$ below) about the saddlepoint $\widetilde{\lambda}$ and

$\widetilde{\theta}$, by WLLN and the continuous mapping theorem we have

$$\widetilde{D} = n^{-1}\sum_{i=1}^{n}\partial\psi\left(z_i,\theta\right)/\partial\theta\mid_{\theta=\widetilde{\theta}}\xrightarrow{p} D,$$
$$\widetilde{\Sigma} = n^{-1}\sum_{i=1}^{n}\psi\left(z_i,\theta\right)\psi\left(z_i,\theta\right)^T\mid_{\theta=\widetilde{\theta}}\xrightarrow{p}\Sigma,$$

so that the $O_p(1)$ leading term is of the Taylor expansion is given (see also $A5$ above) by

$$W_J = n\left(\sum_{i=1}^{n}\psi\left(z_i,\theta_0\right)/n^{1/2}\right)^T\left(\Sigma^{-1}-\Sigma^{-1}DV^{-1}D^T\Sigma^{-1}\right)\sum_{i=1}^{n}\psi\left(z_i,\theta_0\right)/n^{1/2}+o_p(1),$$

which can be rewritten as

$$W_J = n\left(\Sigma^{-1/2}\sum_{i=1}^{n}\psi\left(z_i,\theta_0\right)/n^{1/2}\right)^T(I-Q)\left(\Sigma^{-1/2}\right)\sum_{i=1}^{n}\psi\left(z_i,\theta_0\right)/n^{1/2}+o_p(1)$$

where $Q=\Sigma^{-1/2}DV^{-1}D^T\Sigma^{-1/2}$ and $\Sigma^{-1/2}$ is the matrix inverse of the symmetric positive definite square root of Σ ; we can then apply a result in Rao (1973, p.187) for quadratic forms in normal variables, to show that $W_J\xrightarrow{d}\chi^2(r-p)$ since Q is an idempotent matrix of rank p. *QED*

Theorem 7 (Theorem 1, Qin and Lawless (1994)) *Under Assumptions A0 to A7, the estimator $\widetilde{p}_i$ has limiting distribution given by:*

$$n^{1/2}\sum\left(\widetilde{p}_i-p_i\right)I\left(z_i\le z\right)\xrightarrow{d}N\left(0,\eta^2\right) \qquad (A6)$$

where $\eta^2=\sigma^2-B^T\Sigma^{-1}\left(I-DV^{-1}D^T\Sigma^{-1}\right)B$, $\sigma^2=p_i\left(1-p_i\right)I\left\{z_i\le z\right\}$, and $B=E\left(\psi\left(z_i,\theta_0\right)I\left\{z_i\le z\right\}\right)$ for some fixed z.

By Taylor expanding $n^{1/2}\left(\widetilde{p}_i-p_i\right)$ about 0 and θ_0, we get:

$$n^{1/2}\left(\widetilde{p}_i-p_i\right)=n^{1/2}\left(\widehat{p}_i-p_i\right)-n^{1/2}B^T\widetilde{\lambda}+o_p(1),$$

where $\widehat{p}_i$ is the non-parametric maximum likelihood estimator of p_i; noting that:

$$n^{-1/2}\begin{bmatrix}\sum_{i=1}^{n}I\left\{z_i\le z\right\}-p_i\\ \psi\left(z_i,\theta_0\right)\end{bmatrix}\xrightarrow{d}N\left(\begin{bmatrix}0\\0\end{bmatrix},\begin{bmatrix}\sigma^2 & B^T\\ B & V^{-1}\end{bmatrix}\right)$$

and recalling $(A5)$ that $n^{1/2}\widetilde{\lambda} = \Sigma^{-1}\left(I - DV^{-1}D^T\Sigma^{-1}\right) n^{-1/2} \sum_{i=1}^{n} \psi(z_i, \theta_0) + o_p(1)$ for $n \to \infty$, we obtain:

$$n^{1/2}(\tilde{p}_i - p_i) = \left[\begin{array}{cc} 1 & -B^T\Sigma^{-1}\left(I - DV^{-1}D^T\Sigma^{-1}\right) \end{array}\right] n^{-1/2} \sum_{i=1}^{n} \left[\begin{array}{cc} I\{z_i \leq z\} & \psi(z_i, \theta_0)^T \end{array}\right] + o_p(1)$$

from which $(A6)$ follows by Slutsky theorem. *QED*

B Arrays of derivatives of W_J

In this appendix, we list the derivatives of the empirical likelihood ratio test as obtained from (11). Noticing that the n $(n = 1, 2, 3, 4)$ number of different block components in the derivatives' arrays of order m $(m = 1, 2, 3)$ is given by:

$$\binom{n+m-1}{n}$$

and recalling that the letters $\alpha, \beta, ...$ and $s, t, ...$ run from 1 to r and $r+1$ to $r+p$, respectively, it is easy to see that:

$$\begin{aligned} k_{abc} &= k_{\alpha\beta\gamma} + [3]\, k_{\alpha\beta s} + [3]\, k_{\alpha st} + k_{stu} \\ k_{abcd} &= k_{\alpha\beta\gamma\delta} + [4]\, k_{\alpha\beta\gamma s} + [6]\, k_{\alpha\beta st} + [4]\, k_{\alpha stu} + k_{stuv} \end{aligned}$$

where the symbol $[j]$ indicates j different permutations of the indices giving rise to the same block of derivatives.

The $(r+p) \times (r+p)$ matrix U_{ab} has block components given by:

$$\begin{aligned} \partial^2_{\lambda^\alpha\lambda^\beta} W &= -\psi^\alpha\psi^\beta / (1 + \lambda^\alpha\psi^\alpha)^2, \qquad \text{(B1)} \\ \partial^2_{\theta^s\lambda^\alpha} W &= \frac{\partial\psi^\alpha}{\partial\theta^s}\left(1 - \lambda^\alpha\psi^\beta / (1 + \lambda^\alpha\psi^\alpha)\right) / (1 + \lambda^\alpha\psi^\alpha), \\ \partial^2_{\theta^s\theta^t} W &= \frac{\partial^2\psi^\alpha}{\partial\theta^s\partial\theta^t}\lambda^\alpha / (1 + \lambda^\alpha\psi^\alpha) - \frac{\partial\psi^\alpha}{\partial\theta^s}\frac{\partial\psi^\beta}{\partial\theta^t}\lambda^\alpha\lambda^\beta / (1 + \lambda^\alpha\psi^\alpha)^2. \end{aligned}$$

The $(r+p)^3$ third order array U_{abc} has block components given by:

$$\partial^3_{\lambda^\alpha\lambda^\beta\lambda^\gamma} W = 2\psi^\alpha\psi^\beta\psi^\gamma / (1 + \lambda^\alpha\psi^\alpha)^3, \qquad \text{(B2)}$$

$$\begin{aligned}
\partial^3_{\lambda^\alpha\lambda^\beta\theta^s} W &= 2\left(-\frac{\partial\psi^\alpha}{\partial\theta^s}\psi^\beta + \frac{\partial\psi^\alpha}{\partial\theta^s}\psi^\alpha\psi^\beta\psi^\gamma / \left(1+\lambda^\alpha\psi^\alpha\right)\right) / \left(1+\lambda^\alpha\psi^\alpha\right)^2, \\
\partial^3_{\lambda^\alpha\theta^s\theta^t} W &= \left(\frac{\partial^2\psi^\alpha}{\partial\theta^s\partial\theta^t} - \left(\frac{\partial^2\psi^\alpha}{\partial\theta^s\partial\theta^t}\psi^\alpha\lambda^\beta + 2\frac{\partial\psi^\alpha}{\partial\theta^s}\frac{\partial\psi^\beta}{\partial\theta^t}\lambda^\beta\right) / \left(1+\lambda^\alpha\psi^\alpha\right) + \right. \\
&\quad \left. 2\frac{\partial\psi^\alpha}{\partial\theta^s}\frac{\partial\psi^\beta}{\partial\theta^t}\psi^\beta\lambda^\alpha\lambda^\beta / \left(1+\lambda^\alpha\psi^\alpha\right)\right) / \left(1+\lambda^\alpha\psi^\alpha\right), \\
\partial^3_{\theta^s\theta^t\theta^u} W &= \left(\frac{\partial^3\psi^\alpha}{\partial\theta^s\partial\theta^t\partial\theta^u}\lambda^\alpha - 3\frac{\partial\psi^\alpha}{\partial\theta^s}\frac{\partial^2\psi^\beta}{\partial\theta^t\partial\theta^u}\lambda^\alpha\lambda^\beta / \left(1+\lambda^\alpha\psi^\alpha\right) + \right. \\
&\quad \left. 2\frac{\partial\psi^\alpha}{\partial\theta^s}\frac{\partial\psi^\beta}{\partial\theta^t}\frac{\partial\psi^\gamma}{\partial\theta^u}\lambda^\alpha\lambda^\beta\lambda^\gamma / \left(1+\lambda^\alpha\psi^\alpha\right)\right) / \left(1+\lambda^\alpha\psi^\alpha\right).
\end{aligned}$$

