EMPIRICAL LIKELIHOOD BASED INFERENCE WITH APPLICATIONS TO SOME ECONOMETRIC MODELS

FRANCESCO BRAVO
University of York

This paper uses the concept of dual likelihood to develop some higher order asymptotic theory for the empirical likelihood ratio test for parameters defined implicitly by a set of estimating equations. The resulting theory is likelihood based in the sense that it relies on methods developed for ordinary parametric likelihood models to obtain valid Edgeworth expansions for the maximum dual likelihood estimator and for the dual/empirical likelihood ratio statistic. In particular, the theory relies on certain Bartlett-type identities that can be used to produce a simple proof of the existence of a Bartlett correction for the dual/empirical likelihood ratio. The paper also shows that a bootstrap version of the dual/empirical likelihood ratio achieves the same higher order accuracy as the Bartlett-corrected dual/empirical likelihood ratio.

1. INTRODUCTION

Empirical likelihood (EL) is introduced by Owen (1988) as a way of extending parametric methods of inference to certain nonparametric situations. In the simplest situation, one is interested in obtaining a confidence region for the mean of the unknown distribution of a sample of independent and identically distributed (i.i.d.) observations. Without specifying a parametric form for the unknown distribution, Owen (1988, 1990) shows that under very weak conditions the EL ratio function can be used to construct confidence regions for the mean using an asymptotic χ^2 calibration. EL confidence regions have a data-determined shape (i.e., they tend to be more concentrated in places where the density of the parameters estimator is greatest), and they are not necessarily ellipsoidal, as in the case of confidence regions based on the normal approximation. Fur-

231

thermore, they are Bartlett-correctable (DiCiccio, Hall, and Romano, 1991), so that the actual coverage error differs from the nominal by a term of order $O(n^{-2})$. These attractive properties have motivated various researchers to extend the EL method to more general statistical models. See the monograph of Owen (2001) for an excellent account of recent developments and applications of EL.

In this paper we consider EL inference for the class of statistical models where the parameters are defined implicitly through a set of estimating equations. This class of models is particularly relevant in econometrics, because most maximum and quasi-maximum likelihood, nonlinear least squares, and M estimators are defined through estimating equations. We shall refer to this set of estimating equations as generalized scores (GS).

The main objective of the paper is to develop some higher order asymptotic theory for the EL ratio (ELR) test of a simple hypothesis about the parameters defined by GS. The approach we follow exploits an important connection relating EL to dual likelihood (DL) (Mykland, 1995): in the case of independent observations, the ELR statistic coincides with the DL ratio (DLR) statistic. The importance of this connection stems from the fact that the DL concept is introduced by Mykland (1995) as a device for using likelihood methods in the context of martingale inference. Thus, in the special case of independent observations, the connection between DL and EL implies the possibility of using likelihood methods in the context of EL inference. This is very convenient because we can then rely on techniques developed for ordinary parametric likelihood models and can use the classical methods of Bhattacharya and Ghosh (1978), Chandra and Ghosh (1979, 1980), and Hill and Davies (1968) to derive rigorously results on the higher order properties of the ELR statistic.

The results obtained in the paper extend and/or complement results of Owen (1988, 1991), DiCiccio and Romano (1989), DiCiccio et al. (1991), Chen (1993, 1994), Mykland (1995), and others. The new results are the following: first, we prove the validity of the Edgeworth expansion for the maximum DL estimator and for the ELR statistic under a sequence of local alternatives. The latter extends to the multiparameter case and to the third order Mykland's asymptotic analysis (1995, Sect. 7) of the alternative hypothesis induced by the DL. Incidentally, we note here that, with the exception of Chen (1994) and Lazar and Mykland (1998), much of the research on the higher order properties of the ELR statistic has been focused on its accuracy rather than on its power. The Edgeworth expansion for the ELR under a sequence of local alternatives derived in the paper fills this gap, because it can be used to approximate the power of the ELR test for particular values of the local alternatives.

Second, we show that, under the null hypothesis, the ELR statistic for the parameters defined by GS admits always a Bartlett correction. This result in itself is not new in the sense that the Bartlett correction factor obtained in the paper coincides in practice[1] with the one obtained originally by DiCiccio et al. (1991) in the case of a multivariate mean (see also, in the case of linear regressions, Chen, 1993; in the case of one parameter M functionals, see Zhang, 1996).

What is new instead is the way in which the Bartlett correction is obtained. Our method relies on certain Bartlett-type identities (Mykland, 1994) (see (5), which follows) and produces an alternative proof of the Bartlett-correctability of the ELR that complements the results of DiCiccio et al. (1991) and, we believe, provides a key to understanding the Bartlett phenomenon in the context of EL inference for GS; see the discussion following Corollary 3 in Section 3. We also propose a simple consistent estimator for the Bartlett correction.

Third, we provide asymptotic justification for a bootstrap version of the ELR statistic originally proposed by Owen (1988) as a "hybrid" method in which EL confidence regions for the parameters of interest are based on the bootstrap distribution of the ELR rather than the asymptotic χ^2 calibration. We show that the bootstrap approximates the distribution of the ELR to an order $o(n^{-1})$ except if the observed sample is contained in a set of probability $o(n^{-1})$, and we also show that the ELR with bootstrap calibration accomplishes the Bartlett correction automatically; i.e., the ELR with bootstrap critical values has the same level of accuracy as the Bartlett-corrected one. The same phenomenon was noted by Beran (1988, p. 694) in the case of parametric bootstrap likelihood ratio tests.

Finally, we prove in the Appendix a general theorem on the density of generalized noncentral quadratic forms, which is of its own independent interest, because it can be used to calculate formal Edgeworth expansions of asymptotic χ^2 test statistics admitting a stochastic expansion under a sequence of local alternatives.

The results in the paper are useful to obtain ELR tests that have a desirable higher order accuracy property. The bootstrap version of the ELR is particularly useful because the bootstrap calibration is likely to be more accurate than the Bartlett-corrected χ^2 calibration. Monte Carlo simulations (see, e.g., Corcoran, Davidson, and Spady, 1995; Baggerly, 1998; and Tables 1 and 2 in Section 4 of this paper) seem to suggest in fact that the χ^2 calibration itself is not very reasonable, implying that the practical usefulness of the Bartlett correction is limited.

The results of the paper can be applied to test simple hypotheses for parameters defined in practice by any GS: e.g., they can be applied to the ELR test for generalized linear models considered by Kolaczyk (1994) and to the ELR test for generalized projection pursuit models considered by Owen (1992). Other examples include ELR tests for "moment" models similar to those considered (albeit in the more general context of weakly dependent observations) by Burnside and Eichenbaum (1996), and for nonlinear and robust regression models; the latter are considered in Section 4 of this paper.

On the other hand, the results of this paper do not cover the case of composite hypothesis. When nuisance parameters are present (even in the form of over-identifying restrictions), it is not clear from DL arguments what the higher order behavior of the ELR statistic should be, because the Bartlett-type identities do not hold for the nuisance parameters. This is the principal way in which dual

(and hence empirical) likelihood shows different behavior from an ordinary parametric likelihood, and this implies that the profile ELR test obtained by maximizing out the nuisance parameters is in general not Bartlett-correctable (Lazar and Mykland, 1999). Whether it is possible to obtain higher order refinements to the distribution of profile ELRs is an open question that is relevant to econometrics, because the latter is typically characterized by (possibly overidentified) models with many nuisance parameters. One possibility is to adjust the critical values rather than the test statistics themselves using generalized Cornish–Fisher expansions (Hill and Davies, 1968). The calculations involved in such expansions are however notoriously difficult: computerized algebra or possibly a suitable bootstrap procedure might be very useful in this respect; see the discussion at the end of Corollary 7 in Section 3.

The remainder of the paper is organized as follows. The next section reviews briefly EL inference for GS, emphasizing its DL interpretation. Section 3 contains the main results of the paper. Section 4 considers two examples that illustrate the theory and reports the results of some Monte Carlo simulations used to assess the effectiveness of the Bartlett correction and bootstrap calibration in finite samples. Section 5 contains some concluding remarks and indications for future research. All the proofs are contained in the Appendix.

Throughout the rest of the paper we follow tensor notation and indicate arrays by their elements. Thus, for any index $1 \leq r_j \leq q$ ($j = 1, 2, \ldots, k$), a_r is an $\Re^q$-valued vector, a_{rs} is an $\Re^{q \times q}$-valued matrix, etc. We also follow the summation convention; i.e., for any two repeated indices, their sum is understood. Finally the sum $\sum$ and product $\prod$ operators are intended, unless otherwise stated, as $\sum_{i=1}^n$ and $\prod_{i=1}^n$.

2. THE RELATIONSHIP BETWEEN EMPIRICAL AND DUAL LIKELIHOOD

Suppose that the observations $(z_i)_{i=1}^n$ are i.i.d. $\Re^m$-valued random vectors from an unknown distribution F. Let $\theta \in \Theta \subseteq \Re^q$ be an unknown parameter vector associated with F and E denote the expectation operator with respect to F. We assume that the information about F and θ is available in the form

$$E[f_r(z, \theta_0)] = 0, \tag{1}$$

for some specified value θ_0 of θ, with the GS $f_r(z, \theta) : \Re^q \times \Theta \to \Re^q$-valued vector of known functionally independent functions, satisfying the following standard regularity conditions.

GS1. (i) $E[f_r(z, \theta)] = 0$ for a unique $\theta_0 \in \text{int}\{\Theta\}$, (ii) $f_r(z, \theta)$ is continuous at θ_0 with probability 1, (iii) $E[\partial f_r(z, \theta_0)/\partial \theta^s]$ is of full column rank q, (iv) $E[f_r(z, \theta_0)f_s(z, \theta_0)]$ is positive definite.

The EL approach to inference for the parameter θ defined by the constraint (1) is based on the profile ELR function

$$\mathcal{R}(\theta) = \sup_{p_i}\left\{\prod np_i \mid p_i \geq 0, \sum p_i f_r(z_i, \theta) = 0, \sum p_i = 1\right\}, \tag{2}$$

where $p_i = \Pr(z = z_i)$, which shows that EL inference may be interpreted as parametric likelihood inference using a data-determined multinomial distribution supported on the observations. Assume that

EL. $\Pr(0 \in ch\{f_r(z_1, \theta_0) \quad \ldots \quad f_r(z_n, \theta_0)\}) = 1$ as $n \to \infty$,

where $ch\{\cdot\}$ denote convex hull of the points $f_r(z_i, \theta)$, and suppose that we are interested in testing the hypothesis $H_0: \theta = \theta_0$. A Lagrange multiplier argument shows (Owen, 1990) that the unique solution of (2) evaluated at $\theta = \theta_0$ is given by $p_i = [n(1 + \lambda^r f_r(z_i, \theta_0))]^{-1}$, where $\lambda^r = \lambda^r(\theta_0)$ is an $\mathfrak{R}^q$-valued vector of Lagrange multipliers determined by $\sum f_r(z_i, \theta_0)/(1 + \lambda^r f_r(z_i, \theta_0)) = 0$. Simple algebra shows that the log ELR test $-2\log\mathcal{R}(\theta_0)$ for the null hypothesis $H_0: \theta = \theta_0$ is given by

$$W(\theta_0) = 2\sum \log(1 + \lambda^r f_r(z_i, \theta_0)), \tag{3}$$

and Owen (1990) proves that, under **GS1**(iv) and **EL**, $W(\theta_0) \xrightarrow{d} \chi_q^2$—similarly, confidence regions for θ may be obtained as the set of points θ such that $W(\theta) \leq c_\alpha$, where $c_\alpha = \Pr(\chi_q^2 \leq c_\alpha) = 1 - \alpha$.

Suppose now that, for a fixed θ_0, λ is regarded as a *free-varying* vector of unknown parameters and consider the following statistic:

$$W_\lambda(\theta_0) = \sum \log(1 + \lambda^r f_r(z_i, \theta_0)), \tag{4}$$

where the subscript λ is used to emphasize that (4) now depends on the unknown parameter λ. In the case of independent observations, $W_\lambda(\theta_0)$ coincides with the logarithm of the DL Mykland (1995); the latter corresponds to the product integral of $\lambda^r f_r(z_i, \theta_0)$ and can be viewed as a likelihood function for the parameter λ because, as long as $f_r(z_i, \theta_0)$ is bounded and λ is contained in a neighborhood of 0,

$$E\left[\prod(1 + \lambda^r f_r(z_i, \theta_0))\right] = 1, \quad \text{and} \quad 1 + \lambda^r f_r(z_i, \theta_0) > 0,$$

for small λ. A standard optimization argument shows that the ELR test of $\theta = \theta_0$ defined in (3) is then equivalent to a DL ratio (DLR) test on the dual hypothesis $H_0: \lambda = 0$, i.e.,

$$W(\theta_0) = 2(W_{\hat{\lambda}_n}(\theta_0) - W_0(\theta_0)),$$

where $W_0(\theta_0) = W_\lambda(\theta_0)|_{\lambda=0}$ and $\hat{\lambda}_n$ is the maximum DL estimator solution of the first-order conditions $\partial W_\lambda(\theta_0)/\partial \lambda^r = 0$; we shall call the maximum DL estimator DLE.

As mentioned in the introduction, the equivalence between the ELR test of $\theta = \theta_0$ and the DLR test of $\lambda = 0$ is crucial in the paper because it implies the possibility of analyzing the higher order properties of the ELR statistic using likelihood methods on the dual parameter λ instead. To elaborate further, let us introduce some additional notation: for $k = (k_1,\ldots,k_l) \in (Z_+)^l$ write $|k| = \sum_{j=1}^l k_j$; let $\partial^k W_\lambda(\cdot) = \partial^k W_\lambda(\cdot)/\partial \lambda^{r_1}\ldots\partial \lambda^{r_k}$ and $\Gamma_\tau = \Gamma(0,\tau)$ be an open sphere centered around 0 with radius $\tau > 0$. Assume that for all $\lambda \in \Gamma_\tau$

DL1. $E[|\partial^k W_\lambda(\theta_0)|] < \infty$ for $|k| \geq 0$,

where the inequality should be interpreted componentwise. For $|k| = 0$, DL1 is a sufficient condition for showing the existence and consistency of the DLE $\hat{\lambda}_n$ solution of $\partial W_\lambda(\theta_0) = 0$. To see this, notice that the $\Re^{q \times q}$-valued matrix $E[\partial^2 W_\lambda(\theta_0)]$ is negative definite as long as $1 + \lambda^r f_r(z_i,\theta_0) \geq 1/n$ (which is implied by the fact that under EL the set $S_\lambda = \{\lambda^r : 1 + \lambda^r f_r(z_i,\theta_0) \geq 1/n\}$ is closed convex and bounded). Therefore, by the implicit function theorem there exists a neighborhood Γ_τ where $\partial W_\lambda(\theta_0) = 0$ have a unique solution. Moreover, $E[W_\lambda(\theta_0)]$ is uniquely maximized at $\lambda = 0$, and $W_\lambda(\theta_0)/n \xrightarrow{P} E[W_\lambda(\theta_0)]$ for all $\lambda \in \Gamma_\tau$. Thus, given the strict concavity of $W_\lambda(\theta_0)$ in Γ_τ, it follows by standard consistency results for concave objective functions (e.g., Theorem 2.7 of Newey and McFadden, 1994) that $\hat{\lambda}_n \xrightarrow{P} 0$.