The $(r+p)^4$ fourth order array U_{abcd} has block components given by:

$$\begin{aligned}
\partial^4_{\lambda^\alpha\lambda^\beta\lambda^\gamma\lambda^\delta} W &= -6\psi^\alpha\psi^\beta\psi^\gamma\psi^\delta / \left(1+\lambda^\alpha\psi^\alpha\right)^4, \qquad \text{(B3)} \\
\partial^4_{\lambda^\alpha\lambda^\beta\lambda^\gamma\theta^s} W &= 6\left(\frac{\partial\psi^\alpha}{\partial\theta^s}\psi^\beta\psi^\gamma - \frac{\partial\psi^\alpha}{\partial\theta^s}\lambda^\alpha\psi^\beta\psi^\gamma\psi^\delta / \left(1+\lambda^\alpha\psi^\alpha\right)^2\right) / \left(1+\lambda^\alpha\psi^\alpha\right)^2, \\
\partial^4_{\lambda^\alpha\lambda^\beta\theta^s\lambda^\gamma} W &= 2\left(2\frac{\partial\psi^\alpha}{\partial\theta^s}\psi^\beta\psi^\gamma + \left(\frac{\partial\psi^\alpha}{\partial\theta^s}\psi^\beta\psi^\gamma - 3\frac{\partial\psi^\alpha}{\partial\theta^s}\psi^\beta\psi^\gamma\psi^\delta\lambda^\alpha\right) / \left(1+\lambda^\alpha\psi^\alpha\right)\right) \times \\
&\quad \left(1+\lambda^\alpha\psi^\alpha\right)^{-3}, \\
\partial^4_{\lambda^\alpha\lambda^\beta\theta^s\theta^t} W &= -2\left(\frac{\partial^2\psi^\alpha}{\partial\theta^s\partial\theta^t}\psi^\beta - \frac{\partial^2\psi^\alpha}{\partial\theta^s\partial\theta^t}\psi^\gamma\psi^\beta\lambda^\gamma / \left(1+\lambda^\alpha\psi^\alpha\right)\right) / \left(1+\lambda^\alpha\psi^\alpha\right)^2 + \\
&\quad 2\frac{\partial\psi^\alpha}{\partial\theta^s}\frac{\partial\psi^\beta}{\partial\theta^t}\left(4\psi^\gamma\lambda^\gamma / \left(1+\lambda^\alpha\psi^\alpha\right) - 3\left(\psi^\gamma\lambda^\gamma\right)^2 / \left(1+\lambda^\alpha\psi^\alpha\right)^2\right) \\
&\quad \left(1+\lambda^\alpha\psi^\alpha\right)^{-2}, \\
\partial^4_{\lambda^\alpha\theta^s\theta^t\theta^u} W &= \frac{\partial^3\psi^\alpha}{\partial\theta^s\partial\theta^t\partial\theta^u}\left(1-\psi^\beta\lambda^\beta / \left(1+\lambda^\alpha\psi^\alpha\right)\right) / \left(1+\lambda^\alpha\psi^\alpha\right) - \\
&\quad 6\frac{\partial\psi^\beta}{\partial\theta^u}\frac{\partial^2\psi^\alpha}{\partial\theta^s\partial\theta^t}\left(\lambda^\beta - \lambda^\beta\lambda^\alpha\psi^\beta / \left(1+\lambda^\alpha\psi^\alpha\right)\right) / \left(1+\lambda^\alpha\psi^\alpha\right)^2 + \\
&\quad 6\frac{\partial\psi^\alpha}{\partial\theta^s}\frac{\partial\psi^\beta}{\partial\theta^t}\frac{\partial\psi^\gamma}{\partial\theta^u}\left(\lambda^\beta\lambda^\gamma - \lambda^\alpha\lambda^\beta\lambda^\gamma\psi^\alpha / \left(1+\lambda^\alpha\psi^\alpha\right)\right) / \left(1+\lambda^\alpha\psi^\alpha\right)^3 \\
\partial^4_{\theta^s\theta^t\theta^u\theta^v} W &= \left(\frac{\partial^4\psi^\alpha}{\partial\theta^s\partial\theta^t\partial\theta^u\partial\theta^v}\lambda^\alpha - 4\frac{\partial^3\psi^\alpha}{\partial\theta^s\partial\theta^t\partial\theta^u}\frac{\partial\psi^\beta}{\partial\theta^v}\lambda^\alpha\lambda^\beta / \left(1+\lambda^\alpha\psi^\alpha\right)\right) \\
&\quad \left(1+\lambda^\alpha\psi^\alpha\right)^{-2} - 3\left(\frac{\partial^2\psi^\alpha}{\partial\theta^s\partial\theta^t}\frac{\partial^2\psi^\beta}{\partial\theta^u\partial\theta^v}\lambda^\alpha\lambda^\beta - \right. \\
&\quad \left. 4\frac{\partial\psi^\alpha}{\partial\theta^s}\frac{\partial^2\psi^\beta}{\partial\theta^t\partial\theta^u}\frac{\partial\psi^\gamma}{\partial\theta^u}\lambda^\alpha\lambda^\beta\lambda^\gamma / \left(1+\lambda^\alpha\psi^\alpha\right)\right) / \left(1+\lambda^\alpha\psi^\alpha\right)^2 -
\end{aligned}$$

$$6\frac{\partial\psi^\alpha}{\partial\theta^s}\frac{\partial\psi^\beta}{\partial\theta^t}\frac{\partial\psi^\gamma}{\partial\theta^u}\frac{\partial\psi^\delta}{\partial\theta^v}\lambda^\alpha\lambda^\beta\lambda^\gamma\lambda^\delta/\left(1+\lambda^\alpha\psi^\alpha\right)^4.$$

Evaluating the above derivatives arrays at $\lambda = 0$ gives the null arrays appearing in (12). Notice that the null derivatives k_{stu} and k_{stuv} are 0.

C Derivation of the stochastic expansion for W_J

Under assumption A8 and A12, by expanding the first order conditions defining $\widetilde{\lambda}$ and $\widetilde{\theta}$ as defined in (6), and inverting this expansion, we get the third order stochastic expansion for $\zeta^a = n^{1/2}\widetilde{\zeta^a}$:

$$\begin{aligned}\zeta^a \;=\; & -\left[U^{a\alpha}U_\alpha\right] - \Big(U^{ab}U_{bcd}\left[U^{c\alpha}U_\alpha\right]\left[U^{d\alpha}U_\alpha\right] + \\ & U^{ab}U_{bcd}\left[U^{c\alpha}U_\alpha\right]U^{de}U_{efg}\left[U^{f\alpha}U_\alpha\right]\left[U^{g\alpha}U_\alpha\right]\Big)/2 + \\ & U^{ab}U_{bcde}\left[U^{c\alpha}U_\alpha\right]\left[U^{d\alpha}U_\alpha\right]\left[U^{e\alpha}U_\alpha\right]/6 + o_p\left(n^{-1}\right)\end{aligned} \qquad \text{(C1)}$$

where U^{ab} is the matrix inverse of U_{ab}.

Next, by Taylor expanding the empirical likelihood ratio W_J under the null hypothesis $\lambda = 0$ about the saddlepoint estimator $\zeta^a = \left[\begin{array}{cc}\widetilde{\lambda^\alpha} & \widetilde{\theta^s}\end{array}\right]$ we obtain:

$$W_J = \widetilde{U}_{ab}\zeta^a\zeta^b - \widetilde{U}_{abc}\zeta^a\zeta^b\zeta^c/3 + \widetilde{U}_{abcd}\zeta^a\zeta^b\zeta^c\zeta^d/12 + o_p\left(n^{-1}\right). \qquad \text{(C2)}$$

Using now assumptions A9-A10, we can then expand the $\widetilde{U}$ arrays under the null hypothesis:

$$\begin{aligned}\widetilde{U}_{ab} &= U_{ab} + U_{abc}\zeta^c + U_{abcd}\zeta^c\zeta^d + o_p\left(n^{-1}\right),\\ \widetilde{U}_{abc} &= U_{abc} + U_{abcd}\zeta^d + o_p\left(n^{-1}\right),\\ \widetilde{U}_{abcd} &= U_{abcd} + o_p\left(n^{-1}\right)\end{aligned}$$

from which, it follows that:

$$W_J = U_{ab}\zeta^a\zeta^b + 2U_{abc}\zeta^a\zeta^b\zeta^c/3 + U_{abcd}\zeta^a\zeta^b\zeta^c\zeta^d/4 + o_p\left(n^{-1}\right). \qquad \text{(C3)}$$

Inserting the stochastic expansion for the saddlepoint estimator $(C1)$ in $(C2)$, we obtain the stochastic expansion for ELJ test for the overidentified model (1) as given in (12).

D Derivation of the CF for W_J

By truncating the sum in (14) at $x + 2y + 3y = 4$, the Edgeworth series distribution is to $o(n^{-1})$:

$$\exp\left\{n^{-1/2}\left(k_{\alpha,bc}\partial_\alpha\partial_{bc} - k_{\alpha,\beta,\gamma}\partial_\alpha\partial_\beta\partial_\gamma/6\right) + n^{-1}\left(k_{\alpha,abc}\partial_\alpha\partial_{abc} + k_{ab,cd}\partial_{ab}\partial_{cd}/2 - k_{\alpha,\beta,ab}\partial_\alpha\partial_\beta\partial_{ab}/2 + k_{\alpha,\beta,\gamma,\delta}\partial_\alpha\partial_\beta\partial_\gamma\partial_\delta/24\right) g\left(U_a, U_{ab}, U_{abc}\right)$$

Applying now the various derivative operators to $g\left(U_a, U_{ab}, U_{abc}\right)$, the full Edgeworth density $h\left(U_\alpha, U_{ab}, U_{abc}\right) = h(\cdot)$ is:

$$\begin{aligned} h(\cdot) = {} & (2\pi)^{-r/2}\left|k^{\alpha,\beta}\right|^{-1/2}\exp\left(-k^{\alpha,\beta}U_\alpha U_\beta/2\right)\left\{1 + n^{-1/2}\left(-k_{\alpha,ab}\left[k^{\alpha,\beta}U_\beta\right]\partial_{ab} + \right.\right. \\ & k_{\alpha,\beta,\gamma}\left(\left[k^{\alpha,\delta}U_\delta\right]\left[k^{\beta,\varepsilon}U_\varepsilon\right]\left[k^{\gamma,\zeta}U_\zeta\right] - 3k^{\alpha,\beta}\left[k^{\gamma,\delta}U_\delta\right]\right)/6\Big) + \qquad\qquad \text{(D1)} \\ & n^{-1}\left(-k_{\alpha,abc}\left[k^{\alpha,\beta}U_\beta\right]\partial_{abc} + k_{ab,cd}\partial_{ab}\partial_{cd}/2 - \right. \\ & k_{\alpha,\beta,ab}\left(\left[k^{\alpha,\gamma}U_\gamma\right]\left[k^{\beta,\delta}U_\delta\right] - k^{\alpha,\beta}\right)\partial_{ab}/2 + k_{\alpha,\beta,\gamma,\delta}\left(\left[k^{\alpha,\varepsilon}U_\varepsilon\right]\left[k^{\beta,\zeta}U_\zeta\right]\times\right. \\ & \left[k^{\gamma,\eta}U_\eta\right]\left[k^{\delta,\vartheta}U_\vartheta\right] - 6k^{\alpha,\beta}\left[k^{\gamma,\eta}U_\eta\right]\left[k^{\delta,\vartheta}U_\vartheta\right] + 3k^{\alpha,\beta}k^{\gamma,\delta}\Big)/24 + \\ & \left(k_{\alpha,ab}k_{\beta,cd}\left(\left[k^{\alpha,\gamma}U_\gamma\right]\left[k^{\beta,\delta}U_\delta\right] - k^{\alpha,\beta}\right)\partial_{ab}\partial_{cd} - k_{\alpha,ab}k_{\beta,\gamma,\delta}\left(\left[k^{\alpha,\varepsilon}U_\varepsilon\right]\left[k^{\beta,\zeta}U_\zeta\right]\times\right.\right. \\ & \left[k^{\gamma,\eta}U_\eta\right]\left[k^{\delta,\vartheta}U_\vartheta\right] - 3k^{\alpha,\beta}\left[k^{\gamma,\eta}U_\eta\right]\left[k^{\delta,\vartheta}U_\vartheta\right] - 3\left[k^{\alpha,\varepsilon}U_\varepsilon\right]k^{\beta,\gamma}\left[k^{\delta,\zeta}U_\zeta\right] + \\ & 3k^{\alpha,\beta}k^{\gamma,\delta}\Big)/\partial_{ab}/3 + k_{\alpha,\beta,\gamma}k_{\delta,\varepsilon,\zeta}\left(\left[k^{\alpha,\eta}U_\eta\right]\left[k^{\beta,\vartheta}U_\vartheta\right]\left[k^{\gamma,\iota}U_\iota\right]\left[k^{\delta,\kappa}U_\kappa\right]\times\right. \\ & \left[k^{\varepsilon,\mu}U_\mu\right]\left[k^{\zeta,\xi}U_\xi\right] - 6k^{\alpha,\beta}\left[k^{\gamma,\iota}U_\iota\right]\left[k^{\delta,\kappa}U_\kappa\right]\left[k^{\varepsilon,\mu}U_\mu\right]\left[k^{\zeta,\xi}U_\xi\right] - \\ & 9k^{\alpha,\delta}\left[k^{\beta,\vartheta}U_\vartheta\right]\left[k^{\gamma,\iota}U_\iota\right]\left[k^{\varepsilon,\mu}U_\mu\right]\left[k^{\zeta,\xi}U_\xi\right] + 9k^{\alpha,\beta}k^{\delta,\varepsilon}\left[k^{\gamma,\iota}U_\iota\right]\left[k^{\zeta,\xi}U_\xi\right] + \\ & 18k^{\alpha,\beta}k^{\gamma,\delta}\left[k^{\varepsilon,\mu}U_\mu\right]\left[k^{\zeta,\xi}U_\xi\right] - 18k^{\alpha,\delta}k^{\beta,\varepsilon}\left[k^{\gamma,\mu}U_\mu\right]\left[k^{\zeta,\xi}U_\xi\right] - 9k^{\alpha,\beta}k^{\gamma,\delta}k^{\varepsilon,\zeta} - \\ & 6k^{\alpha,\delta}k^{\beta,\varepsilon}k^{\gamma,\zeta}\Big)/36\Big)/2\Big\}\prod_{a,b}\delta\left\{U_{ab} - k_{ab}\right\}\prod_{a,b,c}\delta\left\{U_{abc} - n^{-1/2}k_{abc}\right\} + o\left(n^{-1}\right) \end{aligned}$$

To derive the CF of W_J, we apply the Edgeworth density $h(U_\alpha, U_{ab}, U_{abc})$ as derived in $(D1)$ to $\exp(\iota t W_J)$, where $\iota = (-1)^{1/2}$, obtaining:

$$CF_{W_J}(t) = \int \ldots \int \exp(\iota t W_J)\, h(U_\alpha, U_{ab}, U_{abc})\, dU_a dU_{ab} dU_{abc} + o\left(n^{-1}\right).$$

Integrating U_{ab} and U_{abc} out (see Hayakawa (1977, p.365) for details about integrating with respect to derivatives of Dirac δ functions), we obtain the following expansion:

$$\begin{aligned}
CF_{W_J}(t) = {} & (2\pi)^{-r/2}\left|k_{\alpha,\beta}\right|^{-1/2}\exp\left\{-\left(k^{\alpha,\beta}-2\iota t k^{\alpha\beta}\right)U_\alpha U_\beta/2\right\}\times \\
& \left\{1-n^{-1/2}\iota t k_{abc}\left[k^{a\alpha}U_\alpha\right]\left[k^{b\beta}U_\beta\right]\left[k^{c\gamma}U_\gamma\right]/3+\right. \\
& n^{-1}\left(-\iota t k_{abe}k_{cdf}k^{ef}\left[k^{a\alpha}U_\alpha\right]\left[k^{b\beta}U_\beta\right]\left[k^{c\gamma}U_\gamma\right]\left[k^{d\delta}U_\delta\right]/4+\right. \qquad (D2) \\
& \iota t k_{abcd}\left[k^{a\alpha}U_\alpha\right]\left[k^{b\beta}U_\beta\right]\left[k^{c\gamma}U_\gamma\right]\left[k^{d\delta}U_\delta\right]/12- \\
& \left.\left.\left(t^2/2\right)\left(k_{abc}\left[k^{a\alpha}U_\alpha\right]\left[k^{b\beta}U_\beta\right]\left[k^{c\gamma}U_\gamma\right]/3\right)^2\right)\right\}\times \\
& \left\{1+n^{-1/2}\left(\iota t k_{\alpha,ab}\left[k^{\alpha,\beta}U_\beta\right]\left[k^{a\gamma}U_\gamma\right]\left[k^{b\delta}U_\delta\right]+\right.\right. \\
& \left.k_{\alpha,\beta,\gamma}\left(\left[k^{\alpha,\delta}U_\delta\right]\left[k^{\beta,\varepsilon}U_\varepsilon\right]\left[k^{\gamma,\zeta}U_\zeta\right]-3k^{\alpha,\beta}\left[k^{\gamma,\delta}U_\delta\right]\right)\right)/6+ \\
& n^{-1}\left(-\iota t k_{\alpha,ab}k^{ab}k_{cde}\left[k^{\alpha,\beta}U_\beta\right]\left[k^{c\gamma}U_\gamma\right]\left[k^{d\delta}U_\delta\right]\left[k^{e\epsilon}U_\epsilon\right]-\right. \\
& \iota t k_{\alpha,abc}\left[k^{\alpha,\beta}U_\beta\right]\left[k^{a\gamma}U_\gamma\right]\left[k^{b\delta}U_\delta\right]\left[k^{c\epsilon}U_\epsilon\right]/3+ \\
& \iota t k_{ab,cd}\left(\iota t\left[k^{a\alpha}U_\alpha\right]\left[k^{b\beta}U_\beta\right]\left[k^{c\gamma}U_\gamma\right]\left[k^{d\delta}U_\delta\right]+2\left[k^{a\alpha}U_\alpha\right]\left[k^{b\beta}U_\beta\right]k^{cd}\right)/2+ \\
& \iota t k_{\alpha,\beta,ab}\left(\left[k^{\alpha,\gamma}U_\gamma\right]\left[k^{\beta,\delta}U_\delta\right]-k^{\alpha,\beta}\right)\left[k^{a\epsilon}U_\epsilon\right]\left[k^{b\eta}U_\eta\right]/2+ \\
& k_{\alpha,\beta,\gamma,\delta}\left(\left[k^{\alpha,\varepsilon}U_\varepsilon\right]\left[k^{\beta,\zeta}U_\zeta\right]\left[k^{\gamma,\eta}U_\eta\right]\left[k^{\delta,\vartheta}U_\vartheta\right]-6k^{\alpha,\beta}\left[k^{\gamma,\eta}U_\eta\right]\left[k^{\delta,\vartheta}U_\vartheta\right]+\right. \\
& \left.3k^{\alpha,\beta}k^{\gamma,\delta}\right)/24+\iota t k_{\alpha,ab}k_{\beta,cd}\left(\left[k^{\alpha,\varepsilon}U_\epsilon\right]\left[k^{\beta,\zeta}U_\zeta\right]-k^{\alpha,\beta}\right)\times \\
& \left(\iota t\left[k^{a\gamma}U_\gamma\right]\left[k^{b\delta}U_\delta\right]\left[k^{c\eta}U_\eta\right]\left[k^{d\vartheta}U_\vartheta\right]+2\left[k^{a\gamma}U_\gamma\right]\left[k^{b\delta}U_\delta\right]k^{cd}\right)/2+ \\
& \iota t k_{\alpha,ab}k_{\beta,\gamma,\delta}\left(\left[k^{\alpha,\epsilon}U_\epsilon\right]\left[k^{\beta,\zeta}U_\zeta\right]\left[k^{\gamma,\eta}U_\eta\right]\left[k^{\delta,\vartheta}U_\vartheta\right]+\right. \\
& \left.k^{\alpha,\beta}\left[k^{\gamma,\epsilon}U_\epsilon\right]\left[k^{\delta,\eta}U_\eta\right]+\left[k^{\alpha,\epsilon}U_\epsilon\right]k^{\beta,\gamma}\left[k^{\delta,\eta}U_n\right]-k^{\alpha,\beta}k^{\gamma,\delta}\right)\left[k^{a\iota}U_\iota\right]\left[k^{b\kappa}U_\kappa\right]/6+ \\
& k_{\alpha,\beta,\gamma,}k_{\delta,\epsilon,\zeta}\left(\left[k^{\alpha,\eta}U_\eta\right]\left[k^{\beta,\vartheta}U_\vartheta\right]\left[k^{\gamma,\iota}U_\iota\right]\left[k^{\delta,\kappa}U_\kappa\right]\left[k^{\epsilon,\mu}U_\mu\right]\left[k^{\zeta,\nu}U_\nu\right]-\right. \\
& 6k^{\alpha,\beta}\left[k^{\gamma,\iota}U_\iota\right]\left[k^{\delta,\kappa}U_\kappa\right]\left[k^{\epsilon,\mu}U_\mu\right]\left[k^{\zeta,\nu}U_\nu\right]-9k^{\alpha,\delta}\left[k^{\beta,\vartheta}U_\vartheta\right]\left[k^{\gamma,\iota}U_\iota\right]\times \\
& \left[k^{\epsilon,\mu}U_\mu\right]\left[k^{\zeta,\nu}U_\nu\right]+9k^{\alpha,\beta}k^{\delta,\epsilon}\left[k^{\gamma,\iota}U_\iota\right]\left[k^{\zeta,\nu}U_\nu\right]+18k^{\alpha,\beta}k^{\gamma,\delta}\left[k^{\epsilon,\mu}U_\mu\right]\times
\end{aligned}$$

$$\left[k^{\zeta,\nu}U_{\nu}\right] + 18k^{\alpha,\delta}k^{\beta,\epsilon}\left[k^{\gamma,\iota}U_{\iota}\right]\left[k^{\zeta,\nu}U_{\nu}\right]$$
$$-9k^{\alpha,\beta}k^{\gamma,\delta}k^{\epsilon,\zeta} - 6k^{\alpha,\delta}k^{\beta,\epsilon}k^{\gamma,\zeta}\Big)/72\Big\} + o\left(n^{-1}\right)$$

where $k^{\alpha\beta}$ is the $r \times r$ upper left block of the matrix U_{ab} (see $(B1)$) which is different from $k^{\alpha,\beta}$ (i.e. the second Bartlett identity does not hold).