For $|k| \geq 1$, DL1 is a sufficient condition for interchanging the differential and integral operators in the equation $\partial^k E\{\exp[W_\lambda(\theta_0)] - 1\} = 0$; the latter can then be used to produce a new set of regularity conditions known in ordinary parametric likelihood theory as Bartlett identities (Bartlett, 1953). The resulting Bartlett-type identities relate linear combinations of expectations of DL derivatives and can be summarized concisely as

$$\sum_{R_k} E[\partial^{k_1} W_0(\theta_0)\ldots\partial^{k_l} W_0(\theta_0)] = 0, \tag{5}$$

where $\partial^{k_j} W_0(\cdot) = \partial^{k_j} W_\lambda(\cdot)|_{\lambda=0} (j = 1,\ldots,l)$ and the sum is over all partitions $k_1|k_2|\ldots|k_l$ of the set of indices R_k. In practice (5) can be used as in parametric likelihood inference to simplify some of the expressions arising in the calculations of the higher order cumulants of the DLR statistic (see (A.20) in the Appendix), leading to the general result (13) about the Bartlett-correctability property of the DLR (whence of the ELR).

3. HIGHER ORDER ASYMPTOTICS AND THE BOOTSTRAP FOR THE ELR STATISTIC

This section contains the main results of the paper. Section 3.1 derives valid Edgeworth expansions for the DLE and the ELR under both a sequence of local

alternatives and the null hypothesis. Section 3.2 discusses how to estimate in practice the Bartlett correction, and Section 3.3 shows how the bootstrap and EL can be combined to produce highly accurate inference for the parameters defined by GS.

3.1. A Valid Edgeworth Expansion for the DLE and for the ELR Statistic

Let us introduce some additional notation and assumptions. Let $\|\cdot\|$ denote the Euclidean norm; assume that for all $\lambda \in \Gamma_\tau$ and some $\alpha, \beta \in Z_+$

DL2.

(i) $E[|\partial^k W_\lambda(\theta_0)|^\alpha] < \infty$ for $1 \le |k| \le \beta$,
(ii) $E[\max_{|k|=\beta+1} \sup_{\|\bar{\lambda}-\lambda\|\le\tau} |\partial^k W_{\bar{\lambda}}(\theta_0)|^\alpha] < \infty$,

DL3. $\sup_{\|t\|>\delta} |E \exp\{\iota^{\bar{r}} \mathbf{Z}^l_{\bar{r}0}\}| < 1$, for $\iota^2 = -1$ and all $\delta > 0$ and $l \in Z_+$,

where, for $\bar{q}_l = \sum_{k=1}^{l} \binom{q+k-1}{k}$, $1 \le \bar{r} \le \bar{q}$, and $\mathbf{Z}^l_{\bar{r}0} \in \Re^{\bar{q}_l}$ denote the vector containing all the different DL derivatives up to lth order evaluated at $\lambda^r = 0$. Conditions DL2 and DL3 are sufficient for proving the validity of Edgeworth expansions; notice that DL2 is satisfied if $E\|f(z,\theta_0)\|^{\alpha\beta} < \infty$.

Let $\sum_{j=0}^{2} n^{-j/2} \phi_{q,\kappa_{rs}}(x) p_j(x)$ denote the formal Edgeworth expansion (Bhattacharya and Ghosh, 1978, eq. (1.14)) of the distribution of $\hat{\lambda}_n$, where $\phi_{q,\kappa_{rs}}(\cdot)$ denote the q-dimensional normal density with mean 0 and covariance $\kappa_{rs} = E[f_r(z,\theta_0)f_s(z,\theta_0)]$. The following theorem shows that the DLE admits a valid Edgeworth expansion in the sense of Bhattacharya and Ghosh (1978).

THEOREM 1. *Assume that GS1, EL, and DL2 with $\alpha = 4$, $\beta = 3$ hold. Then, for some constant $C > 0$,*

$$\Pr(n^{1/2}\|\hat{\lambda}_n\| < C(\log n)^{1/2}) = 1 - o(n^{-1}). \tag{6}$$

Furthermore, assume that the Cramér condition DL3 holds with $l = 3$. Then,

$$\sup_{B\in\mathcal{B}} \left| \Pr(n^{1/2}\hat{\lambda}_n \in B) - \sum_{j=0}^{2} n^{-j/2} \int_B \phi_{q,\kappa_{rs}}(x) p_j(x)\, dx \right| = o(n^{-1}), \tag{7}$$

where $\mathcal{B}$ is the class of Borel subsets in $\Re^q$ satisfying

$$\sup_{B\in\mathcal{B}} \int_{(\partial B)^\epsilon} \phi_{q,\kappa_{rs}}(x)\, dx = O(\epsilon) \quad \text{as } \epsilon \downarrow 0. \tag{8}$$

Before obtaining an Edgeworth expansion for the ELR under a sequence of local alternatives, we first discuss what type of local alternatives is consistent with the DL approach to EL inference. A simple modification of the argument of Mykland (1995, pp. 410–411) shows that the sequence of local alternatives induced by the DL is of the form $H_n: \lambda^r_n = \kappa_{rs} \widetilde{\lambda^s}/n^{1/2}$ for some finite non-random vector $\widetilde{\lambda^r}$. This type of local alternatives assumes implicitly that the

DL under the alternative hypothesis belongs to the same parametric subfamily specified under the null hypothesis (i.e., the product integral of $\lambda^r f_r(z_i, \theta_0)$), ruling out effectively the possibility of considering local alternatives where *both* λ and θ are allowed to depend on a sequence of local alternatives. This type of local analysis is introduced by Chesher and Smith (1997) as a generalization of the standard likelihood based local analysis (see, e.g., Peers, 1971) and requires the specification of an augmented local alternative of the form $H_n^a: \varphi_n = [\theta_n \quad \lambda_n]$, where $\theta_n = \theta_0 + \bar{\theta}/n^{1/2}$ and $\lambda_n = \lambda(\bar{\theta}/n^{1/2}) = \bar{\lambda}/n^{1/2}$ for some finite nonrandom vector $\bar{\theta}$. This general framework, however, might break down the DL construction for it requires the consideration of alternatives that might not necessarily lie in the (log)DL subfamily defined in (4). Bearing this in mind, it should be clear why the only sequence of local alternatives for the ELR consistent with the DL construction takes the form $H_n: \lambda_n^r = \kappa_{rs} \widetilde{\lambda^s}/n^{1/2}$.

For simplicity we reparameterize H_n as $H_n: \lambda_n^r = \kappa_{rs}^{1/2} \widetilde{\lambda^s}/n^{1/2}$ where $\kappa_{rs}^{1/2}$ is the square root of the symmetric positive definite matrix κ_{rs}, and we define the scaled arrays (moments)

$$\kappa_{r_1 \ldots r_k} = E[(\kappa^{r_1 s_1})^{1/2} f_{s_1}(z, \theta_0) \ldots (\kappa^{r_k s_k})^{1/2} f_{s_k}(z, \theta_0)], \tag{9}$$

so that $\kappa_{rs} = \delta^{rs}$—the Kronecker delta, i.e., $\delta^{rr} = 1$, $\delta^{rs} = 0$ for $r \neq s$. Let $G_{q,\tau}(\cdot)$ denote the distribution function of a noncentral chi square random variable with q degrees of freedom and noncentrality parameter τ. The following theorem shows that the ELR test under a sequence of local alternatives H_n admits a valid Edgeworth expansion in the sense of Chandra and Ghosh (1980).

THEOREM 2. *Suppose that the assumptions of Theorem 1 hold. Then, uniformly in $c \in [c_0, \infty)$ for some $c_0 \geq 0$ ($c_0 = 0$ if $r > 1$),*

$$\sup_{c \in [c_0, \infty)} \left| \Pr(W(\theta_0) \leq c | H_n) - \sum_{j=0}^{2} \sum_{k=0}^{4} p_{jk} G_{q+2k, \tau}(c)/n^{j/2} \right| = o(n^{-1}) \tag{10}$$

uniformly over compact subsets of $\widetilde{\lambda^r}$, where $\tau = \widetilde{\lambda^r} \widetilde{\lambda^r}$, and

$$p_{00} = 1, \qquad p_{10} = -\kappa_{rst} \widetilde{\lambda^r} \widetilde{\lambda^s} \widetilde{\lambda^t}/6, \qquad p_{11} = 0,$$

$$p_{12} = \kappa_{rst} \widetilde{\lambda^r} \widetilde{\lambda^s} \widetilde{\lambda^t}/6, \qquad p_{13} = 0, \qquad p_{14} = 0,$$

$$
\begin{aligned}
p_{20} = &-[B^{rs} - 7\kappa_{rstu} \widetilde{\lambda^t} \widetilde{\lambda^u}/12 + (8\kappa_{rst} \kappa_{tuv} - \kappa_{rtu} \kappa_{stv}) \widetilde{\lambda^u} \widetilde{\lambda^v}/9] \\
&\times (\delta^{rs} - \widetilde{\lambda^r} \widetilde{\lambda^s})/2 - 7\kappa_{rsst} \widetilde{\lambda^r} \widetilde{\lambda^t}/24 - (1-q)\tau/6 - \tau^2/2 \\
&- (7\kappa_{rtt} \delta^{su} + \kappa_{rst} \delta^{tu}) \kappa_{suv} \widetilde{\lambda^r} \widetilde{\lambda^v}/18 - (4\kappa_{rst'} \kappa_{t'tu}/9 - \kappa_{rstu}/4) \widetilde{\lambda^r} \widetilde{\lambda^s} \widetilde{\lambda^t} \widetilde{\lambda^u} \\
&+ (-\kappa_{rvv} + 2\kappa_{rvw} \widetilde{\lambda^v} \widetilde{\lambda^w}) \kappa_{stu} (-\widetilde{\lambda^r} \widetilde{\lambda^s} \widetilde{\lambda^t} \widetilde{\lambda^u}/36 + [3]\delta^{rs} \widetilde{\lambda^t} \widetilde{\lambda^u}/36) \\
&+ \kappa_{rstu}(-[3]\delta^{rs} + \widetilde{\lambda^r} \widetilde{\lambda^s}) \widetilde{\lambda^t} \widetilde{\lambda^u}/12 + [3]5\kappa_{rsu} \kappa_{tvw} \delta^{rs}([3]\delta^{tu} - \widetilde{\lambda^t} \widetilde{\lambda^u}) \\
&\times \widetilde{\lambda^v} \widetilde{\lambda^w}/216 + \kappa_{rtv} \kappa_{suw}([3]\delta^{rs} \delta^{tu} - [6]\delta^{rs} \widetilde{\lambda^t} \widetilde{\lambda^u} + \widetilde{\lambda^r} \widetilde{\lambda^s} \widetilde{\lambda^t} \widetilde{\lambda^u}) \widetilde{\lambda^v} \widetilde{\lambda^w}/72 \\
&+ (\kappa_{rtt} \kappa_{svv} + 4\kappa_{rtu} \kappa_{svw} \widetilde{\lambda^t} \widetilde{\lambda^u} \widetilde{\lambda^v} \widetilde{\lambda^w} - 4\kappa_{rtt} \kappa_{svw} \widetilde{\lambda^v} \widetilde{\lambda^w})(\widetilde{\lambda^r} \widetilde{\lambda^s} - \delta^{rs}),
\end{aligned}
$$

$$p_{21} = [B^{rs} - 7\kappa_{rstu}\widetilde{\lambda}^t\widetilde{\lambda}^u/12 + (8\kappa_{rst}\kappa_{tuv} - \kappa_{rtu}\kappa_{stv})\widetilde{\lambda}^u\widetilde{\lambda}^v/9]$$

$$\times (\delta^{rs} - 2\widetilde{\lambda}^r\widetilde{\lambda}^s)/2 + 7\kappa_{rsst}\widetilde{\lambda}^r\widetilde{\lambda}^t/24 + (-7\kappa_{rtt}\delta^{su} + \kappa_{rst}\delta^{tu})\kappa_{suv}\widetilde{\lambda}^r\widetilde{\lambda}^v/18$$

$$+ (1-q)\tau/6 + \tau^2/2 + (4\kappa_{rst'}\kappa_{t'tu}/9 - \kappa_{rstu}/4)\widetilde{\lambda}^r\widetilde{\lambda}^s\widetilde{\lambda}^t\widetilde{\lambda}^u$$

$$+ (\kappa_{rvv} - 2\kappa_{rvw}\widetilde{\lambda}^v\widetilde{\lambda}^w)\kappa_{stu}(\widetilde{\lambda}^r\widetilde{\lambda}^s\widetilde{\lambda}^t\widetilde{\lambda}^u/12 - [3]\delta^{rs}\widetilde{\lambda}^t\widetilde{\lambda}^u/18)$$

$$+ \kappa_{rstu}([3]\delta^{rs}/6 - \widetilde{\lambda}^r\widetilde{\lambda}^s/4)\widetilde{\lambda}^t\widetilde{\lambda}^u + [3]5\kappa_{rsu}\kappa_{tvw}\delta^{rs}$$

$$\times (-2[3]\delta^{tu} + 3\widetilde{\lambda}^t\widetilde{\lambda}^u)\widetilde{\lambda}^v\widetilde{\lambda}^w/216$$

$$- \kappa_{rtv}\kappa_{suw}(2[3]\delta^{rs}\delta^{tu} - 3[6]\delta^{rs}\widetilde{\lambda}^t\widetilde{\lambda}^u + 4\widetilde{\lambda}^r\widetilde{\lambda}^s\widetilde{\lambda}^t\widetilde{\lambda}^u)\widetilde{\lambda}^v\widetilde{\lambda}^w/72$$

$$+ (\kappa_{rtt}\kappa_{svv} + 4\kappa_{rtu}\kappa_{svw}\widetilde{\lambda}^t\widetilde{\lambda}^u\widetilde{\lambda}^v\widetilde{\lambda}^w - 4\kappa_{rtt}\kappa_{svw}\widetilde{\lambda}^v\widetilde{\lambda}^w)(-2\widetilde{\lambda}^r\widetilde{\lambda}^s + \delta^{rs})/72,$$

$$p_{22} = B^{rs}\widetilde{\lambda}^r\widetilde{\lambda}^s/2 - 7\kappa_{rstu}\widetilde{\lambda}^r\widetilde{\lambda}^s\widetilde{\lambda}^t\widetilde{\lambda}^u/24 + (8\kappa_{rst}\kappa_{tuv} - \kappa_{rtu}\kappa_{stv})\widetilde{\lambda}^r\widetilde{\lambda}^s\widetilde{\lambda}^u\widetilde{\lambda}^v/18$$

$$+ (\kappa_{rtt}\kappa_{svv} + 4\kappa_{rtu}\kappa_{svw}\widetilde{\lambda}^t\widetilde{\lambda}^u\widetilde{\lambda}^v\widetilde{\lambda}^w - 4\kappa_{rtt}\kappa_{svw}\widetilde{\lambda}^v\widetilde{\lambda}^w)\widetilde{\lambda}^r\widetilde{\lambda}^s/72$$

$$+ \kappa_{rtv}\kappa_{suw}([3]\delta^{rs}\delta^{tu} - 3[6]\delta^{rs}\widetilde{\lambda}^t\widetilde{\lambda}^u + 6\widetilde{\lambda}^r\widetilde{\lambda}^s\widetilde{\lambda}^t\widetilde{\lambda}^u)\widetilde{\lambda}^v\widetilde{\lambda}^w/72$$

$$- (\kappa_{rvv} - 2\kappa_{rvw}\widetilde{\lambda}^v\widetilde{\lambda}^w)\kappa_{stu}(-3\widetilde{\lambda}^r\widetilde{\lambda}^s\widetilde{\lambda}^t\widetilde{\lambda}^u + [3]\delta^{rs}\widetilde{\lambda}^t\widetilde{\lambda}^u)/36$$

$$+ [3]5\kappa_{rsu}\kappa_{tvw}\delta^{rs}([3]\delta^{tu} - 3\widetilde{\lambda}^t\widetilde{\lambda}^u)\widetilde{\lambda}^v\widetilde{\lambda}^w/216$$

$$- \kappa_{rstu}([3]\delta^{rs} - 3\widetilde{\lambda}^r\widetilde{\lambda}^s)\widetilde{\lambda}^t\widetilde{\lambda}^u/12,$$

$$p_{23} = \kappa_{rtv}\kappa_{suw}([6]\delta^{rs} - 4\widetilde{\lambda}^r\widetilde{\lambda}^s)\widetilde{\lambda}^t\widetilde{\lambda}^u\widetilde{\lambda}^v\widetilde{\lambda}^w/72 - (\kappa_{rvv} - 2\kappa_{rvw}\widetilde{\lambda}^v\widetilde{\lambda}^w)\kappa_{stu}$$

$$\times \widetilde{\lambda}^r\widetilde{\lambda}^s\widetilde{\lambda}^t\widetilde{\lambda}^u/36 - \kappa_{rstu}\widetilde{\lambda}^r\widetilde{\lambda}^s\widetilde{\lambda}^t\widetilde{\lambda}^u/12 + [3]5\kappa_{rsu}\kappa_{tvw}\delta^{rs}\widetilde{\lambda}^t\widetilde{\lambda}^u\widetilde{\lambda}^v\widetilde{\lambda}^w/216,$$

$$p_{24} = \kappa_{rtv}\kappa_{suw}\widetilde{\lambda}^r\widetilde{\lambda}^s\widetilde{\lambda}^t\widetilde{\lambda}^u\widetilde{\lambda}^v\widetilde{\lambda}^w/72,$$

where

$$B^{rs} = \kappa_{rstt}/2 - \kappa_{rtu}\kappa_{stu}/3, \tag{11}$$

and, e.g., $[3]\delta^{rs}\widetilde{\lambda}^t\widetilde{\lambda}^u = \delta^{rs}\widetilde{\lambda}^t\widetilde{\lambda}^u + \delta^{rt}\widetilde{\lambda}^s\widetilde{\lambda}^u + \delta^{ru}\widetilde{\lambda}^s\widetilde{\lambda}^t$ *is the sum over the three different ways to partition a set of four indices into two subsets of two indices each.*