We now integrate with respect to the r vector $U = U_{\alpha}$. As in Hayakawa (1975), we transform the vector U to $z = \Omega^{-1/2}U$, where $\left(\Omega^{-1/2}\right)^{T}\Omega^{-1/2} = \Omega^{-1}$ is the $(r \times r)$ matrix inverse of

$$\begin{aligned}
\Omega &= (1-2\iota t)^{-1}\Sigma\left(I - 2\iota t\Sigma^{-1}DV^{-1}D^{T}\right),\\
\Omega^{-1} &= \Sigma^{-1} - 2\iota t\left(\Sigma^{-1} - \Sigma^{-1}DV^{-1}D^{T}\Sigma^{-1}\right)
\end{aligned}$$

with the Jacobian of the transformation given by:

$$|\Omega|^{1/2} = (1-2\iota t)^{-(r-p)/2}|\Sigma|^{1/2}$$

so that z is $N(0, I)$, and $\left(\Sigma^{-1} - \Sigma^{-1}DV^{-1}D^{T}\Sigma^{-1}\right)\Omega^{1/2}z$ can be regarded as statistically independent of $\left(V^{-1}D^{T}\Sigma^{-1}\right)\Omega^{1/2}z$. Let $\Gamma = \left(\Sigma^{-1} - \Sigma^{-1}DV^{-1}D^{T}\Sigma^{-1}\right)$; we note that:

$$\begin{aligned}
(1-2\iota t)\Sigma^{-1}\Omega\Sigma^{-1} &= \left(\Sigma^{-1} - 2\iota t\Sigma^{-1}DV^{-1}D^{T}\Sigma^{-1}\right) = \left(k^{\alpha,\beta} - 2\iota t\xi^{\alpha\beta}\right) \quad \text{(D3)}\\
(1-2\iota t)\Gamma\Omega\Gamma &= \Gamma = \gamma_{\alpha\beta} = (1-2\iota t)\Sigma^{-1}\Omega\Gamma,\\
-V^{-1}D^{T}\Sigma^{-1}\Omega\Gamma &= 0, \quad V^{-1}D^{T}\Sigma^{-1}\Omega\left(V^{-1}D^{T}\Sigma^{-1}\right)^{T} = V^{-1} = v^{st},\\
\Sigma^{-1}\Omega\left(V^{-1}D^{T}\Sigma^{-1}\right)^{T} &= -\left(V^{-1}D^{T}\Sigma^{-1}\right)^{T} = \phi_{\alpha s}.
\end{aligned}$$

After extremely lengthy algebra, where essentially the arrays k^{ab}, k_{abc} and k_{abcd} are split according to the components given in $(B1)$-$(B3)$, it can be shown that the resulting CF is given by:

$$CF_{W_J}(t) = (1-2\iota t)^{-(r-p)/2}\left(1 + n^{-1}\left(P_0^{'}\iota t + \sum_{j=0}^{3}P_j/(1-2\iota t)^{j}\right)\right) + o\left(n^{-1}\right) \quad \text{(D4)}$$

where the various P's are functions of the cumulants as described in (13) and of the matrices in $(D3)$ and are given by:

$$
\begin{aligned}
P_0' \;=\;& k_{\gamma,\delta,\varepsilon}\left(k_{\alpha\beta s}\xi^{\delta\varepsilon}\gamma_{\alpha\beta}\phi_{\gamma s} + k_{\alpha st}\gamma_{\alpha\delta}\phi_{\gamma s}\phi_{\varepsilon t}\right)/4 - k_{\alpha,\gamma s}k_{\beta,\delta t}\xi^{\alpha\beta}\gamma_{\gamma\delta}\upsilon^{st}/4 - \\
& k_{\alpha s,\beta t}\gamma_{\alpha\beta}\upsilon^{st}/4 + k_{\alpha st}k_{\beta,\gamma u}\gamma_{\alpha\gamma}\left(\phi_{\beta u}\upsilon^{st} + 2\phi_{\beta s}\upsilon^{tu}\right) + \qquad \text{(D5)} \\
& k_{\alpha st}k_{\beta uv}\gamma_{\alpha\beta}\left(\upsilon^{st}\upsilon^{uv} + 2\upsilon^{su}\upsilon^{tv}\right)/16,
\end{aligned}
$$

$$
\begin{aligned}
P_0 \;=\;& k_{\alpha,\gamma\delta}k_{\beta,\varepsilon\zeta}\xi^{\alpha\beta}\left(\gamma_{\gamma\delta}\gamma_{\varepsilon\zeta} + 2\gamma_{\gamma\varepsilon}\gamma_{\delta\zeta}\right)/8 + \left(-k_{\alpha,\gamma\delta}k_{\beta,\varepsilon\zeta}k^{\alpha,\beta}\left(\gamma_{\gamma\delta}\gamma_{\varepsilon\zeta} + 2\gamma_{\gamma\varepsilon}\gamma_{\delta\zeta} - 4k^{\varepsilon\zeta}\gamma_{\gamma\delta}\right) + \right. \\
& 2k_{\alpha,\gamma s}k_{\beta,\delta\varepsilon}\left(3\gamma_{\alpha\gamma}\gamma_{\delta\varepsilon}\phi_{\beta s} + \gamma_{\beta\gamma}\gamma_{\delta\varepsilon}\phi_{\alpha s} - 4k^{\delta\varepsilon}\gamma_{\beta\gamma}\phi_{\alpha s}\right) - 4k_{\alpha,\gamma\delta}k_{\beta,\varepsilon s}k^{\varepsilon s}\xi^{\alpha\beta}\gamma_{\gamma\delta}\Big)/8 - \\
& \left(2k_{\alpha\beta,\gamma\delta}k^{\gamma\delta}\gamma_{\alpha\beta} + 4k_{\alpha\beta,\gamma s}k^{\gamma s}\gamma_{\alpha\beta} + k_{\alpha,\beta,\gamma\delta}\xi^{\alpha\beta}\gamma_{\gamma\delta} - k_{\alpha,\beta,\gamma\delta}k^{\alpha,\beta}\gamma_{\gamma\delta}\right)/4 - \qquad \text{(D6)} \\
& k_{\alpha,\varepsilon s}k_{\beta,\gamma,\delta}\gamma_{\delta\varepsilon}\left(6\xi^{\alpha\beta}\phi_{\gamma s} + k^{\alpha,\beta}\phi_{\gamma s} + k^{\beta,\gamma}\phi_{\alpha s}\right)/3 - k_{\alpha,\varepsilon\zeta}k_{\beta,\gamma,\delta}\gamma_{\varepsilon\zeta}\left(3\xi^{\alpha\beta}\xi^{\gamma\delta} + k^{\alpha,\beta}\xi^{\gamma\delta} - \right. \\
& \left. k^{\alpha,\beta}k^{\gamma,\delta} + k^{\beta,\gamma}\xi^{\alpha\delta}\right)/12 + k_{\alpha,\beta,\gamma}k_{\delta,\varepsilon,\zeta}\left(3\xi^{\alpha\delta}\xi^{\beta\varepsilon}\xi^{\gamma\zeta} + 2\xi^{\alpha\beta}\xi^{\gamma\delta}\xi^{\varepsilon\zeta} - \right. \\
& \left. 2k^{\alpha,\beta}\left(2\xi^{\gamma\varepsilon}\xi^{\delta\zeta} + \xi^{\gamma\delta}\xi^{\varepsilon\zeta}\right)\right)/24 + k_{\alpha,\beta,\gamma}k_{\delta,\varepsilon,\zeta}\left(-2k^{\alpha,\delta}\xi^{\beta\varepsilon}\xi^{\gamma\zeta} - k^{\alpha,\delta}\xi^{\beta\gamma}\xi^{\delta\zeta} + \right. \\
& \left. 2k^{\alpha,\beta}k^{\gamma,\delta}\xi^{\varepsilon\zeta} + k^{\alpha,\beta}k^{\delta,\varepsilon}\xi^{\gamma\zeta}\right)/8 + k_{\alpha,\beta,\gamma,\delta}\left(\xi^{\alpha\beta}\xi^{\gamma\delta} - 2k^{\alpha,\beta}\xi^{\gamma\delta} + k^{\alpha,\beta}k^{\gamma,\delta}\right)/8 + \\
& k_{\delta,\varepsilon,\zeta}k^{\delta,\varepsilon}\left(k_{\alpha\beta\gamma}\left(\gamma_{\alpha\beta}\gamma_{\gamma\zeta} + 2\gamma_{\alpha\gamma}\gamma_{\beta\zeta}\right) - 3k_{\alpha\beta s}\gamma_{\alpha\beta}\phi_{\gamma s} - 3k_{\alpha st}\gamma_{\alpha\zeta}\upsilon^{st}\right)/24 + \\
& \left(k_{\alpha\beta,\gamma\delta}\left(2\gamma_{\alpha\gamma}\gamma_{\beta\delta} + \gamma_{\alpha\beta}\gamma_{\gamma\delta}\right) - k_{\alpha s,\beta t}\gamma_{\alpha\beta}\upsilon^{st}\right)/8 + k_{\alpha,\gamma s}k_{\beta,\delta t}\left(-2\gamma_{\gamma\delta}\phi_{\alpha s}\phi_{\beta t} - \right. \\
& \left. 4k^{\delta t}\gamma_{\beta\gamma}\phi_{\alpha s} + 2\gamma_{\alpha\delta}\gamma_{\beta\gamma}\upsilon^{st} - 2\xi^{\alpha\beta}\gamma_{\gamma\delta}\upsilon^{st} + k^{\alpha,\beta}\gamma_{\gamma\delta}\upsilon^{st}\right)/8 + \\
& \gamma_{\alpha\beta}\left(4k_{\alpha,\beta st}\upsilon^{st} + 9k_{\alpha st}k_{\beta uv}k^{tv}\upsilon^{su}\right)/8 + \left(-3k_{\alpha\beta s}k_{\gamma,\delta\varepsilon}\gamma_{\alpha\beta}\gamma_{\delta\varepsilon}\phi_{\gamma s} - 2k_{\alpha\beta\gamma}k_{\delta,\varepsilon s}\gamma_{\alpha\beta}\gamma_{\gamma\delta}\phi_{\varepsilon s} - \right. \\
& \left. 3k_{\alpha st}k_{\beta,\gamma\delta}\gamma_{\alpha\beta}\gamma_{\gamma\delta}\upsilon^{st} - 6k_{\alpha\beta s}k_{\gamma,\delta t}\gamma_{\alpha\beta}\gamma_{\gamma\delta}\upsilon^{st} + 2k_{\alpha st}k_{\beta,\gamma u}\gamma_{\alpha\gamma}\left(\phi_{\beta u}\upsilon^{st} + 2\phi_{\beta s}\upsilon^{tu}\right)\right)/4 + \\
& \left(4k_{\alpha,\gamma\delta}k_{\beta,\varepsilon\zeta}k^{\alpha,\beta}k^{\varepsilon s}\gamma^{\gamma\delta} + k_{\alpha,\gamma s}k_{\beta,\delta t}\left(-2\phi_{\alpha s}\phi_{\beta t}\gamma_{\gamma\delta} - 4k^{\varepsilon t}\gamma_{\beta\gamma}\phi_{\alpha s} + \upsilon^{st}\times\right.\right. \\
& \left.\left. \left(\gamma_{\alpha\beta}\gamma_{\gamma\delta} + 2\gamma_{\alpha\gamma}\gamma_{\beta\delta}\right)\right)\right)/8 + \left(-24k_{\alpha\beta st}\gamma_{\alpha\beta}\upsilon^{st} + 6k_{\alpha\beta s}k_{\gamma\delta t}\gamma_{\alpha\gamma}\gamma_{\beta\delta}\upsilon^{st} + 3k_{\alpha\beta s}k_{\gamma\delta t}\gamma_{\alpha\beta}\gamma_{\gamma\delta}\upsilon^{st}\right. \\
& \left. +4k_{\alpha\beta\gamma}k_{\delta st}\gamma_{\alpha\gamma}\gamma_{\beta\delta}\upsilon^{st} + 2k_{\alpha\beta\gamma}k_{\delta st}\gamma_{\alpha\beta}\gamma_{\gamma\delta}\upsilon^{st} - 3k_{\alpha st}k_{\beta uv}\gamma_{\alpha\beta}\left(\upsilon^{st}\upsilon^{uv} + 2\upsilon^{su}\upsilon^{tv}\right)\right)/96,
\end{aligned}
$$