The preceding expansion is useful for two reasons: first, it can be used to compute the approximate local power function of the test $H_0: \lambda = 0$ versus alternatives of the form $H_n: \lambda_n^r = \kappa_{rs}^{1/2}\widetilde{\lambda}^s/n^{1/2}$ (i.e., $\Pr(W(\theta_0) > c_\alpha | H_n)$ where $c_\alpha = \Pr(\chi_q^2 \geq c_\alpha) = \alpha$). Second, it can be used to obtain a valid Edgeworth expansion of the ELR statistic under H_0 with remainder of order $o(n^{-1})$. To improve the latter to the order $O(n^{-2})$ we need to strengthen DL2 and DL3 as in the following corollary; let $g_q(\cdot)$ denote the density of a chi square random variable with q degrees of freedom and let $B = B^{rr}$.

COROLLARY 3. *Assume that GS1, EL, and DL2 with $\alpha = 6$ and $\beta = 5$ and DL3 with $l = 4$ hold. Then, for some $c_0 \geq 0$*

$$\sup_{c \in [c_0, \infty)} |\Pr(W(\theta_0) \leq c) - G_q(c) + Bcg_q(c)/(nq)| = O(n^{-2}). \tag{12}$$

Let $W^B(\theta_0) = W(\theta_0)/(1 + B/(nq))$ and $c_\alpha = \Pr(\chi_q^2 \geq c_\alpha) = \alpha$. Then

$$\Pr(W^B(\theta_0) \leq c_\alpha) = 1 - \alpha + O(n^{-2}). \tag{13}$$

Expansion (12) shows clearly that the Bartlett factor B is the main error term for the ELR statistic; moreover because B enters linearly in the expansion it is clear that scaling $W(\theta_0)$ by the same factor improves the coverage error to the order $O(n^{-2})$, as in (13). This remarkable property of the ELR statistic is a direct consequence of the connection relating dual and empirical likelihood in the case of independent observations: because the ELR can be interpreted as a DLR for λ, and a DLR inherits properties of an ordinary parametric likelihood ratio statistic via the Bartlett-type identities (5), it is perhaps not surprising that the ELR shares the Bartlett-correctability property of an ordinary likelihood ratio statistic. In particular, as shown in the proof of Theorem 2 in the Appendix, the Bartlett-type identities (5) imply that, for DL2 with $\alpha = 5$ and $\beta = 4$ and DL3 with $l = 3$, the signed square root $W_r(\theta_0)$ of the ELR (i.e., an $\Re^q$-valued random vector such that $W_r(\theta_0)W_r(\theta_0) = W(\theta_0) + O_p(n^{-3/2})$) is asymptotically

$$W_r(\theta_0) \sim N(c_r/n^{1/2}, \delta^{rs} + c_{rs}/n) + O(n^{-3/2}), \tag{14}$$

where c_r and $c_{r,s}$ are constants defined as k_r^2 and $k_{r,s}^3$ (with $\widetilde{\lambda^r} = 0$) in (A.20) in the Appendix. This result was originally proved by DiCiccio and Romano (1989), using a different technique involving some lengthy algebra, and shows that $W_r(\theta_0)$ can be mean and variance corrected so that the resulting adjusted statistic is $N(0, \delta^{rs}) + O(n^{-3/2})$, as is typically the case for the signed square root of ordinary parametric likelihood ratios. The existence of a Bartlett correction for the ELR then follows from this result combined with an Edgeworth expansion argument. The latter shows in fact that the density of $W_r(\theta_0)W_r(\theta_0) + O_p(n^{-3/2})$ is proportional to $\exp(-x^2/2)(x^2)^{q/2-1}[1 + \psi(x^2)/n] + O(n^{-2})$ where $\psi(x^2)$ is a *linear* function in x^2, so that scaling $W(\theta_0)$ by a factor $1 + B/(nq)$ with $B = c_r c_r + c_{rr}$ eliminates the coefficient of n^{-1} in the expansion of $W^B(\theta_0)$ yielding (13).

3.2. Estimation of the Bartlett Correction

The Bartlett correction for GS

$$B = \kappa_{rrss}/2 - \kappa_{rst}\kappa_{rst}/3 \tag{15}$$

depends on the third and fourth multivariate moments of $f_r(z, \theta_0)$. The computation of the moments involved in the threefold summation in (15) takes $O(nq^3)$ computing time, so unless q is very large the computational cost of the Bartlett

correction is not very high. Recall that B is based on the scaled moments $\kappa_{r_1 \ldots r_k} = E[(\kappa^{r_1 s_1})^{1/2} f_{s_1}(z, \theta_0) \ldots (\kappa^{r_k s_k})^{1/2} f_{s_k}(z, \theta_0)]$ and suppose that $\hat{\theta} \overset{P}{\to} \theta_0$; then a simple estimator of (15) can be based on

$$\hat{B} = \hat{K}_{rstu}(\hat{\theta}) \widehat{K^{rs}}(\hat{\theta}) \widehat{K^{tu}}(\hat{\theta})/2 - \hat{K}_{rtv}(\hat{\theta}) \hat{K}_{suw}(\hat{\theta}) \widehat{K^{rs}}(\hat{\theta}) \widehat{K^{tv}}(\hat{\theta}) \widehat{K^{uw}}(\hat{\theta})/3, \quad \textbf{(16)}$$

where $\hat{K}_{r_1 \ldots r_k}(\theta) = \sum f_{r_1}(z_i, \theta) \ldots f_{r_k}(z_i, \theta)/n$. The following theorem shows that $\hat{B}$ is consistent for (15) and that replacing B with $\hat{B}$ in Corollary 3 does not alter the order of magnitude of the approximation error (13). Let $\mathcal{N}$ denote a neighborhood of θ_0 and make the following assumption.

GS2. (i) $E[\sup_{\theta \in \mathcal{N}} \|f(z, \theta)\|^4] < \infty$, (ii) $\partial f_r(z, \theta)/\partial \theta^s$ is continuous with probability 1 at θ_0, (iii) $E[\sup_{\theta \in \mathcal{N}} |\partial f_r(z, \theta)/\partial \theta^s|] < \infty$.

THEOREM 4. *Assume that GS2 holds. Then,* $\hat{B} \overset{P}{\to} B$ *for any* $\hat{\theta} \overset{P}{\to} \theta_0$. *Furthermore, suppose that the assumptions of Corollary 3 hold and that* $n^{1/2} \|\hat{\theta} - \theta_0\| = O_p(1)$. *Then, for some* $c_0 \geq 0$

$$\sup_{c \in [c_0, \infty)} |\Pr(W^{\hat{B}}(\theta_0) \leq c) - G_q(c)| = O(n^{-2}). \quad \textbf{(17)}$$

Remark. The assumption of i.i.d. sampling can be relaxed by considering $(z_{in})_{i \leq n, n \geq 1}$ as a triangular array of $\Re^m$-valued random vectors as in Owen (1991); i.e., for each n, $f_r(z_{1n}, \theta), \ldots, f_r(z_{nn}, \theta)$ are independent but not identically distributed $\Re^q$ valued random vectors. The results presented in Sections 3.1 and 3.2 are still valid by replacing some of the previous regularity conditions with the following uniform (in n) version.

GS1. (i) $E[f_r(z_{in}, \theta_0)] = 0$ for a unique $\theta_0 \in \Theta$ for all n, (ii) $\lim \inf_{n \to \infty} \zeta_n/n > 0$, where ζ_n is the smallest eigenvalue of $\sum E[f_r(z_{in}, \theta_0) f_s(z_{in}, \theta_0)]$,

GS2.

  (i) $\lim \sup_{n \to \infty} \sum E[\sup_{\theta \in \mathcal{N}} \|f(z_{in}, \theta)\|^4/n] < \infty$,
  (iii) $\lim \sup_{n \to \infty} \sum E[\sup_{\theta \in \mathcal{N}} |\partial f_r(z_{in}, \theta)/\partial \theta^s|] < \infty$.

EL. $\Pr(0 \in ch\{f_r(z_{1n}, \theta_0), \ldots, f_r(z_{nn}, \theta_0)\}) = 1$ as $n \to \infty$,

DL2.

  (i) $\lim \sup_{n \to \infty} E[|\partial^k W_{n\lambda}(\theta_0)|^\alpha] < \infty$ for $1 \leq |k| \leq \beta$ for any $\lambda^r \in \Gamma_\tau$,
  (ii) $\lim \sup_{n \to \infty} E[\max_{|k|=\beta+1} \sup_{\|\bar{\lambda} - \lambda\| \leq \tau} |\partial^k W_{n\lambda^r}(\theta_0)|^\alpha] < \infty$, and
  (iii) $\lim \sup_{n \to \infty} E[|\mathbf{Z}_{n\bar{r}0}^l|]^{\beta+1} I\{\mathbf{Z}_{n\bar{r}0}^l > \epsilon n^{1/2}\} = 0$ for every $\epsilon > 0$ and some $l \in Z_+$,

DL3. $\lim \sup_{n \to \infty} \sup_{\|t\| \geq \delta} |E[\exp(\iota t^{\bar{r}} \mathbf{Z}_{n\bar{r}0}^l)]| < 1$ for all $\delta > 0$ and some $l \in Z_+$,

where $W_{n\lambda}(\cdot)$ and $\mathbf{Z}_{n\bar{r}0}^l$ are, respectively, the log DL defined in (4) and the $\Re^{\bar{q}_l}$-valued vector of DL derivatives evaluated at $\lambda = 0$ for the triangular array

$f_r(z_{in}, \theta_0)$ $n \geq 1$. In particular GS1, GS2(i), EL, and DL2(iii) with $\beta = 3$ and $\mathbf{Z}^l_{n\bar{r}0} = Z_{nr0}$ for $l = 1$ imply Theorem 2 of Owen (1991), whereas DL2(i)–(ii) and DL3 imply that we can use Theorem 20.6 of Bhattacharya and Rao (1976) together with Theorem 3.2 and Remarks 3.3 and 3.4 in Skovgaard (1981) to justify the validity of the Edgeworth expansions in Theorems 1 and 2. The results shown in Section 3.3, which follows, are also valid by replacing BDL with uniform versions similar to those just shown.

3.3. A Valid Edgeworth Expansion for the Bootstrap DLE and the Bootstrap ELR Statistic

In this section we consider the higher order properties of bootstrap dual/EL inference. Let $\chi_n^* = (z_1^*, \ldots, z_n^*)$ denote a bootstrap sample, i.e., a resample drawn independently and uniformly from the observed sample χ_n with the property that $\Pr(z_i^* = z_j | \chi_n) = 1/n$ for $1 \leq i, j \leq n$, and let $f_r(z_i^*, \theta_0)$ denote the corresponding bootstrap GS.

We first consider bootstrapping the DLE $\hat{\lambda}_n$. Let

$$W_{\lambda^*}^*(\theta_0) = \sum \log(1 + \lambda^{*r} f_r(z_i^*, \theta_0))$$

denote the bootstrap (log) DL where, in analogy to Section 2, the dual parameter λ^{*r} is free-varying. The bootstrap DLE (BDLE) $\widehat{\lambda_n^*}$ solves $0 = \partial W_{\lambda^*}^*(\theta_0)/\partial \lambda^r$. Assume that for some $\alpha, \beta \in Z_+$ and $\gamma > 0$

BDL.

(i) $E[\sup_{\lambda \in \Gamma_\tau} |\partial^k W_\lambda(\theta_0)|^{\alpha+\gamma}] < \infty$ for $1 \leq |k| \leq \beta$,
(ii) $E[\sup_{\lambda \in \Gamma_\tau} \max_{|k|=\beta+1} \sup_{\|\bar{\lambda}-\lambda\| \leq \tau} |\partial^k W_{\bar{\lambda}}(\theta_0)|^{\alpha+\gamma}] < \infty$,

and let $\Pr^*$ denote the bootstrap probability conditional on χ_n. The following theorem establishes the higher order equivalence between the original and BDLE.

THEOREM 5. *Assume that GS1, EL, and BDL with $\alpha = 6$, $\beta = 3$, and some $\gamma > 0$ hold. Then, for some constant $C > 0$*

$$\Pr^*(n^{1/2} \| \widehat{\lambda_n^*} - \hat{\lambda}_n \| > C \log n^{1/2}) = o(n^{-1}), \tag{18}$$

except if χ_n is contained in a set with probability $o(n^{-1})$. Furthermore, assume that the Cramér condition DL3 holds with $l = 3$. Then, for every class $\mathcal{B}$ of Borel subsets in $\Re^q$ satisfying (8)

$$\sup_{B \in \mathcal{B}} |\Pr^*(n^{1/2}(\widehat{\lambda_n^*} - \hat{\lambda}_n) \in B) - \Pr(\hat{\lambda}_n \in B)| = o(n^{-1}), \tag{19}$$

except if χ_n is contained in a set of probability $o(n^{-1})$.

Notice that another way to express (18) (and analogously (19) and also (21), which follows) is

$$\Pr(|\Pr^*(n^{1/2}\|\widehat{\lambda_n^*} - \hat{\lambda}_n\| > C \log n^{1/2})| > n^{-1}) = o(n^{-1}).$$

We now consider bootstrapping the ELR. It is important to note that the validity of the bootstrap in the context of EL inference depends crucially on the validity of the bootstrap moment condition $E^*[f_r(z_i^*, \theta)] = 0$ at $\theta = \theta_0$, where E^* is the bootstrap expectation under $\Pr^*$. A straightforward method to achieve this is to consider the centered bootstrap GS $f_r^*(z_i, \theta) = f_r(z_i^*, \theta) - E^*[f_r(z_i^*, \theta_0)]$, which mimics the original GS as defined in (1) because by construction the bootstrap moment condition $E^*[f_r^*(z_i, \theta_0)] = 0$. Alternatively, we can consider the so-called biased bootstrap GS $f_r(z_i^\dagger, \theta)$, where the $z_i^\dagger$'s are resamples drawn independently from the original sample χ_n with the property that $\Pr(z_i^\dagger = z_j | \chi_n) = p_j$ for $1 \leq i, j \leq n$, and the p_js are the estimated EL probabilities. This resampling procedure was proposed, originally in 1992, by Brown and Newey (2001) in the context of bootstrapping for generalized method of moments (GMM); the term *biased* used here is borrowed from Hall and Presnell (1999). Notice that $f_r(z_i^\dagger, \theta)$ does not need to be centered, because by definition $E^\dagger[f_r(z_i^\dagger, \theta_0)] = 0$ where $E^\dagger$ is the expectation under $\Pr^\dagger$. Thus both methods lead to unbiased (conditional on χ_n) bootstrap GS at $\theta = \theta_0$. The latter fact is the key to the validity of bootstrap ELR because it implies that the distribution of the bootstrap ELR (BELR) resembles the null distribution of the original ELR regardless of whether the null hypothesis is true or not. More important, at least in the context of this paper, the results of Brown and Newey (2001) (see also Horowitz, 2001) imply that biased bootstrap leads to the same theoretical improvements to the finite-sample distribution of the ELR statistic as those obtained using the standard (uniform) bootstrap (cf. Corollary 7). Bearing this in mind, in the remaining part of this section we consider bootstrap EL inference based on the centered bootstrap GS $f_r^*(z_i, \theta)$.

An immediate consequence of centering is that the bootstrap equivalent of assumption EL holds, i.e.,

BEL. $\Pr^*(0 \in ch\{f_r^*(z_1, \theta_0) \quad \ldots \quad f_r^*(z_n, \theta_0)\}) = 1$ as $n \to \infty$.

The latter implies that the bootstrap analogue of the profile ELR function defined in (2),

$$\mathcal{R}^*(\theta) = \sup_{p_i}\left\{\prod np_i \,|\, p_i \geq 0, \sum p_i f_r^*(z_i, \theta) = 0, \sum p_i = 1\right\}, \tag{20}$$

admits a unique solution at $\theta = \theta_0$ and that such a solution can be obtained by the same Lagrangian argument of Section 2. Let $-2 \log(\mathcal{R}^*(\theta_0))$ denote the resulting log BELR. Solving (20) and calculating $-2 \log(\mathcal{R}^{*b}(\theta_0))$ for $b = 1, \ldots, B$ bootstrap samples χ_n^* yields an estimator of the distribution of the ELR

that can be used to form critical values for test statistics (and/or confidence regions) in the usual way.

As in Section 2, the BELR $-2 \log(\mathcal{R}^*(\theta_0))$ can be interpreted as a bootstrap DLR test for the dual parameter λ^{*r}, i.e.,

$$-2 \log(\mathcal{R}^*(\theta_0)) = 2(W_{\widehat{\lambda}_n^*}^*(\theta_0) - W_0^*(\theta_0)),$$

where $W_{\widehat{\lambda}_n^*}^*(\theta_0) = \sum \log(1 + \widehat{\lambda}^{*r} f_r^*(z_i, \theta_0))$ is the bootstrap DL based on the centered bootstrap GS $f_r^*(z_i, \theta_0)$, $W_0^*(\theta_0)$ is the "restricted" bootstrap DL evaluated at $\lambda^* = 0$, and $\widehat{\lambda}_n^*$ is the BDLE that solves $0 = \partial W_{\lambda^*}^*(\theta_0)/\partial \lambda^r$ using centered bootstrap GS $f_r^*(z_i, \theta_0)$. Note that because of the centering $\widehat{\lambda}_n^*$ converges (in bootstrap probability) to 0 and not to the DLE $\hat{\lambda}_n$ as in Theorem 5. Because $W_0^*(\theta_0) = 0$, let $W^*(\theta_0) = 2W_{\widehat{\lambda}_n^*}^*(\theta_0)$; the following theorem shows that the BELR statistic approximates the distribution of the ELR statistic under the null hypothesis up to the order $o(n^{-1})$.

THEOREM 6. *Suppose that the assumptions of Theorem 2 with $\alpha = 6$ hold. Then for some $c_0 \geq 0$*

$$\sup_{c \in [c_0, \infty)} |\Pr^*(W^*(\theta_0) \leq c) - \Pr(W(\theta_0) \leq c)| = o(n^{-1}), \tag{21}$$

except if χ_n is contained in a set with probability $o(n^{-1})$.

Notice that the moment assumption of Theorem 6 (which is slightly stronger than the one assumed in Theorem 2) is necessary to ensure that the BELR test is accurate up to $o(n^{-1})$; i.e., the BELR test has rejection probabilities that are correct up to the same order. To see this note that (21) is equivalent to

$$\Pr\left(\sup_{c \in [c_0, \infty)} |\Pr^*(W^*(\theta_0) \leq c) - \Pr(W(\theta_0) \leq c)| > n^{-1}\right) = o(n^{-1});$$

thus using the same arguments as Andrews (2002, pp. 149–150), it follows that for $c_\alpha^* = \inf\{c \in [c_0, \infty) : \Pr^*(W^*(\theta_0) \geq c) \geq \alpha\}$

$$\Pr(|1 - \alpha - \Pr(W(\theta_0) \leq c_\alpha^*)| > n^{-1}) = o(n^{-1}),$$

which implies that, for large n, $|1 - \alpha - \Pr(W(\theta_0) \leq c_\alpha^*)| \leq n^{-1}$, or equivalently

$$\Pr(W(\theta_0) > c_\alpha^*) = \alpha + o(n^{-1}). \tag{22}$$

Thus, with the bootstrap calibration the level of ELR test is correct through the order $O(n^{-1})$. In contrast, as shown in Theorem 2, the ELR test with χ^2 calibration is accurate up to $O(n^{-1})$. On the other hand, using a slightly refined argument based on generalized Cornish–Fisher expansions (Hill and Davies, 1968), the approximation error in (22) can be improved to $O(n^{-2})$, as next corollary shows.

COROLLARY 7. *Suppose that the assumptions of Corollary 3 with $\alpha = 8$ hold. Then,*

$$\Pr(W(\theta_0) > c_\alpha^*) = \alpha + O(n^{-2}). \tag{23}$$

Corollary 7 shows that bootstrapping the ELR delivers the same type of higher order accurate inference implied by the existence of a Bartlett correction. The same phenomenon was noted by Beran (1988) in the context of bootstrap likelihood ratio inference. Unless $\dim[f(z,\theta)]$ is very large, it is clear that the Bartlett-corrected ELR is computationally more convenient than the BELR. On the other hand, as already mentioned in the Introduction and further illustrated in Section 4 by some simulations, it seems that the BELR has better finite-sample accuracy than the Bartlett-corrected ELR. Moreover, the higher order refinements delivered by the bootstrap approach to ELR inference do not depend in general on Bartlett-type identities. This should be important when considering the ELR statistic with nuisance parameters. To see why, let η be a finite-dimensional vector of nuisance parameters and $W(\theta_0, \hat{\eta})$ denote the resulting profile ELR (PELR) where $\hat{\eta} = \max_\eta W(\theta_0, \eta)$. As shown by Lazar and Mykland (1999) the fact that the Bartlett-type identities (5) do not hold for η implies that in general $W(\theta_0, \hat{\eta})$ is not Bartlett-correctable. Suppose however that $W(\theta_0, \hat{\eta})$ admits a valid asymptotic expansion in the sense of Chandra and Ghosh (1979) and that $\Pr(\|\hat{\eta} - \eta_0\| > C(\log)^{-1}) = O(n^{-3/2})$. Then, at least theoretically, generalized Cornish–Fisher expansions can be brought to bear to show that $W(\theta_0, \hat{\eta})$ with adjusted critical values is accurate to an order $O(n^{-2})$, independent of whether the Bartlett-type identities hold or not. Furthermore, using arguments similar to those used in the proofs of Theorem 6 and Corollary 7 in the Appendix it might be possible to obtain higher order refinements to the distribution of the PELR using critical values based on the bootstrap PELR $W^*(\theta_0, \hat{\eta})$.

4. SOME ECONOMETRIC APPLICATIONS

In this section, we illustrate the theory developed in the paper by considering EL inference for two classes of widely used econometric models that, as far as we know, have not been considered previously in the literature on EL. The finite-sample effectiveness of the two proposed methods to improve the accuracy of the ELR (i.e., Bartlett correction and bootstrap critical values) is illustrated by a small Monte Carlo study.

Example I. Nonlinear regressions

We consider nonlinear regression models $y = g(x^r, \theta_0^r) + \varepsilon$ for a given (differentiable) parametric function $g(\cdot)$ and $E(\varepsilon | x^r) = 0$. Following Newey (1990), the GS is given by

$$f_r(z, \theta) = [\partial g(x^r, \theta^r)/\partial \theta^s] \varepsilon, \tag{24}$$

which corresponds to an IV based estimating function. Notice that as long as the regression function is correctly specified, we do not need to model explicitly the (conditional) variance $\sigma^2(x)$ of ε; clearly if we knew the functional form of $\sigma^2(x)$ we could use an IV based GS with instruments equal to $\partial g(x^r, \theta^r)/\partial \theta^s$ and $\sigma^2(x)^{-1}$ to gain efficiency, but we stress here our view of the EL method as a technique to improve the accuracy of inference.

For this class of models the bootstrap calibration is based on the resamples (x_i^{r*}, y_i^*), and the centered bootstrap GS is

$$f_r^*(z_i, \theta) = [\partial g(x_i^{*r}, \theta^r)/\partial \theta^s]\varepsilon_i^* - E^*[\partial g(x_i^{*r}, \theta_0^r)/\partial \theta^s]\varepsilon_i^*,$$

where $\varepsilon_i^* = y_i^* - g(x_i^{*r}, \theta_0^r)$.

In the Monte Carlo study, we specify $g(\cdot)$ as $\exp(\theta^1 + \theta^2 x)$ with $x \sim N(0,1)$ and $\varepsilon \sim N(0,1)$ or $\sim t(5)$ or $\chi_4^2 - 4$. The values of θ^1 and θ^2 are both set equal to 1. Table 1 reports the finite-sample sizes of the ELR test (3), of the Bartlett-corrected ELR with the estimated Bartlett correction (16), and of the ELR with bootstrap critical values (23) of the null hypothesis $H_0 : [\theta^1 \quad \theta^2] = [1 \quad 1]$ at 0.10, 0.05, and 0.01 nominal size. The results are based on 5,000 replications, and the bootstrap critical values are calculated from 1,000 bootstrap replications.

TABLE 1. Nonlinear regression models

Nominal size	0.100			0.050			0.010		
Model with $N(0,1)$ errors									
$n = 50$	0.128[a]	0.116[b]	0.111[c]	0.080[a]	0.071[b]	0.062[c]	0.033[a]	0.030[c]	0.026[c]
$n = 100$	0.114[a]	0.110[b]	0.109[c]	0.070[a]	0.065[b]	0.060[c]	0.026[a]	0.024[b]	0.019[c]
$n = 200$	0.110[a]	0.109[b]	0.107[c]	0.063[a]	0.061[b]	0.058[c]	0.024[a]	0.022[b]	0.017[c]
$n = 500$	0.108[a]	0.107[b]	0.106[c]	0.062[a]	0.060[b]	0.057[c]	0.022[a]	0.021[b]	0.016[c]
Model with $t(5)$ errors									
$n = 50$	0.161[a]	0.147[b]	0.134[c]	0.105[a]	0.090[b]	0.074[c]	0.042[a]	0.034[b]	0.027[c]
$n = 100$	0.135[a]	0.127[b]	0.120[c]	0.084[a]	0.074[b]	0.069[c]	0.036[a]	0.031[b]	0.024[c]
$n = 200$	0.129[a]	0.123[b]	0.115[c]	0.073[a]	0.069[b]	0.062[c]	0.031[a]	0.026[b]	0.023[c]
$n = 500$	0.125[a]	0.119[b]	0.114[c]	0.067[a]	0.064[b]	0.062[c]	0.026[a]	0.026[b]	0.022[c]
Model with $\chi_4^2 - 4$ errors									
$n = 50$	0.183[a]	0.178[b]	0.148[c]	0.144[a]	0.132[b]	0.088[c]	0.056[a]	0.051[b]	0.038[c]
$n = 100$	0.177[a]	0.170[b]	0.132[c]	0.130[a]	0.120[b]	0.078[c]	0.051[a]	0.048[b]	0.032[c]
$n = 200$	0.152[a]	0.144[b]	0.126[c]	0.121[a]	0.109[b]	0.073[c]	0.044[a]	0.041[b]	0.028[c]
$n = 500$	0.146[a]	0.140[b]	0.123[c]	0.115[a]	0.099[b]	0.072[c]	0.040[a]	0.039[b]	0.027[c]

[a]ELR (3).
[b]ELR with estimated Bartlett correction (16).
[c]ELR with bootstrap critical values (23).
Note: Underlined values indicate that the empirical level is not statistically different from the nominal at the 0.05 level.

Example II. Robust regressions

We consider robust regression models with fixed regressors x_i^r that are assumed to satisfy $\max_{1 \le i \le n} \| x_i^r \| = O(1)$. The GS in this case is given by

$$f_r(z_i, \theta) = x_i^r \psi(y_i - x_i^s \theta^s) \tag{25}$$

for the psi function $\psi : \Re \to \Re$ satisfying $E[\psi(y_i - x_i^r \theta_0^r)] = 0$. For this model, the centered bootstrap GS is

$$f_r^*(z_i, \theta) = x_i^r [\psi(y_i^* - x_i^s \theta^s) - \varepsilon_n^*],$$

where $y_i^* = x_i^r \theta_0^r + \varepsilon_i^*$ is the bootstrap pseudo-observation, ε_i^* is the bootstrap sample drawn from $\varepsilon_i = y_i - x_i^r \theta_0^r$, and $\varepsilon_n^* = E^*[\psi(\varepsilon_i^*)]$.

Following Huber (1973), we specify the psi function $\psi(\cdot)$ as

$$(y_i - x_i^r \theta^r) I\{|y_i - x_i^r \theta^r| \le k\} + k \cdot \mathrm{Sgn}\{y_i - x_i^r \theta^r\} I\{|y_i - x_i^r \theta^r| > k\}$$

with the constant $k = 1.4$, the scale parameter $\sigma^2 = 1$, and $\mathrm{Sgn}\{\cdot\}$ denoting the sign function.

Table 2 reports some Monte Carlo results for a simple two covariates design with an intercept and a single fixed regressor x_i generated as an equally spaced

TABLE 2. Robust regression model

Nominal size	0.100			0.050			0.010		
Model with $N(0,1)$ errors									
$n = 50$	0.162[a]	0.151[b]	0.136[c]	0.097[a]	0.087[b]	0.079[c]	0.039[a]	0.034[b]	0.028[c]
$n = 100$	0.145[a]	0.132[b]	0.128[c]	0.083[a]	0.072[b]	<u>0.066</u>[c]	0.031[a]	0.027[b]	0.023[c]
$n = 200$	0.124[a]	<u>0.120</u>[b]	<u>0.118</u>[c]	0.070[a]	<u>0.064</u>[b]	<u>0.059</u>[c]	0.028[a]	0.026[b]	<u>0.020</u>[c]
$n = 500$	0.123[a]	<u>0.119</u>[b]	<u>0.115</u>[c]	0.069[a]	<u>0.065</u>[b]	<u>0.059</u>[c]	0.027[a]	0.024[b]	<u>0.019</u>[c]
Model with $t(5)$ errors									
$n = 50$	0.173[a]	0.162[b]	0.153[c]	0.112[a]	0.093[b]	0.084[c]	0.055[a]	0.045[b]	0.037[c]
$n = 100$	0.152[a]	0.137[b]	0.132[c]	0.101[a]	0.085[b]	0.073[c]	0.045[a]	0.039[b]	0.035[c]
$n = 200$	0.140[a]	0.130[b]	<u>0.120</u>[c]	0.084[a]	0.077[b]	0.069[c]	0.039[a]	0.031[b]	0.024[c]
$n = 500$	0.133[a]	<u>0.123</u>[b]	<u>0.118</u>[c]	0.079[a]	0.072[b]	<u>0.066</u>[c]	0.032[a]	0.030[b]	0.023[c]
Model with $\chi_4^2 - 4$ errors									
$n = 50$	0.191[a]	0.184[b]	0.166[c]	0.141[a]	0.123[b]	0.104[c]	0.061[a]	0.053[b]	0.034[c]
$n = 100$	0.180[a]	0.175[b]	0.154[c]	0.132[a]	0.110[b]	0.091[c]	0.055[a]	0.053[b]	0.029[c]
$n = 200$	0.172[a]	0.164[b]	0.143[c]	0.119[a]	0.109[b]	0.083[c]	0.050[a]	0.047[b]	0.027[c]
$n = 500$	0.165[a]	0.161[b]	0.138[c]	0.115[a]	0.108[b]	0.080[c]	0.046[a]	0.040[b]	0.027[c]

[a] ELR (3).
[b] ELR with estimated Bartlett correction (16).
[c] ELR with bootstrap critical values (23).
Note: Underlined values indicate that the empirical level is not statistically different from the nominal at the 0.05 level.

grid of numbers between -1 and 1 and points at -3 and 3, so that we have a rather substantial leverage effect. As with the previous example, the error ε is specified to be $N(0,1)$ or $t(5)$, or $\chi_4^2 - 4$, and the two parameters θ^1 and θ^2 are set equal to 1. Table 2 reports the finite-sample sizes of the ELR test (3), of the Bartlett-corrected ELR with estimated Bartlett correction (16), and of the ELR with bootstrap critical values (23) of the null hypothesis $H_0: [\theta^1 \quad \theta^2] = [1 \quad 1]$ at 0.10, 0.05, and 0.01 nominal size. The results are based on 5,000 replications, and the bootstrap critical values are calculated from 1,000 bootstrap replications.