$$
P_1 \;=\; k_{\alpha,\gamma\delta}k_{\beta,\varepsilon\zeta}\left(-4\gamma_{\alpha\gamma}\gamma_{\beta\delta}\gamma_{\varepsilon\zeta} - 6\xi^{\alpha\beta}\gamma_{\gamma\varepsilon}\gamma_{\delta\zeta} - 3\xi^{\alpha\beta}\gamma_{\gamma\delta}\gamma_{\varepsilon\zeta} + k^{\alpha,\beta}\left(\gamma_{\gamma\delta}\gamma_{\varepsilon\zeta} + 2\gamma_{\gamma\varepsilon}\gamma_{\delta\zeta}\right)\right)/8 +
$$

$$
\begin{aligned}
&\left(k_{\alpha,\gamma\delta}k_{\beta,\varepsilon\zeta}\left(k^{\alpha,\beta}\left(\gamma_{\gamma\delta}\gamma_{\varepsilon\zeta}+2\gamma_{\gamma\varepsilon}\gamma_{\delta\zeta}-4k^{\varepsilon\zeta}\gamma_{\gamma\delta}\right)-4k^{\varepsilon\zeta}\left(\gamma_{\alpha\gamma}\gamma_{\beta\delta}-\xi^{\alpha\beta}\gamma_{\gamma\delta}\right)\right)-\times\right.\\
&k_{\alpha,\gamma s}k_{\beta,\delta\varepsilon}\left(6\gamma_{\alpha\gamma}\gamma_{\delta\varepsilon}\phi_{\beta_s}+2\gamma_{\beta\gamma}\gamma_{\delta\varepsilon}\phi_{\alpha_s}-4k^{\delta\varepsilon}\gamma_{\beta\gamma}\phi_{\alpha s}\right)+2k_{\alpha,\gamma\delta}k_{\beta,\varepsilon s}\left(4k^{\varepsilon s}\xi^{\alpha\beta}\gamma_{\gamma\delta}+\right. \qquad \text{(D7}\\
&\left.\left.2k^{\alpha,\beta}k^{\varepsilon s}\gamma_{\gamma\delta}\right)\right)/4+\left(k_{\alpha\beta,\gamma\delta}k^{\gamma\delta}\gamma_{\alpha\beta}+2k_{\alpha\beta,\gamma s}k^{\gamma s}\gamma_{\alpha\beta}-k_{\alpha,\beta,\gamma\delta}\left(\gamma_{\alpha\gamma}\gamma_{\beta\delta}-\right.\right.\\
&\left.\left.-\xi^{\alpha\beta}\gamma_{\gamma\delta}+k^{\alpha,\beta}\gamma_{\gamma\delta}\right)\right)/2+k_{\alpha,\varepsilon s}k_{\beta,\gamma,\delta}\left(12\xi^{\alpha\beta}\gamma_{\delta\varepsilon}\phi_{\gamma s}-6k^{\alpha,\beta}\gamma_{\delta\varepsilon}\phi_{\gamma s}+k^{\alpha,\beta}\gamma_{\delta\varepsilon}\phi_{\gamma s}+\right.\\
&\left.k^{\beta,\gamma}\gamma_{\delta\varepsilon}\phi_{\alpha s}\right)/3-k_{\alpha,\varepsilon\zeta}k_{\beta,\gamma,\delta}\left(12\xi^{\gamma\delta}\gamma_{\alpha\varepsilon}\gamma_{\beta\zeta}-9\xi^{\alpha\beta}\xi^{\gamma\delta}\gamma_{\varepsilon\zeta}+2k^{\alpha,\beta}\gamma_{\gamma\varepsilon}\gamma_{\delta\zeta}+\right.\\
&\left.k^{\alpha,\beta}\xi^{\gamma\delta}\gamma_{\varepsilon\zeta}+2k^{\beta,\gamma}\gamma_{\alpha\varepsilon}\gamma_{\delta\zeta}-2k^{\beta,\gamma}\xi^{\alpha\delta}\gamma_{\varepsilon\zeta}+k^{\beta,\gamma}k^{\alpha,\delta}\gamma_{\varepsilon\zeta}+3k^{\beta,\gamma}\xi^{\alpha\beta}\gamma_{\varepsilon\zeta}\right)/12\\
&-k_{\alpha,\beta,\gamma}k_{\delta,\varepsilon,\zeta}\left(9\xi^{\alpha\delta}\xi^{\beta\varepsilon}\xi^{\gamma\zeta}+6\xi^{\alpha\beta}\xi^{\gamma\delta}\xi^{\varepsilon\zeta}-8k^{\alpha,\beta}\xi^{\gamma\varepsilon}\xi^{\delta\zeta}-6k^{\alpha,\beta}\xi^{\gamma\delta}\xi^{\varepsilon\zeta}-3k^{\alpha,\delta}\xi^{\beta\varepsilon}\xi^{\gamma\zeta}-\right.\\
&-3k^{\beta,\varepsilon}\xi^{\alpha\delta}\xi^{\gamma\zeta}-2k^{\gamma,\delta}\xi^{\alpha\beta}\xi^{\varepsilon\zeta}+2k^{\alpha,\beta}k^{\gamma,\delta}\xi^{\varepsilon\zeta}+4k^{\alpha,\beta}k^{\gamma,\varepsilon}\xi^{\delta\zeta}-3k^{\gamma,\zeta}\xi^{\alpha\delta}\xi^{\gamma\zeta}+4k^{\alpha,\beta}k^{\delta,\zeta}\xi^{\gamma\varepsilon}+\\
&\left.2k^{\varepsilon,\zeta}\xi^{\alpha\beta}\xi^{\gamma\delta}+2k^{\alpha,\beta}k^{\varepsilon,\zeta}\xi^{\gamma\delta}\right)/24+\left(-k_{\alpha,\beta,\gamma}k_{\delta,\varepsilon,\zeta}\left(-4k^{\alpha,\delta}\xi^{\beta\varepsilon}\xi^{\gamma\zeta}-2k^{\alpha,\delta}\xi^{\beta\gamma}\xi^{\varepsilon\zeta}+\right.\right.\\
&\left.k^{\alpha,\delta}k^{\beta,\gamma}\xi^{\varepsilon\zeta}+2k^{\alpha,\delta}k^{\beta,\varepsilon}\xi^{\gamma\zeta}+2k^{\alpha,\beta}k^{\gamma,\delta}\xi^{\varepsilon\zeta}+2k^{\alpha,\delta}k^{\gamma,\zeta}\xi^{\beta\varepsilon}+k^{\alpha,\beta}k^{\delta,\varepsilon}\xi^{\gamma\zeta}-k^{\alpha,\beta}k^{\gamma,\zeta}k^{\delta,\varepsilon}\right)+\\
&\left.k_{\alpha,\beta,\gamma,\delta}\left(-2\xi^{\alpha\beta}\xi^{\gamma\delta}+3k^{\alpha,\beta}\xi^{\gamma\delta}+k^{\gamma,\delta}\xi^{\alpha\beta}-2k^{\alpha,\beta}k^{\gamma,\delta}\right)\right)/8-k_{\alpha,\gamma\delta}k_{\beta,\varepsilon s}k^{\alpha,\beta}k^{\varepsilon s}\gamma_{\gamma\delta}/2\\
&-k_{\alpha s,\beta t}\gamma_{\alpha\beta}\upsilon^{st}/8+\left(k_{\alpha,\gamma s}k_{\beta,\delta t}\left(\phi_{\alpha s}\phi_{\beta t}\gamma_{\gamma\delta}+2k^{\delta t}\gamma_{\alpha\gamma}\gamma_{\beta\delta}+\upsilon^{st}\left(\gamma_{\alpha\beta}\gamma_{\gamma\delta}+2\gamma_{\alpha\gamma}\gamma_{\beta\delta}\right)\right)-\right.\\
&\left.k_{\alpha\beta,\gamma\delta}\left(\gamma_{\alpha\beta}\gamma_{\gamma\delta}+2\gamma_{\alpha\gamma}\gamma_{\beta\delta}\right)\right)/4+k_{\alpha,\gamma s}k_{\beta,\delta t}\left(2\gamma_{\gamma\delta}\phi_{\alpha s}\phi_{\beta t}+4k^{\delta t}\gamma_{\beta\gamma}\phi_{\alpha s}\right.\\
&\left.-4\gamma_{\alpha\delta}\gamma_{\beta\gamma}\upsilon^{st}+\gamma_{\gamma\delta}\upsilon^{st}\left(3\xi^{\alpha\beta}-2k^{\alpha,\beta}\right)\right)/8+k_{\delta,\varepsilon,\zeta}\left(15k_{\alpha\beta\gamma}\gamma_{\alpha\beta}\gamma_{\gamma\delta}\xi^{\varepsilon\zeta}+\right.\\
&k_{\alpha\beta s}\left(12\gamma_{\alpha\delta}\gamma_{\beta\varepsilon}\phi_{\zeta s}-9\gamma_{\alpha\beta}\xi^{\varepsilon\zeta}\phi_{\delta s}+6k^{\varepsilon,\zeta}\gamma_{\alpha\beta}\phi_{\delta s}\right)-6k_{\alpha st}\left(\gamma_{\alpha\varepsilon}\phi_{\delta s}\phi_{\varepsilon t}+\gamma_{\alpha\delta}\gamma_{\varepsilon\zeta}\upsilon^{st}\right)\\
&\left.+5k^{\varepsilon,\zeta}k_{\alpha\beta\gamma}\gamma_{\alpha\beta}\gamma_{\gamma\delta}+k_{\delta,\varepsilon,\zeta}k^{\delta,\varepsilon}\left(-6k_{\alpha,\beta,\gamma}\gamma_{\alpha\beta}\gamma_{\gamma\delta}+3k_{\alpha\beta s}\gamma_{\alpha\beta}\phi_{\gamma s}+3k_{\alpha st}\gamma_{\alpha\zeta}\upsilon^{st}\right)\right)/24+\\
&\left(\left(\gamma_{\alpha\beta}\gamma_{\delta\varepsilon}+2\gamma_{\alpha\delta}\gamma_{\beta\varepsilon}\right)\left(18k_{\alpha\beta\gamma}k_{\delta\varepsilon s}k^{\gamma s}+3k_{\alpha\beta\varepsilon}k_{\gamma\delta\zeta}k^{\varepsilon\zeta}+27k_{\alpha\beta s}k_{\gamma\delta t}k^{st}+4k_{\alpha,\beta\gamma\delta}\right)\right.\\
&\left.-12k_{\alpha,\beta st}\gamma_{\alpha\beta}\upsilon^{st}-27k_{\alpha st}k_{\beta uv}k^{tv}\gamma_{\alpha\beta}\upsilon^{su}\right)/24+\left(-k_{\alpha\beta\gamma}k_{\delta,\varepsilon\zeta}\left(4\gamma_{\alpha\delta}\gamma_{\beta\varepsilon}\gamma_{\gamma\zeta}+\gamma_{\alpha\beta}\gamma_{\gamma\delta}\gamma_{\varepsilon\zeta}\right)+\right.\\
&+6k_{\alpha\beta s}k_{\gamma,\delta\varepsilon}\gamma_{\alpha\beta}\gamma_{\delta\varepsilon}\phi_{\gamma s}+4k_{\alpha\beta\gamma}k_{\delta,\varepsilon s}\gamma_{\alpha\beta}\gamma_{\gamma\delta}\phi_{\varepsilon s}+6k_{\alpha st}k_{\beta,\gamma\delta}\gamma_{\alpha\beta}\gamma_{\gamma\delta}\upsilon^{st}+\\
&\left.+12k_{\alpha\beta s}k_{\gamma,\delta t}\gamma_{\alpha\beta}\gamma_{\gamma\delta}\upsilon^{st}-2k_{\alpha st}k_{\beta,\gamma u}\gamma_{\alpha\gamma}\left(\phi_{\beta u}\upsilon^{st}+2\phi_{\beta s}\upsilon^{tu}\right)\right)/4+\\
&\left(12k_{\alpha\beta\gamma\delta}\gamma_{\alpha\beta}\gamma_{\gamma\delta}+k_{\alpha\beta\gamma}k_{\delta\varepsilon\zeta}\left(2\gamma_{\alpha\delta}\gamma_{\beta\varepsilon}\gamma_{\gamma\zeta}+3\gamma_{\alpha\beta}\gamma_{\gamma\delta}\gamma_{\varepsilon\zeta}\right)+24k_{\alpha\beta st}\gamma_{\alpha\beta}\upsilon^{st}\right.\\
&-6k_{\alpha\beta s}k_{\gamma\delta t}\upsilon^{st}\left(\gamma_{\alpha\beta}\gamma_{\gamma\delta}+2\gamma_{\alpha\gamma}\gamma_{\beta\delta}\right)-4k_{\alpha\beta\gamma}k_{\delta st}\upsilon^{st}\left(\gamma_{\alpha\beta}\gamma_{\gamma\delta}+2\gamma_{\alpha\gamma}\gamma_{\beta\delta}\right)\\
&\left.\left.+3k_{\alpha st}k_{\beta uv}\gamma_{\alpha\beta}\left(\upsilon^{st}\upsilon^{uv}+2\upsilon^{su}\upsilon^{tv}\right)\right)\right)/96
\end{aligned}
$$