Bearing in mind that the scale of the simulation study is small, the results of Tables 1 and 2 seem to indicate the following. First, with the exception of the nonlinear regression case with $N(0,1)$ errors, the ELR test is characterized by a noticeable size distortion, although the distortion tends to diminish when the sample size increases. Not surprisingly, the size distortion is more severe for the skewed $\chi_4^2 - 4$ errors. Second, in the case of errors from a symmetric distribution, both the Bartlett correction and the bootstrap reduce the finite-sample size distortion of the ELR, although some size distortion is still present, especially at the 0.01 nominal level. On the other hand, for skewed errors the effectiveness of the Bartlett correction is reduced considerably. Finally, the ELR with bootstrap calibration has smaller size distortion than the ELR with a Bartlett-corrected χ^2 calibration. The first point supports the findings of Corcoran et al. (1995) and Baggerly (1998) about the relative poor quality of the χ^2 approximation to the distribution of the ELR statistic. The second point depends clearly on the form of the Bartlett correction (15), which implies that nonzero skewness typically reduces the magnitude of the Bartlett correction itself. The last point is related to the first one and should not therefore come as a surprise: the effectiveness of the Bartlett correction depends crucially on the quality of the χ^2 approximation. Thus for data sets for which the latter is not reasonable there should be significant gains by using the bootstrap.

To conclude this section, it is worth mentioning that the EL framework can easily incorporate additional information, most noticeably information about the second moment. For example, the conditional variance $\sigma^2(x)$ of the innovations ε in Example I can be parameterized by a known function $h(x^r, \eta^a)$ that depends on an additional $\Re^p$-valued vector of parameters η^a—which may include θ^r also—and, similarly, in Example II one can introduce an estimating equation for the scale parameter σ^2.

5. CONCLUSIONS

In this paper we have developed some higher order asymptotic theory for the ELR test for parameters defined implicitly by GS. By exploiting the connection between empirical and dual likelihood we have obtained valid Edgeworth expansions for the distribution of the DLE and of the ELR. The latter is used to (i) derive an explicit expression of the third-order power function of the ELR

under a sequence of local alternatives, (ii) justify rigorously the existence of a Bartlett correction for the ELR, and (iii) show that the "hybrid" bootstrap procedure suggested by Owen (1988) yields the same type of higher order accurate inference implied by the existence of a Bartlett correction for the ELR. The derivation of (i) and (ii) relies on a set of Bartlett-type identities that, apart from simplifying as in parametric likelihood inference some calculations, provide a simple explanation of the Bartlett-correctability phenomenon in the context of EL inference for GS.

The application of bootstrap methods in the context of EL inference seems very promising and suggests some directions for future research. First, ELRs with bootstrap calibration appear to have better finite-sample size properties than ELRs with a Bartlett-corrected χ^2 calibration. This is certainly true in the Monte Carlo study reported in the paper; however, more simulation studies are required to support this conclusion in full generality.

Second, bootstrap methods might be used to deliver higher order refinements to the distribution of ELRs when nuisance parameters are present. We have conjectured that this might be the case as long as we can obtain Cornish–Fisher expansions for the critical values of the resulting profiled ELRs. This possibility is certainly of interest and is left for future research.

Finally, bootstrap methods can be used in the context of weakly dependent processes. Kitamura (1997) shows that it is possible to obtain higher order refinements to the distribution of the ELR statistic for smooth functions of means of α-mixing processes, using blockwise resampling techniques similar to those used in the bootstrap literature. The same methods could be used in the context of GS for time series models.

NOTE

1. Our calculations support Keith Baggerly's conjecture, as reported in the Errata section of Owen's empirical likelihood Web page: http://www-stat.stanford.edu/~owen/empirical, that there is a mistake in the formula of DiCiccio et al. (1991, p. 1056) of the Bartlett correction. Specifically the term t_2 should be 0, whereas the term t_1 should be $\frac{1}{3}$. With this correction, the formula of DiCiccio (1991) coincides with the one obtained in this paper (cf. (15)) in the case of GS.

REFERENCES

Abramowitz, M. & I. Stegun (1970) *Handbook of Mathematical Functions*. Dover.

Andrews, D. (2002) Higher-order improvements of a computationally attractive *k*-step bootstrap for extremum estimators, *Econometrica* 70, 119–162.

Babu, G. & K. Singh (1984) On one term Edgeworth corrections by Efron's bootstrap. *Sankhyā A* 46, 219–232.

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

Barndorff-Nielsen, O. & P. Hall (1988) On the level-error after Bartlett adjustment of the likelihood ratio statistic. *Biometrika* 75, 374–378.

Bartlett, M. (1953) Approximate confidence intervals II: More than one unknown parameter. *Biometrika* 40, 306–317.

Beran, R. (1988) Prepivoting test statistics: A bootstrap view of asymptotic refinements. *Journal of the American Statistical Society* 83, 687–697.

Bhattacharya, R. & J. Ghosh (1978) On the validity of formal Edgeworth expansion. *Annals of Statistics* 6, 434–451.

Bhattacharya, R. & R. Rao (1976) *Normal Approximation and Asymptotic Expansions*. Wiley.

Brown, B. & W. Newey (2001) GMM, Efficient Bootstrapping, and Improved Inference. Manuscript, MIT.

Burnside, C. & M. Eichenbaum (1996) Small sample properties of generalized method of moments based Wald tests. *Journal of Business and Economic Statistics* 14, 294–308.

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

Chandra, T. & J. Ghosh (1980) Valid asymptotic expansions for the likelihood ratio and other statistics under contiguous alternatives. *Sankhyā A* 42, 170–184.

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

Chen, S.X. (1994) Comparing empirical likelihood and bootstrap hypothesis tests. *Journal of Multivariate Analysis* 51, 277–293.

Chesher, A. & R. Smith (1997) Likelihood ratio specification tests. *Econometrica* 65, 627–646.

Corcoran, S., A. Davison, & R. Spady (1995) Reliable Inference from Empirical Likelihoods. Technical report, Department of Statistics, Oxford University.

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

DiCiccio, T. & J. Romano (1989) On adjustments based on the signed root of the empirical likelihood ratio statistic. *Biometrika* 76, 447–456.

Hall, P. & B. Presnell (1999) Intentionally biased bootstrap methods. *Journal of the Royal Statistical Society B* 61, 143–158.

Hill, G. & A. Davies (1968) Generalized asymptotic expansions of Cornish-Fisher type. *Annals of Mathematical Statistics* 39, 1264–1273.

Horowitz, J. (2001) The bootstrap. In J. Heckman & E. Leamer (eds.), *Handbook of Econometrics*, vol. 5, pp. 3159–3228. North-Holland.

Huber, P. (1973) Robust regression: Asymptotics, conjectures, and Monte Carlo. *Annals of Statistics* 1, 799–821.

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

Kolaczyk, E. (1994) Empirical likelihood for generalized linear models. *Statistica Sinica* 4, 199–218.

Lazar, N. & P. Mykland (1998) An evaluation of the power and conditionality properties of empirical likelihood. *Biometrika* 85, 523–534.

Lazar, N. & P. Mykland (1999) Empirical likelihood in the presence of nuisance parameters. *Biometrika* 86, 203–211.

Muirhead, R. (1982) *Aspects of Multivariate Statistical Theory*. Wiley.

Mykland, P. (1994) Bartlett type of identities for martingales. *Annals of Statistics* 22, 21–38.

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

Newey, W. (1990) Efficient instrumental variables estimation of nonlinear models. *Econometrica* 58, 809–837.

Newey, W. & D. McFadden (1994) Large sample estimation hypothesis testing. In R. Engle & D. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2113–2445. North-Holland.

Owen, A. (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 36, 237–249.

Owen, A. (1990) Empirical likelihood ratio confidence regions. *Annals of Statistics* 18, 90–120.

Owen, A. (1991) Empirical likelihood for linear models. *Annals of Statistics* 19, 1725–1747.

Owen, A. (1992) Empirical Likelihood and Generalized Projection Pursuit. Technical report 393, Department of Statistics, Stanford University.

Owen, A. (2001) *Empirical Likelihood*. Chapman and Hall.

Peers, H. (1971) Likelihood ratio and associated test criteria. *Biometrika* 58, 577–587.

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

Zhang, B. (1996) On the accuracy of empirical likelihood confidence intervals for *m*-functionals. *Journal of Nonparametric Statistics* 6, 311–321.

APPENDIX

A Theorem on the Density of Generalized Noncentral Quadratic Forms and Its Application to Hermite Polynomials. The following theorem derives a general formula useful to express the density of generalized noncentral quadratic forms (i.e., of scalar random variables obtained by contracting nonzero mean normal random vectors over multidimensional arrays of constants) in terms of finite linear combinations of noncentral chi square random variables. The formula can be applied to obtain asymptotic expansions for test statistics under a sequence of local alternatives starting from the Edgeworth expansion of their corresponding signed square root, rather than from their approximate moment generating function.

Let $\phi_{q,\gamma}(\cdot)$ denote the q-dimensional normal density with mean γ and identity covariance matrix and let $g_{q,\tau}(\cdot)$ denote the density of a noncentral chi square random variate with q degrees of freedom and noncentrality parameter $\tau < \infty$.

THEOREM 8. *Let $w^{R_k} = w^{r_1}\dots w^{r_k}$ where each $w^{r_j} \sim \phi_{q,\gamma}(w)$ ($j = 1,2,\dots,k$), $b^{R_k} = b^{r_1 \cdots r_k}$ an $\mathfrak{R}^{q^k}$-valued array of constants not depending on n, t^r an $\mathfrak{R}^q$-valued vector of auxiliary variables, and $\partial^k(\cdot) = \partial^k(\cdot)/\partial t^{r_1}\dots \partial t^{r_k}$. Consider the function $f(\cdot;t^r): \mathfrak{R}^q \to \mathfrak{R}$ and make the following assumptions.*

GQF1. $f(\cdot;t^r) \in \mathcal{C}_{\mathcal{N}}^k$, the space of k times continuously differentiable functions on an open set $\mathcal{N}$ of $t^r = 0$,

GQF2. $\int \sup_{t^r \in \mathcal{N}} |\partial^k f(\cdot,v^r;t^r)| \, dv^r < \infty$,

where the $\mathfrak{R}^q$-valued vector v^r is defined subsequently. Then, for any arbitrary noncentral k form $w^{R_k} b^{R_k}$, the following holds:

$$w^{R_k} b^{R_k} \phi_{q,\gamma}(w) = \sum_{\Upsilon} b^{R_k} \gamma^{R_{k_1}} \delta^{R_{k_2}} g_{q+2\#_{R_k},\tau}(x), \tag{A.1}$$

where $\sum_{j=1}^2 k_j = k$, $x = w^r w^r$, $\tau = \gamma^r \gamma^r$, $\gamma^{R_{k_1}} = \gamma^{r_1}\dots \gamma^{r_{k_1}}$, $\delta^{R_{k_2}} = \delta^{r_{k_1+1} r_{k_1+2}}\dots \delta^{r_{k_2-1} r_{k_2}}$, the symbol $\#_{R_k}$ denotes the number of different indices in the set R_k, and the sum is over $\Upsilon = \{\nu_2,\dots,\nu_k\}$—the number of ways of partitioning a set of R_k indices into ν_j subsets containing j indices ($j = 2,\dots,k$) such that the resulting homogeneous polynomial in $\gamma^{R_{k_1}}$ is even or odd according to the number of indices in the set R_k.

Proof. We use the transformation from $\Re^q$ to $\Re^{q+1}$

$$T: w^r \rightarrow (x, v^r),$$

where $x = w^r w^r$, $v^r = w^r/(w^s w^s)^{1/2}$ and the following identity:

$$w^{R_k} b^{R_k} \phi_{q, \gamma}(w) \equiv \sum_{j=1}^{k} b^{R_j} \partial^j [\phi_{q, \gamma}(w) \exp(w^r t^r)]|_{t=0}.$$

Using T, the density for x is obtained by integrating out the vector $v^r \in V_{1,q}$ (i.e., over the unit sphere $v^r v^r = 1$ in $\Re^q$), i.e.,

$$(2\pi)^{-q/2} \int_{V_{1,q}} \sum_{j=1}^{k} b^{R_j} \exp[-(x+\tau)/2]|J_T| \partial^j [\exp(x^{1/2} v^r t^r)]|_{t=0}(dv^r), \qquad \textbf{(A.2)}$$

where $|J_T| = x^{q/2-1}/2$ is the Jacobian of the transformation T and (dv^r) denotes the unnormalized Haar measure on the Stiefel manifold $V_{1,q}$. Upon normalizing the Haar measure on $V_{1,q}$ by the constant $2\pi^{q/2}/\Gamma(q/2)$ and interchanging differentiation and integration in (A.2), which is permissible under GQF2 for $f(\cdot, v^r; t^r) = \exp(x^{1/2} v^r t^r)$, we can then use Theorem 7.4.1 in Muirhead (1982) to obtain

$$C(x, q, \tau) \sum_{j=1}^{k} b^{R_j} \partial^j [{}_0F_1(; q/2; x(\tau + t^r t^r + 2\gamma^r t^r)/4)]|_{t=0} \qquad \textbf{(A.3)}$$

with $C(x, q, \tau) = x^{q/2-1} \exp[-(x+\tau)/2]/[2^{q/2}\Gamma(q/2)]$, ${}_0F_1(; c; z) = \sum_{j=0}^{\infty} z^j/(c)_j j!$, and $(c)_j = \Gamma(c+j)/\Gamma(c)$. Differentiating (A.3) and evaluating the resulting derivatives at $t = 0$ yields a polynomial in x of degree at most k with coefficients given by

$${}_0F_1(; q/2 + \#_{R_k}; x\tau/4) C(b, \gamma, \#_{R_k})/[2^{\#_{R_k}}(q/2)_{\#_{R_k}}] \qquad k = 1, 2, \dots \qquad \textbf{(A.4)}$$