$$
P_2 = k_{\alpha\beta,\gamma\delta}\left(\gamma_{\alpha\beta}\gamma_{\gamma\delta}+2\gamma_{\alpha\gamma}\gamma_{\beta\delta}\right)/8+\left(-18k_{\alpha\beta\gamma}k_{\delta\varepsilon s}k^{\gamma s}\left(\gamma_{\alpha\beta}\gamma_{\delta\varepsilon}+2\gamma_{\alpha\delta}\gamma_{\beta\varepsilon}\right)-\right.
$$

$$
\begin{aligned}
&\left(\gamma_{\alpha\beta}\gamma_{\gamma\delta}+2\gamma_{\alpha\gamma}\gamma_{\beta\delta}\right)\left(3k_{\alpha\beta\varepsilon}k_{\gamma\delta\zeta}k^{\varepsilon\zeta}+27k_{\alpha\beta s}k_{\gamma\delta t}v^{st}+4k_{\alpha,\beta\gamma\delta}\right)\Big)/24\\
&+k_{\alpha,\beta,\gamma\delta}\left(k^{\alpha,\beta}\gamma_{\gamma\delta}-\xi^{\alpha\beta}\gamma_{\gamma\delta}+2\gamma_{\alpha\beta}\gamma_{\gamma\delta}\right)/4+ \qquad \text{(D8)}\\
&k_{\alpha,\gamma\delta}k_{\beta,\varepsilon\zeta}\left(-24\gamma_{\alpha\delta}\gamma_{\beta\varepsilon}\gamma_{\gamma\zeta}+\left(\gamma_{\gamma\delta}\gamma_{\varepsilon\zeta}+2\gamma_{\gamma\varepsilon}\gamma_{\delta\zeta}\right)\left(3\xi^{\alpha\beta}+k^{\alpha,\beta}\right)\right)/8+\\
&4k_{\alpha,\gamma\delta}k_{\beta,\varepsilon\zeta}\left(k^{\varepsilon\zeta}\left(2\gamma_{\alpha\gamma}\gamma_{\beta\delta}-\xi^{\alpha\beta}\gamma_{\gamma\delta}\right)-k^{\alpha,\beta}\left(\gamma_{\gamma\delta}\gamma_{\varepsilon\zeta}+2\gamma_{\gamma\varepsilon}\gamma_{\delta\zeta}-k^{\varepsilon\zeta}\gamma_{\gamma\delta}\right)\right)\\
&+2k_{\alpha,\gamma s}k_{\beta,\delta\varepsilon}\left(\phi_{\alpha s}\gamma_{\beta\gamma}\gamma_{\delta\varepsilon}+3\phi_{\beta s}\gamma_{\alpha\gamma}\gamma_{\delta\varepsilon}\right)+4k_{\alpha,\gamma\delta}k_{\beta,\varepsilon s}k^{\varepsilon s}\times\\
&\left(2\gamma_{\alpha\gamma}\gamma_{\beta\delta}-\gamma_{\gamma\delta}\left(\xi^{\alpha\beta}-k^{\alpha,\beta}\right)\right)\Big)/8+2k_{\alpha,\varepsilon s}k_{\beta,\gamma,\delta}\gamma_{\delta\varepsilon}\phi_{\gamma s}\left(k^{\alpha,\beta}-\xi^{\alpha\beta}\right)+\\
&k_{\alpha,\varepsilon\zeta}k_{\beta,\gamma,\delta}\left(24\xi^{\gamma\delta}\gamma_{\alpha\varepsilon}\gamma_{\beta\zeta}-9\xi^{\alpha\beta}\xi^{\gamma\delta}\gamma_{\varepsilon\zeta}+2k^{\alpha,\beta}\gamma_{\gamma\varepsilon}\gamma_{\delta\zeta}+5k^{\alpha,\beta}\xi^{\gamma\delta}\gamma_{\varepsilon\zeta}+\right.\\
&\left.k^{\beta,\gamma}\gamma_{\delta\zeta}\left(2\gamma_{\alpha\varepsilon}+\xi^{\alpha\varepsilon}+k^{\alpha,\varepsilon}\right)-2k^{\gamma,\delta}\left(6\gamma_{\alpha\varepsilon}\gamma_{\beta\zeta}-3\xi^{\alpha\beta}\gamma_{\varepsilon\zeta}+k^{\alpha,\beta}\gamma_{\varepsilon\zeta}\right)\right)/12+\\
&k_{\alpha\beta\gamma}k_{\delta,\varepsilon,\zeta}k^{\delta,\varepsilon}\left(\gamma_{\alpha\beta}\gamma_{\gamma\zeta}+2\gamma_{\alpha\gamma}\gamma_{\beta\zeta}\right)/24+k_{\alpha,\beta,\gamma}k_{\delta,\varepsilon,\zeta}k^{\alpha,\delta}\left(-\xi^{\beta\gamma}\xi^{\varepsilon\zeta}-2\xi^{\beta\varepsilon}\xi^{\gamma\zeta}+\right.\\
&\left.k^{\beta,\gamma}\xi^{\varepsilon\zeta}+2k^{\beta,\varepsilon}\xi^{\gamma\zeta}+2k^{\gamma,\zeta}\xi^{\beta\varepsilon}-2k^{\beta,\varepsilon}k^{\gamma,\zeta}+k^{\varepsilon,\zeta}\xi^{\beta\gamma}-k^{\beta,\gamma}k^{\varepsilon,\zeta}\right)/8\\
&+k_{\alpha,\beta,\gamma}k_{\delta,\varepsilon,\zeta}\left(9\xi^{\alpha\delta}\xi^{\beta\varepsilon}\xi^{\gamma\zeta}+6\xi^{\alpha\beta}\xi^{\gamma\delta}\xi^{\varepsilon\zeta}-4k^{\alpha,\beta}\xi^{\gamma\varepsilon}\xi^{\delta\zeta}-6k^{\alpha,\beta}\xi^{\gamma\delta}\xi^{\varepsilon\zeta}-6k^{\alpha,\delta}\xi^{\beta\varepsilon}\xi^{\gamma\zeta}\right.\\
&-6k^{\beta,\varepsilon}\xi^{\alpha\delta}\xi^{\gamma\delta}+3k^{\alpha,\delta}k^{\beta,\varepsilon}\xi^{\gamma\zeta}-4k^{\gamma,\delta}\xi^{\alpha\beta}\xi^{\varepsilon\zeta}+4k^{\alpha,\beta}k^{\gamma,\delta}\xi^{\varepsilon\zeta}+4k^{\alpha,\beta}k^{\gamma,\varepsilon}\xi^{\delta\zeta}-\\
&-6k^{\gamma,\zeta}\xi^{\alpha\delta}\xi^{\beta\varepsilon}+3k^{\alpha,\delta}k^{\gamma,\zeta}\xi^{\beta\varepsilon}+3k^{\beta,\varepsilon}k^{\gamma,\zeta}\xi^{\alpha\delta}+4k^{\alpha,\beta}k^{\delta,\zeta}\xi^{\gamma\varepsilon}-4k^{\alpha,\beta}k^{\gamma,\varepsilon}k^{\delta,\zeta}-\\
&\left.4k^{\varepsilon,\zeta}\xi^{\alpha\beta}\xi^{\gamma\delta}+4k^{\alpha,\beta}k^{\varepsilon,\zeta}\xi^{\gamma\delta}+2k^{\gamma,\delta}k^{\varepsilon,\zeta}\xi^{\alpha\beta}-2k^{\alpha,\beta}k^{\gamma,\delta}k^{\varepsilon,\zeta}\right)/24+\\
&k_{\alpha,\beta,\gamma,\delta}\left(\xi^{\alpha\beta}\xi^{\gamma\delta}-k^{\alpha,\beta}\xi^{\gamma\delta}-k^{\gamma,\delta}\xi^{\alpha\beta}+k^{\alpha,\beta}k^{\gamma,\delta}\right)/8+\left(12k_{\alpha\beta\gamma\delta}\gamma_{\alpha\beta}\gamma_{\gamma\delta}-\right.\\
&2k_{\alpha\beta\gamma}k_{\delta\varepsilon\zeta}\left(2\gamma_{\alpha\delta}\gamma_{\beta\varepsilon}\gamma_{\gamma\zeta}+3\gamma_{\alpha\beta}\gamma_{\gamma\delta}\gamma_{\varepsilon\zeta}\right)+3k_{\alpha\beta s}k_{\gamma\delta t}v^{st}\left(\gamma_{\alpha\beta}\gamma_{\gamma\delta}+2\gamma_{\alpha\gamma}\gamma_{\beta\delta}\right)+\\
&\left.k_{\alpha\beta\gamma}k_{\delta st}v^{st}\left(\gamma_{\alpha\beta}\gamma_{\gamma\delta}+2\gamma_{\alpha\gamma}\gamma_{\beta\delta}\right)\right)/96+\\
&k_{\alpha,\gamma s}k_{\beta,\delta t}v^{st}\left(2\gamma_{\alpha\gamma}\gamma_{\beta\delta}+\gamma_{\alpha\beta}\gamma_{\gamma\delta}-\xi^{\alpha\beta}\gamma_{\gamma\delta}+k^{\alpha,\beta}\gamma_{\gamma\delta}\right)/8+\\
&k_{\delta,\varepsilon,\zeta}\left(5k_{\alpha\beta\gamma}\gamma_{\alpha\beta}\gamma_{\gamma\delta}\left(-3\xi^{\varepsilon\zeta}+2k^{\varepsilon,\zeta}\right)+3k_{\alpha\beta s}\left(-2\gamma_{\alpha\delta}\gamma_{\beta\varepsilon}\phi_{\zeta s}+\gamma_{\alpha\beta}\xi^{\varepsilon\zeta}\phi_{\delta s}\right.\right.\\
&\left.\left.-k^{\varepsilon,\zeta}\gamma_{\alpha\beta}\phi_{\delta s}\right)\right)/24+\left(2k_{\alpha\beta\gamma}k_{\delta,\varepsilon\zeta}\left(\gamma_{\alpha\beta}\gamma_{\gamma\delta}\gamma_{\varepsilon\zeta}+4\gamma_{\alpha\delta}\gamma_{\beta\varepsilon}\gamma_{\gamma\zeta}\right)\right.\\
&-3k_{\alpha\beta s}\gamma_{\alpha\beta}\left(k_{\delta,\varepsilon\zeta}\phi_{\delta s}\gamma_{\varepsilon\zeta}+2k_{\delta,\varepsilon t}v^{st}\gamma_{\delta\varepsilon}\right)-\\
&\left.2k_{\alpha\beta\gamma}k_{\delta,\varepsilon s}\phi_{\varepsilon s}\gamma_{\alpha\beta}\gamma_{\gamma\delta}-3k_{\alpha st}k_{\gamma,\varepsilon\zeta}v^{st}\gamma_{\alpha\beta}\gamma_{\gamma\varepsilon}\right)/4,
\end{aligned}
$$

$$
\begin{aligned}
P_3 \;=\;& k_{\alpha\beta\gamma}k_{\delta\varepsilon\zeta}\left(3\gamma_{\alpha\beta}\gamma_{\gamma\delta}\gamma_{\varepsilon\zeta}+2\gamma_{\alpha\delta}\gamma_{\beta\varepsilon}\gamma_{\gamma\zeta}\right)/96+k_{\alpha,\gamma\delta}k_{\beta,\varepsilon\zeta}\times\\
&\left(12\gamma_{\alpha\gamma}\gamma_{\beta\delta}\gamma_{\varepsilon\zeta}-\xi^{\alpha\beta}\left(\gamma_{\gamma\delta}\gamma_{\varepsilon\zeta}+2\gamma_{\gamma\varepsilon}\gamma_{\delta\zeta}\right)+k^{\alpha,\beta}\left(\gamma_{\gamma\delta}\gamma_{\varepsilon\zeta}+2\gamma_{\gamma\varepsilon}\gamma_{\delta\zeta}\right)\right)/8- \qquad \text{(D9)}
\end{aligned}
$$

$$k_{\alpha\beta\gamma}k_{\delta,\varepsilon\zeta}\left(\gamma_{\alpha\beta}\gamma_{\gamma\delta}\gamma_{\varepsilon\zeta}+4\gamma_{\alpha\delta}\gamma_{\beta\varepsilon}\gamma_{\gamma\zeta}\right)/4+k_{\alpha,\varepsilon\zeta}k_{\beta,\gamma,\delta}\Big(-4\gamma_{\alpha\varepsilon}\gamma_{\beta\zeta}\xi^{\gamma\delta}+\xi^{\alpha\beta}\xi^{\gamma\delta}\gamma_{\varepsilon\zeta}-$$

$$-k^{\alpha,\beta}\xi^{\gamma\delta}\gamma_{\varepsilon\zeta}+4k^{\gamma,\delta}\gamma_{\alpha\varepsilon}\gamma_{\beta\zeta}-k^{\gamma,\delta}\xi^{\alpha\beta}\gamma_{\varepsilon\zeta}+k^{\alpha,\beta}k^{\gamma,\delta}\gamma_{\varepsilon\zeta}\Big)/4+$$

$$5k_{\alpha\beta\gamma}k_{\delta,\varepsilon,\zeta}\left(\xi^{\varepsilon\zeta}\gamma_{\alpha\beta}\gamma_{\gamma\delta}-k^{\varepsilon,\zeta}\gamma_{\alpha\beta}\gamma_{\gamma\delta}\right)/24+k_{\alpha,\beta,\gamma}k_{\delta,\varepsilon,\zeta}\times\Big(-3\xi^{\alpha\delta}\xi^{\beta\varepsilon}\xi^{\gamma\zeta}-$$

$$2\xi^{\alpha\beta}\xi^{\gamma\delta}\xi^{\varepsilon\zeta}+2k^{\alpha,\beta}\xi^{\gamma\delta}\xi^{\varepsilon\zeta}+3k^{\alpha,\delta}\xi^{\beta\varepsilon}\xi^{\gamma\zeta}+3k^{\beta,\varepsilon}\xi^{\alpha\delta}\xi^{\gamma\zeta}-3k^{\alpha,\delta}k^{\beta,\varepsilon}\xi^{\gamma\zeta}+$$

$$2k^{\gamma,\delta}\xi^{\alpha\beta}\xi^{\varepsilon\zeta}-2k^{\alpha,\beta}k^{\gamma,\delta}\xi^{\varepsilon\zeta}+3k^{\gamma,\zeta}\xi^{\alpha\delta}\xi^{\beta\varepsilon}-3k^{\alpha,\delta}k^{\gamma,\zeta}\xi^{\beta\varepsilon}-3k^{\beta,\varepsilon}k^{\gamma,\zeta}\xi^{\alpha\delta}+$$

$$3k^{\alpha,\delta}k^{\beta,\varepsilon}k^{\gamma,\zeta}+2k^{\varepsilon,\zeta}\xi^{\alpha\beta}\xi^{\gamma\delta}-2k^{\alpha,\beta}k^{\varepsilon,\zeta}\xi^{\gamma\delta}-2k^{\gamma,\delta}k^{\varepsilon,\zeta}\xi^{\alpha\beta}+2k^{\alpha,\beta}k^{\gamma,\delta}k^{\varepsilon,\zeta}\Big)/24.$$

E Proof of Theorem 3

The proof is standard; from the two asymptotic expansion (17) and (20), it is easy to see that

$$\Pr\{W_J \geq x\} - \Pr{}^*\{W_J^* \geq x\} = o\left(n^{-1}\right)$$

uniformly over x; as P_0' and P_i $(i = 0, ..., 3)$ differ from their bootstrap analogues by $o_p(1)$, equation (22) follows by replacing x by $x^*_{cv_\alpha}$. *QED*

References

BACK, K., and D. BROWN (1993): "Implied Probabilities in GMM Estimation," *Econometrica*, 61, 971–975.