The constant $C(b, \gamma, \#_{R_k})$ in (A.4) is itself an even or odd polynomial in $\gamma^{R_{k_1}}$ with coefficients obtained by contracting accordingly the components of $\gamma^{R_{k_1}}$, $\delta^{R_{k_2}}$, and b^{R_k}. Elementary symmetry considerations show that the number of such contractions can be found using standard combinatorial results on partitions reported, e.g., in Abramowitz and Stegun (1970, Table 24.2, p. 831). Combining (A.3) and (A.4), because $g_{q,\tau}(x) = \exp[-(x+\tau)/2] x^{q/2-1} {}_0F_1(; q/2; x\tau/4)/2^{q/2}\Gamma(q/2)$, (A.1) follows immediately. ∎

For generalized quadratic forms up to $k = 6$ (corresponding to the terms appearing in third-order Edgeworth expansions), (A.1) yields

$$b^r w^r \phi_{q, \gamma}(w) = b^r \gamma^r g_{q+2, \tau}(x),$$

$$b^{rs} w^r w^s \phi_{q, \gamma}(w) = b^{rs} \gamma^r \gamma^s g_{q+4, \tau}(x) + b^{rs} \delta^{rs} g_{q+2, \tau}(x),$$

$$b^{rst} w^r \dots w^t \phi_{q, \gamma}(w) = b^{rst} \gamma^r \gamma^s \gamma^t g_{q+6, \tau}(x) + [3] b^{rst} \gamma^t \delta^{rs} g_{q+4, \tau}(x),$$

$$b^{r \dots u} w^r \dots w^u \phi_{q, \gamma}(w) = b^{r \dots u} \gamma^r \dots \gamma^u g_{q+8, \tau}(x) + [6] b^{r \dots u} \gamma^r \gamma^s \delta^{tu} g_{q+6, \tau}(x)$$
$$+ [3] b^{r \dots u} \delta^{rs} \delta^{tu} g_{q+4, \tau}(x),$$

$$b^{r \dots v} w^r \dots w^v \phi_{q, \gamma}(w) = b^{r \dots v} \gamma^r \dots \gamma^v g_{q+10, \tau}(x) + [10] b^{r \dots v} \gamma^r \gamma^s \gamma^t \delta^{uv} g_{q+8, \tau}(x)$$
$$+ [15] b^{r \dots v} \gamma^r \delta^{st} \delta^{uv} g_{q+6, \tau}(x),$$

$$b^{r \dots w} w^r \dots w^w \phi_{q, \gamma}(w) = b^{r \dots w} \gamma^r \dots \gamma^w g_{q+12, \tau}(x) + [15] b^{r \dots w} \gamma^r \dots \gamma^u \delta^{vw} g_{q+10, \tau}(x)$$
$$+ [45] b^{r \dots w} \gamma^r \gamma^s \delta^{tu} \delta^{vw} g_{q+8, \tau}(x)$$
$$+ [15] b^{r \dots w} \delta^{rs} \delta^{tu} \delta^{vw} g_{q+6, \tau}(x),$$

from which the generalized quadratic forms $b^{R_k} h^{R_k}$ $(1 \le k \le 6)$ based on the multivariate Hermite polynomial $h^{R_k} = h^{r_1 \cdots r_k} = (-1)^k \partial^k \phi_{q,\gamma}(w)/\partial w^{r_1} \ldots \partial w^{r_k}$ have densities

$$b^r h^r = b^r \gamma^r \nabla g_{q,\tau}(x), \qquad b^{rs} h^{rs} = b^{rs} \gamma^r \gamma^s \nabla^2 g_{q,\tau}(x) + b^{rs} \delta^{rs} \nabla g_{q,\tau}(x),$$

$$b^{rst} h^{rst} = b^{rst} \gamma^r \gamma^s \gamma^t \nabla^3 g_{q,\tau}(x) + [3] b^{rst} \gamma^t \delta^{rs} \nabla^2 g_{q,\tau}(x),$$

$$b^{r \cdots u} h^{r \cdots u} = b^{r \cdots u} \gamma^r \ldots \gamma^u \nabla^4 g_{q,\tau}(x) + [6] b^{r \cdots u} \gamma^r \gamma^s \delta^{tu} \nabla^3 g_{q,\tau}(x)$$
$$+ [3] b^{r \cdots u} \delta^{rs} \delta^{tu} \nabla^2 g_{q,\tau}(x),$$

$$b^{r \cdots v} h^{r \cdots v} = b^{r \cdots v} \gamma^r \ldots \gamma^v \nabla^5 g_{q,\tau}(x) + [10] b^{r \cdots v} \gamma^r \gamma^s \gamma^t \delta^{uv} \nabla^4 g_{q+8,\tau}(x)$$
$$+ [15] b^{r \cdots v} \gamma^r \delta^{st} \delta^{uv} \nabla^3 g_{q,\tau}(x),$$

$$b^{r \cdots w} h^{r \cdots w} = b^{r \cdots w} \gamma^r \ldots \gamma^w \nabla^6 g_{q,\tau}(x) + [15] b^{r \cdots w} \gamma^r \ldots \gamma^u \delta^{vw} \nabla^5 g_{q,\tau}(x)$$
$$+ [45] b^{r \cdots w} \gamma^r \gamma^s \delta^{tu} \delta^{vw} \nabla^4 g_{q,\tau}(x) + [15] b^{r \cdots w} \delta^{rs} \ldots \delta^{vw} \nabla^3 g_{q,\tau}(x),$$

$$\text{(A.5)}$$

where $\nabla^k g_{q,\tau}(\cdot)$ is the kth-difference operator applied to the density $g_{q,\tau}(\cdot)$, i.e., $\nabla^k g_{q,\tau}(\cdot) = \sum_{j=0}^{k} (-1)^j \binom{k}{j} g_{q+2(k-j),\tau}(\cdot)$.

Proofs. Let C denote a generic positive constant not depending on n that may vary from one (in)equality to another. For simplicity of notation let $W(\theta_0) = W$ and $W_\lambda(\theta_0) = W_\lambda$.

Proof of Theorem 1. We first establish (6). Let $\lambda^{R_k} = \lambda^{r_1} \ldots \lambda^{r_k}$; a Taylor expansion of the DL first-order condition $0 = \partial W_\lambda$ about 0 gives, for any $\lambda^r \in \Gamma_\tau$,

$$0 = \partial W_0/n + \sum_{k=2}^{3} \partial^k W_0 \lambda^{R_{k-1}}/(k-1)!n + R_n(\bar{\lambda}), \qquad \text{(A.6)}$$

where $R_n(\bar{\lambda}) \le C \|\lambda^r\|^3 \max_{|k|=4} \sup_{\|\bar{\lambda} - \lambda\| \le \tau} |\partial^k W_{\bar{\lambda}}|/n$. By DL2 for some $C_\varepsilon < \infty$ and all $\varepsilon > 0$

$$\Pr(\|\partial W_0/n^{1/2}\| > C(\log n)^{1/2}) \le Cn^{-1}(\log n)^{-2},$$

$$\Pr(|\partial^k W_0 - E[\partial^k W_0]| > \varepsilon n) = o(n^{-1}) \quad \text{for } 1 \le |k| \le 3,$$

$$\Pr\left(\max_{|k|=4} \sup_{\|\bar{\lambda} - \lambda\| \le \tau} |\partial^k W_{\bar{\lambda}}|/n > C_\varepsilon\right) = o(n^{-1}), \qquad \text{(A.7)}$$

where the first inequality follows by a moderate deviations result of Bhattacharya and Rao (1976, Corollary 17.12), the second by Rosenthal and Markov inequality, and the third by combining the second equality with the triangle inequality. Let S_n denote the union of the sets in (A.7); clearly $\Pr(S_n) = o(n^{-1})$, so that on the set S_n^c we can write (A.6) as $\lambda^r = g_n(\lambda^r)$ where $g_n(\cdot)$ is a continuous function from $\mathfrak{R}^q \to \mathfrak{R}^q$ satisfying $\|g_n(u)\| \le Cn^{-1/2}(\log n)^{1/2}$ for all $\|u\| \le Cn^{-1/2}(\log n)^{1/2}$. Application of Brower's fixed point theorem as in Theorem 2.1 of Bhattacharya and Ghosh (1978) and the concavity of W_λ show that there exists a unique sequence λ_n^r such that $\Pr(\|\lambda_n^r\| < Cn^{-1/2}(\log n)^{1/2}) \ge 1 - Cn^{-1}(\log n)^{-2}$, whence (6). To derive the Edgeworth expansion (7), we first derive a stochastic expansion for $\hat{\lambda}_n$. For notational convenience let $\lambda = n^{1/2} \hat{\lambda}_n$; define the following $O_p(1)$ random arrays:

$$Z_{r_1 \ldots r_k} = \sum f_{r_1}(z_i, \theta_0) \ldots f_{r_k}(z_i, \theta_0)/n^{1/2} - n^{1/2} \kappa_{r_1 \ldots r_k}, \tag{A.8}$$

where $\kappa_{r_1 \ldots r_k} = E[f_{r_1}(z, \theta_0) \ldots f_{r_k}(z, \theta_0)]$, and let

$$\kappa_{R_k, R_l, \ldots, R_v} = E[Z_{R_k}, Z_{R_l}, \ldots, Z_{R_v}] \tag{A.9}$$

denote the vth-order (mixed) cumulant of $Z_{R_j} = Z_{r_1 \ldots r_j}$.

Combining (A.7) and (A.8) yields

$$0 = Z_r - Z_{rs}\lambda^r/n^{1/2} - \kappa_{rs}\lambda^r + \partial^3 W_0 \lambda^{R_3}/(2n^{3/2}) + R_n(\bar{\lambda}), \tag{A.10}$$

for $\|\bar{\lambda}\| \le \|\lambda\|$, where $\Pr(\|R_n(\bar{\lambda})\| > Cn^{-2}(\log n)^{5/2}) = o(n^{-1})$. Because $\lambda^r = \kappa^{rs}Z_s + \zeta^r_{1n}$, where $\Pr(\|\zeta^r_{1n}\| > Cn^{-1/2}(\log n)) = o(n^{-1})$, and $\partial^3 W_0/n \xrightarrow{p} 2\kappa_{rst}$ by the weak law of large numbers, (A.10) can be written as

$$\lambda^r = \kappa^{rs}Z_s + (-\kappa^{rs}\kappa^{tu}Z_{st}Z_u + \kappa^{ru}\kappa^{sv}\kappa^{tw}\kappa_{uvw}Z_s Z_t)/n^{1/2} + \zeta^r_{2n},$$

where $\Pr(\|\zeta^r_{2n}\| > Cn^{-1}(\log n)^{3/2}) = o(n^{-1})$. Upon substituting this last stochastic expansion into (A.10), because $\partial^4 W_0/n \xrightarrow{p} -6\kappa_{rstu}$, we obtain the stochastic expansion

$$\lambda^r = Z^r + (-\kappa^{rs}Z_{st}Z^t + \kappa^{rs}\kappa_{stu}Z^t Z^u)/n^{1/2}$$

$$+ (\kappa^{rs}\kappa^{tu}Z_{st}Z_{uv}Z^v - 2\kappa^{ru}\kappa^{tw}\kappa_{uvw}Z^v Z_{tu'}Z^{u'} + \kappa^{rs}Z_{stu}Z^t Z^u$$

$$- \kappa^{rs}\kappa^{tu}\kappa_{uvw}Z_{st}Z^v Z^w - \kappa^{rs}\kappa_{stuv}Z^t Z^u Z^v$$

$$+ 2\kappa^{rr'}\kappa^{ss'}\kappa^{t'u'}\kappa_{r's't'}\kappa_{u'uv}Z_s Z^u Z^v)/n + \zeta^r_{3n}$$

$$= Z^r + \Lambda^r_n + \zeta^r_{3n}, \tag{A.11}$$

where $Z^r = \kappa^{rs}Z_s$ and $\Pr(\|\zeta^r_{3n}\| > Cn^{-3/2}(\log n)^2) = o(n^{-1})$. To derive the Edgeworth expansion of λ^r on $B \in \mathcal{B}$ (the class of all Borel subsets of $\mathfrak{R}^q$), notice that λ^r is a function of the three random arrays Z_r, Z_{rs}, and Z_{rst} and recall that $\mathbf{Z}^3_{\bar{r}}$ is the $\mathfrak{R}^{\bar{q}_3}$-valued vector containing all the different DL derivatives under H_0 up to third order. Let $\gamma_{\mathbf{Z}^3_{\bar{r}}}(\mathbf{B})$ denote the (signed) measure corresponding to the characteristic function of $\mathbf{Z}^3_{\bar{r}}$ for $\mathbf{B} \in \bar{\mathcal{B}}$ (the class of all Borel subsets of $\mathfrak{R}^{\bar{q}_3}$) and consider the following continuous transformation: $\mathbf{U}_{\bar{r}} = \chi(\mathbf{Z}^3_{\bar{r}})$, where

$$\chi : U_r = Z_r, \qquad U_{rs} = -Z_{rs} + \kappa_{r,st}Z^t, \qquad U_{rst} = 2Z_{rst} - 2\kappa_{r,stu}Z^u, \tag{A.12}$$

and $\kappa_{r,st}$ and $\kappa_{r,stu}$ are defined in (A.9). Let $r_1, \ldots, r_k$ denote indices each with $q = \sum_{j=2}^3 \binom{q+j-1}{j}$ components, $E(\mathbf{U}_{\bar{r}}\mathbf{U}_{\bar{s}}) = \mathbf{\Sigma}_{\bar{r}\bar{s}}$, $E(U_r U_s) = \Sigma_{rs} = \kappa_{rs}$, $E(\mathbf{U}_{\underline{r}}\mathbf{U}_{\underline{s}}) = \mathbf{\Sigma}_{\underline{r}\underline{s}}$, $\kappa_{\bar{r}_1 \ldots \bar{r}_k} = CUM(\mathbf{U}_{\bar{r}_1}, \ldots, \mathbf{U}_{\bar{r}_k})$, the kth-order multivariate cumulant of $\mathbf{U}_{\bar{r}}$, and $h_{\bar{r}_1 \ldots \bar{r}_k}(\mathbf{u}) = (-1)^k (\partial^k/\partial \mathbf{U}_{\bar{r}_1} \ldots \partial \mathbf{U}_{\bar{r}_k})\phi_{\bar{q}, \mathbf{\Sigma}_{\bar{r}\bar{s}}}(\mathbf{U}_{\bar{r}})$. Given the continuity of (A.12) and the independence of U_r from U_{rs} and U_{rst} we have that, for $\mathbf{Z}^3_{\bar{r}} \in \chi^{-1}(\mathbf{B})$, $\gamma_{\mathbf{U}_{\bar{r}}}(\mathbf{B}) = \gamma_{\mathbf{Z}^3_{\bar{r}}}[\chi^{-1}(\mathbf{B})]$ is given by

$$\gamma_{\mathbf{U}_{\bar{r}}}(\mathbf{B}) = \int_{\mathbf{B}} \phi_{q,\Sigma_{rs}}(U_r)\phi_{q,\Sigma_{\underline{rs}}}(\mathbf{U}_{\underline{r}})\{1 + [\kappa_{\bar{r}}h_{\bar{r}}(\mathbf{u}) + \kappa_{\overline{rst}}h_{\overline{rst}}(\mathbf{u})/6]/n^{1/2}$$

$$+ [(\kappa_{\bar{r}}\kappa_{\bar{s}} + \kappa_{\overline{rs}})h_{\overline{rs}}(\mathbf{u})/2 + (\kappa_{\bar{r}}\kappa_{\overline{stu}}/6 + \kappa_{\overline{rstu}}/24)h_{\overline{rstu}}(\mathbf{u})$$

$$+ \kappa_{\overline{rst}}\kappa_{\overline{uvw}}h_{\overline{rstuvw}}(\mathbf{u})/72]/n\} \, d\mathbf{u}$$

$$= \int_{\mathbf{B}} \varsigma_n(\mathbf{u}) \, d\mathbf{u}. \tag{A.13}$$

Consider the following continuous transformation: $\mathbf{X}_{\bar{r}} = \psi(\mathbf{U}_{\bar{r}})$,

$$\psi : X_r = g(U_r), \qquad X_{rs} = U_{rs}, \qquad X_{rst} = U_{rst} \tag{A.14}$$

for

$$g(U_r) = \kappa^{rs}U_s + (\kappa^{rs}U_{st}U^t - \kappa^{rs}\kappa_{s,\,tv}U^tU^v + \kappa^{rst}U_sU_t)/n^{1/2}$$

$$+ (\kappa^{rs}\kappa^{tu}U_{st}U_{uv}U^v + 2\kappa^{rs}\kappa^{tu}\kappa_{u,vw}U_{st}U^vU^w + \kappa^{rs}\kappa^{tu}\kappa_{s,\,tv}\kappa_{u,wz}U^vU^wU^z$$

$$+ \kappa^{rs}\kappa^{uv}\kappa_{stu}U_{vw}U^tU^w - \kappa^{rs}\kappa^{tuv}\kappa_{s,\,tw}U_uU_vU^w + \kappa^{rs}U_{stu}U^tU^u/2$$

$$+ \kappa^{rs}\kappa_{s,\,tuv}U^tU^uU^v/2 - \kappa^{rs}\kappa^{tuv}U_{st}U_uU_v - 2\kappa^{rst}\kappa_{t,uv}U_sU^uU^v$$

$$+ 2\kappa^{rs}\kappa^{uvw}\kappa_{stu}U^tU_vU_w - \kappa^{rstu}U_sU_tU_u)/n,$$

where $U^r = \kappa^{rs}U_s$, $\kappa^{rst} = \kappa^{rr'}\kappa^{ss'}\kappa^{tt'}\kappa_{r's't'}$, and $\kappa^{rstu} = \kappa^{rr'}\kappa^{ss'}\kappa^{tt'}\kappa^{uu'}\kappa_{r's't'u'}$. By (A.7) there exists a $C > 0$ such that on the set $S_n^{\mathbf{U}} = \{\mathbf{U} : \|\mathbf{U}_{\bar{r}}\| < C(\log n)^{1/2}\}$ the transformation ψ of (A.14) is a C^∞ diffeomorphism on $S_n^{\mathbf{U}}$ onto its image. Set $\gamma_{\mathbf{U}_{\bar{r}}}[\cdot] = \gamma_{\mathbf{U}_{\bar{r}}}[\psi^{-1}(B \times \Re^q)]$; clearly

$$\gamma_{\mathbf{U}_{\bar{r}}}[\cdot] = \int_{\psi^{-1}(B \times \Re^q)} \varsigma_n(\mathbf{u}) \, d\mathbf{u} = \int_{S_n^{\mathbf{U}} \cap \psi^{-1}(B \times \Re^q)} \varsigma_n(\mathbf{u}) \, d\mathbf{u} + o(n^{-1})$$

$$= \int_{\psi(S_n^{\mathbf{U}}) \cap (B \times \Re^q)} \varsigma_n[\psi^{-1}(\mathbf{x})]|J_{\mathbf{x}}| \, d\mathbf{x} + o(n^{-1}),$$

where $|J_{\mathbf{x}}|$ is the Jacobian of the transformation (A.14). Because $\varsigma_n[\psi^{-1}(\mathbf{x})]|J_{\mathbf{x}}|$ can be expressed as

$$\gamma_{\mathbf{U}_{\bar{r}}}[\cdot] = \int_{\psi(S_n^{\mathbf{U}}) \cap (B \times \Re^q)} \phi_{q,\Sigma_{rs}}(X_r)\phi_{\underline{q},\Sigma_{\underline{rs}}}(\mathbf{X}_{\underline{r}})\left[1 + \sum_{j=1}^2 q_j(\mathbf{x})/n^{j/2}\right] d\mathbf{X}_{\bar{r}} + o(n^{-1})$$

$$= \int_{B} \phi_{q,\Sigma_{rs}}(X_r)\left\{\int_{\Re^q} \phi_{\underline{q},\Sigma_{\underline{rs}}}(\mathbf{X}_{\underline{r}})\left[1 + \sum_{j=1}^2 q_j(\mathbf{x})/n^{j/2}\right] d\mathbf{X}_{\underline{r}}\right\} dX_r + o(n^{-1}),$$

for some polynomials $q_j(\cdot)$, we can integrate $\mathbf{X}_{\underline{r}}$ out, obtaining

$$\gamma_{\mathbf{U}_{\bar{r}}}[\cdot] = \int_{B} \phi_{q,\Sigma_{rs}}(X_r)\left[1 + \sum_{j=1}^2 q_j'(X_r)/n^{j/2}\right] dX_r + o(n^{-1})$$

$$= \int_{B} \varsigma_n(X_r) \, dX_r + o(n^{-1}) \tag{A.15}$$

for some polynomials $q_j'(\cdot)$. Thus by (A.15) we have that for $\lambda_n^r \in B^\varepsilon$ with $\varepsilon = Cn^{-3/2}(\log n)^2$

$$\sup_{B \in \mathcal{B}} \left| \Pr(\lambda^r \in B) - \int_B \varsigma_n(X_r)\, dX_r \right| \leq \sup_{B \in \mathcal{B}} \int_{(\partial B)^\varepsilon} |\varsigma_n(X_r)|\, dX_r + o(n^{-1})$$

$$= O(\varepsilon) + o(n^{-1}).$$

It remains to show that the density $\phi_{q, \kappa_{rs}}(x)[1 + \sum_{j=1}^2 p_j(x)/n^{j/2}]$ of the formal two-terms Edgeworth expansion of the DLE λ^r obtained by the delta method coincides with $\varsigma_n(X_r)$. Because λ^r is a maximum likelihood type estimator, its cumulants of order bigger than 4 are $o(n^{-1})$, whence $\sup_{\|\eta\|<1} |E\{\exp[\iota\eta^r(Z^r + \Lambda_n^r)]\} - \hat\varsigma_n(\eta)| = o(n^{-1})$, where $\hat\varsigma_n(\cdot)$ is the Fourier transform of $\varsigma_n(\cdot)$ defined in (A.15). By Cauchy's estimates for derivatives of analytic functions (Bhattacharya and Rao, 1976, Lemma 9.2), we then have that all the derivatives of $E\{\exp[\iota\eta^r(Z^r + \Lambda_n^r)]\}$ and $\hat\varsigma_n(\eta)$ differ by $o(n^{-1})$ at $\eta = 0$, implying that the two expansions are identical. ∎

Proof of Theorem 2. A Taylor expansion of W about the DLE λ^r and (A.7) show that on the same set S_n^c of Theorem 1 $W = \lambda^r\lambda^s(\partial^2 W_\lambda)p_n(Z_{r_1 \ldots r_k}, \lambda^r) + \zeta_{4n}$, where $p_n(\cdot)$ is a quartic polynomial in λ^r and the random variable ζ_{4n} is such that $\Pr(|\zeta_{4n}| > Cn^{-3/2}(\log n)^{5/2}) = o(n^{-1})$. By a further Taylor expansion of W about the normalized deviation $n^{1/2}\eta^r = \eta^r = (\lambda^r - \lambda_n^r)$ and the second inequality in (A.7), we have that on S_n^c with $\Pr(S_n^c) = 1 - o(n^{-1})$ W admits the following stochastic expansion,

$$W = \kappa_{rs}\lambda^r\lambda^s/2 - [\partial^2 W_\lambda/2 + E(\partial^3 W_\lambda)\eta^t - E(\partial^3 W_\lambda)\lambda^t/3]\lambda^r\lambda^s/n^{1/2}$$

$$- [\partial^3 W_\lambda \eta^t - \partial^3 W_\lambda \lambda^t/3 + E[\partial^4 W_\lambda]\eta^t\eta^u/2 - E(\partial^4 W_\lambda)\lambda^t\eta^u/3$$

$$+ E[\partial^4 W_\lambda(\theta_0)]\lambda^t\lambda^u/12]\lambda^r\lambda^s/n, \tag{A.16}$$

uniformly over compact subsets of $\widetilde{\lambda^r}$. Define the scaled arrays

$$Z_{r_1 \ldots r_k} = \sum (\kappa^{r_1 s_1})^{1/2} f_{s_1}(z_i, \theta_0) \ldots (\kappa^{r_k s_k})^{1/2} f_{s_k}(z_i, \theta_0)/n^{1/2} - n^{1/2}\kappa_{r_1 \ldots r_k},$$

where $\kappa_{r_1 \ldots r_k} = E[(\kappa^{r_1 s_1})^{1/2} f_{s_1}(z, \theta_0) \ldots (\kappa^{r_k s_k})^{1/2} f_{s_k}(z, \theta_0)]$. Substituting the stochastic expansion $\lambda^r - \lambda_n^r = Z_r + \Lambda_n^r + o_p(n^{-1})$ (where Λ_n^r is as in (A.11)) in (A.16), some algebra shows that the stochastic expansion for the ELR statistic W under a sequence of local alternatives has signed square root W_r (i.e., $W_r W_r + o_p(n^{-1}) = W$) given by

$$W_r = Z_r + \widetilde{\lambda^r} + (\kappa_{rst} Z_s Z_t/3 - Z_{rs} Z_s/2 - Z_{rs}\widetilde{\lambda^s}/2 + 2\kappa_{rst} Z_s \widetilde{\lambda^t}/3 + \kappa_{rst}\widetilde{\lambda^s}\widetilde{\lambda^t}/3)/n^{1/2}$$

$$+ [Z_{rst} Z_s Z_t/3 + 3Z_{rs} Z_{st} Z_t/8 - 5\kappa_{stu} Z_{rs} Z_t Z_u/6$$

$$- (\kappa_{rstu}/4 - 4\kappa_{rst'}\kappa_{t'tu}/9)Z_s Z_t Z_u - 5\kappa_{stu} Z_{rs} Z_t \widetilde{\lambda^u}/3 + 2Z_{rst} Z_s \widetilde{\lambda^t}/3$$

$$+ 3Z_{rs} Z_{st}\widetilde{\lambda^t}/8 - (3\kappa_{rstu}/4 - 4\kappa_{rst'}\kappa_{t'tu}/3)Z^s Z^t\widetilde{\lambda^u} + Z_{rst}\widetilde{\lambda^s}\widetilde{\lambda^t}/3$$

$$- (3\kappa_{rstu}/4 - 4\kappa_{rst'}\kappa_{t'tu}/3)Z^s\widetilde{\lambda^t}\widetilde{\lambda^u} - 5\kappa_{stu} Z_{rs}\widetilde{\lambda^t}\widetilde{\lambda^u}/6 - \widetilde{\lambda^r}\tau/2$$

$$- [2]\delta^{st} Z_r Z_s\widetilde{\lambda^t}/2 + [3]\delta^{st}(Z_r Z_s + 2Z_r\widetilde{\lambda^s} + \widetilde{\lambda^r}\widetilde{\lambda^s})\widetilde{\lambda^t}/3$$

$$- [2]\delta^{st} Z_r\widetilde{\lambda^s}\widetilde{\lambda^t}/2 - (\kappa_{rstu}/4 - 4\kappa_{rst'}\kappa_{t'tu}/9)\widetilde{\lambda^s}\widetilde{\lambda^t}\widetilde{\lambda^u}]/n. \tag{A.17}$$

To prove (10), we first show that W_r admits a valid Edgeworth expansion in the sense of Bhattacharya and Ghosh (1978). Because W_r is a function of the first three DL derivatives, let $\gamma_{\mathbf{Z}_{\bar{r}}^3}(\mathbf{B})$ denote the measure corresponding to the characteristic function of the vector $\mathbf{Z}_{\bar{r}}^3$ (as in the proof of Theorem 1) and consider the same continuous transformation defined in (A.12), $\mathbf{U}_{\bar{r}} = \chi(\mathbf{Z}_{\bar{r}}^3)$ where

$$\chi : U_r = Z_r^{\tilde{\lambda}}, \qquad U_{rs} = -Z_{rs} + \kappa_{r,st} Z_t, \qquad U_{rst} = 2Z_{rst} - 2\kappa_{r,stu} Z_u, \tag{A.18}$$

and $Z_r^{\tilde{\lambda}} = Z_r + \widetilde{\lambda^r}$. Clearly (A.18) implies that $\gamma_{\mathbf{U}_{\bar{r}}}(\mathbf{B}) = \gamma_{\mathbf{Z}_{\bar{r}}^3}[\chi^{-1}(\mathbf{B})]$ can be expressed using the same Edgeworth measure as in (A.13), the only difference being that the normal distribution appearing in the leading term has now nonzero mean $\widetilde{\lambda^r}$. Proceeding now as in the proof of Theorem 1, define $\mathbf{X}_{\bar{r}} = \psi(\mathbf{U}_{\bar{r}})$

$$\psi : X_r = g_{\tilde{\lambda}}(U_r), \qquad X_{rs} = U_{rs}, \qquad X_{rst} = U_{rst}, \tag{A.19}$$

where

$$g_{\tilde{\lambda}}(U_r) = U_r + (U_{rs} U_s - \kappa_{r,st} U_s U_t)/n^{1/2}$$

$$+ [U_{rst} U_s U_t/6 + \kappa_{r,stu} U_s U_t U_u/6 - U_{rst} \widetilde{\lambda^s} \widetilde{\lambda^t}/6 - \kappa_{r,stu} U_s \widetilde{\lambda^t} \widetilde{\lambda^u}/6$$

$$+ 3U_{rs} U_{st} U_t/8 - 3\kappa_{s,tu} U_{rs} U_t U_u/4 + 3\kappa_{r,st} \kappa_{t,uv} U_s U_u U_v/4$$

$$+ 5\kappa_{stu} U_{rs} U_t U_u/6 - 5\kappa_{r,st} \kappa_{suv} U_t U_u U_v/6$$

$$- (\kappa_{rstu}/4 - 4\kappa_{rst'} \kappa_{t'tu}/9) U_s U_t U_u + \widetilde{\lambda^r} \tau/2 + (9\kappa_{rstu} - 16\kappa_{rst'} \kappa_{t'tu})$$

$$\times U_s \widetilde{\lambda^t} \widetilde{\lambda^u}/18 + [3](\delta^{st} U_r \widetilde{\lambda^s} \widetilde{\lambda^t} - \delta^{rs} \delta^{tu} \widetilde{\lambda^s} \widetilde{\lambda^t} \widetilde{\lambda^u})/3]/n.$$