BARNDORFF-NIELSEN, O., and P. HALL (1988): "On the Level-Error After Bartlett Adjustment of the Likelihood Ratio Statistics," *Biometrika*, 75, 374–378.

BHATTACHARYA, R., and J. GHOSH (1978): "On the Validity of Formal Edgeworth Expansion," *Annals of Statistics*, 6, 434–451.

BRAVO, F. (1999): "Empirical Likelihood Based Inference with Applications to some Econometric Models," Submitted for publication.

BROWN, B., and W. NEWEY (1995): "Bootstrapping for GMM," Unpublished manuscript, MIT.

CHAMBERLAIN, G. (1987): "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions," *Journal of Econometrics*, 34, 305–334.

CHANDRA, T. (1985): "Asymptotic Expansion of Perturbed Chi-Square Variables," *Sankhyā A*, 47, 100–110.

CHANDRA, T., and J. GHOSH (1979): "Valid Asymptotic Expansions for the Likelihood Ratio and Other Perturbed Chi-Square Variables," *Sankhyā A*, 41, 22–47.

CORCORAN, S. A. (1998): "Bartlett Adjustments of Empirical Discrepancy Statistics," *Biometrika*, 85, 967–972.

GÖTZE, F., and C. HIPP (1983): "Asymptotic Expansion for Sums of Weakly Dependent Random Vectors," *Zeitschrift für wahrscheinlichkeiteorie und verwandte gebiete*, 64, 211–239.

——— (1994): "Asymptotic Distribution of Statistics in Time Series," *Annals of statistics*, 22, 2062–2088.

HALL, P. (1986): "On the Number of Bootstrap Simulations Required to Construct a Confidence Interval," *Annals of Statistics*, 14, 1453–1462.

HALL, P., and J. HOROWITZ (1996): "Bootstrap Critical Values for Tests Based on Generalized-Method of Moment Estimators," *Econometrica*, 64, 891–916.

HALL, P., and B. PRESNELL (1999): "Intentionally Biased Bootstrap Methods," *Journal of the Royal Statistical Society B*, 61, 143–158.

HANSEN, L. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–1054.

HANSEN, L., J. HEATON, and A. YARON (1996): "Finite Sample Properties of some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262–280.

HAYAKAWA, T. (1975): "The Likelihood Ratio Criterion for a Composite Hypothesis under a Local Alternative," *Biometrika*, 62, 451–460.

——— (1977): "The Likelihood Ratio Criterion and the Asymptotic Expansion of its Distribution," *Annals of the Institute of Statistical Mathematics*, 29, 359–378.

IMBENS, G., R. SPADY, and P. JOHNSON (1998): "Information Theoretic Approaches to Inference in Moment Condition Models," *Econometrica*, 66, 333–357.

KITAMURA, Y. (1997): "Empirical Likelihood Methods with Weakly Dependent Processes," *Annals of Statistics*, 25, 2084–2102.

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

Mykland, P. (1994): "Bartlett Type of Identities," *Annals of Statistics*, 22, 21–38.

——— (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.

——— (1990): "Empirical Likelihood Ratio Confidence Regions," *Annals of Statistics*, 18, 90–120.

Peers, H. (1971): "Likelihood Ratio and Associated Test Criteria," *Biometrika*, 58, 577–587.

Phillips, P., and J. Park (1988): "On Formulation of Wald Tests of Nonlinear Restrictions," *Econometrica*, 56, 1065–1084.

Qin, J., and J. Lawless (1994): "Empirical Likelihood and General Estimating Equations," *Annals of Statistics*, 22, 300–325.

Rao, C. (1973): *Linear Statistical Inference and Its Applications*. New York: Wiley, 2nd edn.

Sargan, J. (1980): "Some Approximations to the Distribution of Econometric Criteria Which are Asymptotically Distributed as Chi-Squared," *Econometerica*, 48, 1107–1138.

Smith, R., A. Chesher, and S. Peters (1999): "Asymptotic Approximations to the Distribution of Conditional Moment Test Statistics in Alikelihood Framework," Mimeo, University of Bristol.

Stein, C. (1956): "Efficient Nonparametric Testing and Estimation," in *Proceedings of Third Berkley Symposium in Mathematical Statistics and Probability*, ed. by J. Neyman, vol. 1, pp. 187–195. University of California Press, Berkeley.

White, H. (1982): "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50, 1–25.

Footnotes

1 That is to say, the information of the resulting parametric subfamily at the true unknown distribution is no greater than for the full original nonparametric problem.

2 It is not difficult to see that (under the null hypothesis) -see also Appendix B- the so called signed square root decomposition of W_J (cf. eq. (12) is given by:

$$U_a - U_{a\alpha\beta}\left[U^{\alpha\gamma}U_{\gamma}\right]\left[U^{\beta\delta}U_{\delta}\right]/3 - 2U_{a\alpha s}\left[U^{\alpha\gamma}U_{\gamma}\right]\left[U^{s\delta}U_{\delta}\right]/3 + o_p\left(n^{-1/2}\right),$$

from which Sargan's (1980) Lemma 1 follows with $k = 2$, as $U_s = 0$.
Notice that we do not require the existence of all moments of U_s, but we assume that the Cramér condition holds jointly for the first 4 arrays of derivatives.

3 Actually the coverage error can be shown to be $O\left(n^{-2}\right)$ by strengthening the regularity conditions so that there exists an Edgeworth expansion of the J statistic with error $O\left(n^{-3/2}\right)$ and exploiting the even-odd property of the Hermite polynomials (tensors) appearing in it (Barndorff-Nielsen and Hall (1988)).

4 We set the number B of bootstrap replications to 200; increasing the number of replications did not change significantly the results (for an explanation of the phenomenon, see Hall (1986)).

5 We use $\exp\{\max(z)\}$ rather than an (inefficient) $n^{1/2}$ consistent GMM, because in the case of the bootstrap the maximum empirical likelihood estimator $\widetilde{\theta}$ would be inconsistent.

Tables and figures

TABLE I

EMPIRICAL LEVELS OF J TESTS FOR MODEL

(10) SAMPLE SIZE $n = 50$

	$z \sim N(1,2)$			$z \sim \chi^2(1)$		
size	W_J	J_{IT}	J_{CU}	W_J	J_{IT}	J_{CU}
0.10	0.126	0.131	0.130	0.261	0.299	0.288
0.05	0.060	0.065	0.060	0.183	0.213	0.202
0.01	0.029	0.030	0.029	0.129	0.152	0.143

W_J empirical likelihood ratio J test, J_I and J_{CU} J tests based on iterated and continuously updated GMM estimators, respectively (cf. Hansen, Heaton, and Yaron (1996)).

TABLE II

EMPIRICAL LEVELS OF J TESTS FOR MODEL

(10) SAMPLE SIZE $n = 100$

	$z \sim N(1,2)$			$z \sim \chi^2(1)$		
size	W_J	J_{IT}	J_{CU}	W_J	J_{IT}	J_{CU}
0.10	0.116	0.121	0.120	0.181	0.209	0.208
0.05	0.058	0.061	0.060	0.143	0.162	0.154
0.01	0.021	0.025	0.024	0.094	0.112	0.109

W_J empirical likelihood ratio J test, J_I and J_{CU} J tests based on iterated and continuously updated GMM estimators, respectively (cf. Hansen, Heaton, and Yaron (1996)).

TABLE III

EMPIRICAL LEVELS OF J TEST FOR MODEL (23)

size	$n = 50$			$n = 100$		
0.10	0.171^a (0.228)	0.132^b	0.124^c (0.197)	0.165^a (0.194)	0.123^b	$\underline{0.120}^c$ (0.164)
0.05	0.142^a (0.171)	0.090^b	0.080^c (0.149)	0.113^a (0.132)	0.082^b	$\underline{0.065}^c$ (0.113)
0.01	0.101^a (0.102)	0.048^b	0.048^c (0.095)	0.072^a (0.067)	$\underline{0.032}^b$	$\underline{0.031}^c$ (0.063)

a asymptotic, *b* Edgeworth corrected, *c* bootstrapped critical values. The values reported in parenthesis are those reported i Hall and Horowitz (1996). Underlined values indicate a level not significantly different from nominal level at 0.05 level.

Figure 1: Densities of ELJ test

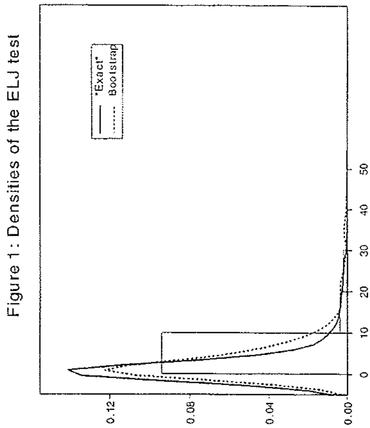

Figure 2: Densities of $\widetilde{\theta}$