On the set $S_n^{\mathbf{U}} = \{\mathbf{U} : \|\mathbf{U}_{\bar{r}}\| < C(\log n)^{1/2}\}$, the transformation ψ defined in (A.19) is a C^∞ diffeomorphism on $S_n^{\mathbf{U}}$ onto its image. Then, as in the proof of Theorem 1, the measure $\gamma_{\mathbf{U}_{\bar{r}}}[\psi^{-1}(B \times \Re^q)]$ corresponding to the characteristic function of the inverse transformation of (A.19) can be expressed as in (A.15) for some polynomials $q_j''(\cdot)$ whose degree is even or odd according to j. Integrating out $\mathbf{X}_r$, the joint expansion for $\mathbf{U}_{\bar{r}}$ is then reduced to that of $W_r - \widetilde{\lambda^r} \in B$. It remains to calculate the formal Edgeworth expansion of $W_r - \widetilde{\lambda^r}$. Using the delta method, some algebra shows that the approximate cumulants for W_r are

$$k_r = \widetilde{\lambda^r} + (k_r^2/n^{1/2} + k_r^3/n) + o(n^{-1}),$$

$$k_{r,s} = \delta^{rs} + (k_{r,s}^2/n^{1/2} + k_{r,s}^3/n) + o(n^{-1}),$$

$$k_{r,s,t} = k_{r,s,t}^3/n + o(n^{-1}), \qquad k_{r,s,t,u} = k_{r,s,t,u}^4/n + o(n^{-1}),$$

$$k_{r_1,\dots,r_l} = o(n^{-1}) \qquad l \geq 5,$$

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

$$k_r^2 = -\kappa_{r,s,s}/6 + \kappa_{rst}\,\widetilde{\lambda^s}\,\widetilde{\lambda^t}/3, \qquad k_{r,s}^2 = (2\kappa_{rst} - \kappa_{r,s,t})\,\widetilde{\lambda^t}/3,$$

$$k_r^3 = (9\kappa_{rs,st} + 16\kappa_{r,sst} - 18\kappa_{rsst})\,\widetilde{\lambda^t}/24 - q\widetilde{\lambda^r}/6 + \widetilde{\lambda^r}/6$$

$$+ [-(4\kappa_{r,t,t} + 3\kappa_{r,tt})\delta^{sv} + (9\kappa_{r,st} - 8\kappa_{r,s,t})\delta^{tv}]\kappa_{svw}\,\widetilde{\lambda^w}/18$$

$$+ (4\kappa_{rst'}\kappa_{t'tu}/9 - \kappa_{rstu}/4)\,\widetilde{\lambda^s}\,\widetilde{\lambda^t}\,\widetilde{\lambda^u} + \widetilde{\lambda^r}\tau/2,$$

$$k_{r,s}^3 = (\kappa_{rs,tt} + 2\kappa_{r,stt} - \kappa_{r,s,t,t})/4$$

$$+ (9\kappa_{r,tu}\kappa_{s,tu} + 4\kappa_{rtu}\kappa_{stu} - 24\kappa_{r,tu}\kappa_{stu} - \kappa_{r,t,u}\kappa_{s,t,u})/36$$

$$+ (27\kappa_{r,tt}\kappa_{s,uu} + 32\kappa_{rtt}\kappa_{suu} - 60\kappa_{r,tt}\kappa_{suu})/36$$

$$+ (8\kappa_{r,stu} + 3\kappa_{rs,tu} - 18\kappa_{rstu})\,\widetilde{\lambda^t}\,\widetilde{\lambda^u}/12$$

$$+ (8\kappa_{rst}\kappa_{tuv} + 16\kappa_{rtu}\kappa_{stv} + \kappa_{r,t,u}\kappa_{stv} - 18\kappa_{r,tu}\kappa_{stv})\,\widetilde{\lambda^u}\,\widetilde{\lambda^v}/9,$$

$$k_{r,s,t}^3 = \kappa_{r,s,t} - [3]\kappa_{r,st} + 2\kappa_{rst} + [3(\kappa_{r,s,tu} - \kappa_{rstu}) - \kappa_{r,s,t,u}]\widetilde{\lambda^u}/2$$

$$+ [3][4(9\kappa_{r,s,u} + 32\kappa_{rsu} - 39\kappa_{r,su})\kappa_{tvw} + 3(2\kappa_{r,s,u} - 3\kappa_{r,su})\kappa_{t,vw}]\delta^{uv}\,\widetilde{\lambda^w}/36,$$

$$k_{r,s,t,u}^4 = \kappa_{r,s,t,u} + 2[4]\kappa_{r,stu} + [3]\kappa_{rs,tu} - [6]\kappa_{r,s,tu} - 6\kappa_{rstu}$$

$$+ [4](\kappa_{r,s,t} + 2\kappa_{r,st} - [2]\kappa_{r,st})\kappa_{uvv}/3$$

$$- (3[4]\kappa_{r,s,t} + 4[6]\kappa_{rst} - 6[6]\kappa_{r,st})\kappa_{u,vv}/6$$

$$+ (2\kappa_{rtv} - [3]\kappa_{r,tv} + \kappa_{r,t,v})\kappa_{suv}. \tag{A.20}$$

Applying now the Bartlett-type identities (5) to (A.20), it is straightforward to verify that $\kappa_{r,s,t} - [3]\kappa_{r,st} + 2\kappa_{rst} = 0$ and $k_{r,s,t,u}^4 = 0$. Thus, as in the conclusion of Theorem 1, for the class $\mathcal{B}$ of Borel subsets in $\mathfrak{R}^q$ satisfying (8), the formal Edgeworth expansion of $\widetilde{W}_r = W_r - \widetilde{\lambda^r}$,

$$\sup_{B \in \mathcal{B}}\left| \Pr(\widetilde{W}_r \in B) - \sum_{j=0}^{2} n^{-j/2} \int_B \phi_q(x)p_j'(x)\,dx \right| = o(n^{-1}), \tag{A.21}$$

is valid in the sense of Bhattacharya and Ghosh (1978). The validity of expansion (10) follows by verifying that the conditions (a)–(c) of a theorem in Chandra and Ghosh (1980, p. 173) that we will denote as Theorem CG hold. By (A.17), $W = Z_r Z_r p_n(Z_{r_1 \ldots r_k}, \widetilde{\lambda^r})$ for $1 \le |k| \le 3$ and $p_n(\cdot)$ is a polynomial; therefore condition (a) of Theorem CG holds—in particular $\mathbf{A}_s$ (iv) (see p. 172) holds on the same set S_n^c of Theorem 1. Condition (b) of Theorem CG is a Cramér condition as assumed in DL3, whereas condition (c) of Theorem CG holds by the validity of (A.21). Expansion (10) is therefore valid in the sense of Chandra and Ghosh (1980). It remains to calculate the polynomials p_{jk} in (10). By (A.21) and the linear transformation $T : y_r = x_r + \widetilde{\lambda^r}$, we have that on $S_c = \{y_r : y_r y_r \le c\}$ for any $c \in [c_0, \infty)$ with $c_0 \ge 0$

$$\Pr(W \le c | H_n) = \int_{S_c} \{1 + [\partial k_r^2 + \partial^2 k_{r,s}^2/2]/n^{1/2} + [\partial k_r^3 + \partial^2(k_{r,s}^3 + k_r^2 k_s^2)/2$$

$$+ \partial^3(k_{r,s,t}^3/6 + k_r^2 k_{s,t}^2/2) + \partial^4 k_{r,s}^2 k_{t,u}^2/8]/n\}\phi_q(y)\,dy_r, \tag{A.22}$$

where $\partial^k = \partial^k(\cdot)/\partial y_{r_1} \dots \partial y_{r_k}$ and the cumulants k_r^j, $k_{r,s}^j$ ($j = 2,3$), and $k_{r,s,t}^3$ are defined in (A.20). Integrating (A.22) and using (A.5) yields

$$\Pr(W \le c \mid H_n) = G_{q,\tau}(c) + P_1/n^{1/2} + P_2/n + o(n^{-1}),$$

where

$$P_1 = [2(\kappa_{rss} - \kappa_{r,s,s} + \kappa_{rst}\,\widetilde{\lambda^s}\,\widetilde{\lambda^t})\nabla G_{q,\tau}(c) + (2\kappa_{rst} - \kappa_{r,s,t})\widetilde{\lambda^s}\,\widetilde{\lambda^t}\nabla^2 G_{q,\tau}(c)]\widetilde{\lambda^r}/6,$$

$$
\begin{aligned}
P_2 = \{&(9\kappa_{rs,st} + 16\kappa_{r,sst} - 18\kappa_{rsst})\widetilde{\lambda^r}\,\widetilde{\lambda^t}/24 - q\tau/6 + \tau/6 + \tau^2/2\\
&+ [-(4\kappa_{r,t,t} + 3\kappa_{r,tt})\delta^{sv} + (9\kappa_{r,st} - 8\kappa_{r,s,t})\delta^{tv}]\kappa_{svw}\widetilde{\lambda^r}\,\widetilde{\lambda^w}/18\\
&+ (4\kappa_{rst'}\kappa_{t'tu}/9 - \kappa_{rstu}/4) \times \widetilde{\lambda^r}\,\widetilde{\lambda^s}\,\widetilde{\lambda^t}\,\widetilde{\lambda^u} + B^{rs}\delta^{rs}/2\\
&+ (8\kappa_{r,stu} + 3\kappa_{rs,tu} - 18\kappa_{rstu})\delta^{rs}\widetilde{\lambda^t}\,\widetilde{\lambda^u}/24\\
&+ (8\kappa_{rst}\kappa_{tuv} + 16\kappa_{rtu}\kappa_{stv} + \kappa_{r,t,u}\kappa_{stv} - 18\kappa_{r,tu}\kappa_{stv})\delta^{rs}\widetilde{\lambda^u}\,\widetilde{\lambda^v}/18\\
&+ (\kappa_{r,t,t}\kappa_{s,v,v} + 4\kappa_{rtu}\kappa_{svw}\widetilde{\lambda^t}\,\widetilde{\lambda^u}\,\widetilde{\lambda^v}\,\widetilde{\lambda^w} - 4\kappa_{r,t,t}\kappa_{svw}\widetilde{\lambda^v}\,\widetilde{\lambda^w})\delta^{rs}/72\}\nabla G_{q,\tau}(c)\\
&+ \{B^{rs}\widetilde{\lambda^r}\,\widetilde{\lambda^s}/2 + (8\kappa_{r,stu} + 3\kappa_{rs,tu} - 18\kappa_{rstu})\widetilde{\lambda^r}\,\widetilde{\lambda^s}\,\widetilde{\lambda^t}\,\widetilde{\lambda^u}/24\\
&\quad + (8\kappa_{rst}\kappa_{tuv} + 16\kappa_{rtu}\kappa_{stv} + \kappa_{r,t,u}\kappa_{stv} - 18\kappa_{r,tu}\kappa_{stv})\widetilde{\lambda^r}\,\widetilde{\lambda^s}\,\widetilde{\lambda^u}\,\widetilde{\lambda^v}/18\\
&\quad + (\kappa_{r,t,t}\kappa_{s,v,v} + 4\kappa_{rtu}\kappa_{svw}\widetilde{\lambda^t}\,\widetilde{\lambda^u}\,\widetilde{\lambda^v}\,\widetilde{\lambda^w} - 4\kappa_{r,t,t}\kappa_{svw}\widetilde{\lambda^v}\,\widetilde{\lambda^w})\widetilde{\lambda^r}\,\widetilde{\lambda^s}/72\\
&\quad + [3](-\kappa_{r,u,u} + 2\kappa_{ruv}\widetilde{\lambda^u}\,\widetilde{\lambda^v})(2\kappa_{stw} - \kappa_{s,t,w})\delta^{rs}\widetilde{\lambda^t}\,\widetilde{\lambda^w}/36\\
&\quad + [3][3][4(9\kappa_{r,s,u} + 32\kappa_{rsu} - 39\kappa_{r,su})\kappa_{tvw}\\
&\qquad\qquad + 3(2\kappa_{r,s,u} - 3\kappa_{r,su})\kappa_{t,vw}]\delta^{rs}\delta^{uw}\widetilde{\lambda^t}\,\widetilde{\lambda^w}/216\\
&\quad + [3][3(\kappa_{r,s,tu} - \kappa_{rstu}) - \kappa_{r,s,t,u}]\delta^{rs}\widetilde{\lambda^t}\,\widetilde{\lambda^u}/12\\
&\quad + [3](4\kappa_{rtv}\kappa_{suw} - 4\kappa_{rtv}\kappa_{s,u,w} + \kappa_{r,t,v}\kappa_{s,u,w})\delta^{rs}\delta^{tu}\widetilde{\lambda^v}\,\widetilde{\lambda^w}/72\}\nabla^2 G_{q,\tau}(c)\\
&+ \{6(-\kappa_{r,u,u} + 2\kappa_{ruv}\widetilde{\lambda^u}\,\widetilde{\lambda^v})(2\kappa_{stw} - \kappa_{s,t,w})\widetilde{\lambda^r}\,\widetilde{\lambda^s}\,\widetilde{\lambda^t}\,\widetilde{\lambda^w}\\
&\quad + 18[3(\kappa_{r,s,tu} - \kappa_{rstu}) - \kappa_{r,s,t,u}]\widetilde{\lambda^r}\,\widetilde{\lambda^s}\,\widetilde{\lambda^t}\,\widetilde{\lambda^u}\\
&\quad + [3][4(9\kappa_{r,s,u} + 32\kappa_{rsu} - 39\kappa_{r,su})\kappa_{tvw}\\
&\qquad\qquad + 3(2\kappa_{r,s,u} - 3\kappa_{r,su})\kappa_{t,vw}]\delta^{uw}\widetilde{\lambda^r}\,\widetilde{\lambda^s}\,\widetilde{\lambda^t}\,\widetilde{\lambda^w}\\
&\quad + 3[6](4\kappa_{rtv}\kappa_{suw} - 4\kappa_{rtv}\kappa_{s,u,w} + \kappa_{r,t,v}\kappa_{s,u,w})\delta^{rs}\widetilde{\lambda^t}\,\widetilde{\lambda^u}\,\widetilde{\lambda^v}\,\widetilde{\lambda^w}\}\nabla^3 G_{q,\tau}(c)/216\\
&+ (4\kappa_{rtv}\kappa_{suw} - 4\kappa_{rtv}\kappa_{s,u,w} + \kappa_{r,t,v}\kappa_{s,u,w})\widetilde{\lambda^r}\,\widetilde{\lambda^s}\,\widetilde{\lambda^t}\,\widetilde{\lambda^u}\,\widetilde{\lambda^v}\,\widetilde{\lambda^w}\nabla^4 G_{q,\tau}(c)/72,
\end{aligned}
\tag{A.23}
$$

with

$$
\begin{aligned}
B^{rs} = 2\kappa_{r,stt} - 3\kappa_{rstt}/2 &+ (9\kappa_{r,tu}\kappa_{s,tu} + 4\kappa_{rtu}\kappa_{stu} - 24\kappa_{r,tu}\kappa_{stu}\\
&- \kappa_{r,t,u}\kappa_{s,t,u} + 27\kappa_{r,tt}\kappa_{s,uu} + 32\kappa_{rtt}\kappa_{suu} - 60\kappa_{r,tu}\kappa_{stu})/36,
\end{aligned}
$$

from which (10) and (11) follow after some algebra, noting that $\kappa_{r,stt} = \kappa_{rstt}$ and $\kappa_{r,st} = \kappa_{rst}$. ■

Proof of Corollary 3. By (essentially) the same arguments of Theorem 2, it is possible to show that, except on a set S_n with $\Pr(S_n) = o(n^{-2})$, W_r admits a stochastic

expansion that depends on the first four DL derivatives $\mathbf{Z}_{\tilde{r}}^4$ and that there exists a transformation ψ' similar to the one given in (A.19) such that the measure corresponding to the characteristic function of $(\psi')^{-1}$ can be expressed as the product of a multivariate normal and some polynomials $q_j'''(\cdot)/n^{j/2}$ for $j = 1,\ldots,4$. Then, the same integration argument giving (A.15) in the proof of Theorem 1 shows that W_r has a valid Edgeworth expansion up to $o(n^{-2})$ whose coefficients cannot be written down explicitly. Thus, using the moment condition and Cauchy's estimates for derivatives of analytic functions (Bhattacharya and Rao, 1976, Lemma 9.2) it is possible to show that this expansion agrees with the formal four-terms Edgeworth expansion $\sum_{j=0}^{4} \phi_q(x)p_j(x)/n^{j/2}$, i.e., for the class $\mathcal{B}$ of Borel subsets in $\Re^q$ satisfying (8), the following expression,

$$\sup_{B \in \mathcal{B}} \left| \Pr(W_r \in B) - \sum_{j=0}^{4} n^{-j/2} \int_B \phi_q(x)p_j(x)\,dx \right| = o(n^{-2}), \tag{A.24}$$

is a valid Edgeworth expansion in the sense of Bhattacharya and Ghosh (1978). Thus to show (12) it suffices to calculate $\int_{S_c} \phi_q(x)p_j(x)\,dx$ for $j = 1,\ldots,4$, where S_c is a sphere in $\Re^q$ with radius $c^{1/2}$ for some $c \in [c_0,\infty)$. Integrating over the sphere S_c, using (A.5) (with $\gamma = 0$) and the standard argument based on the oddness/evenness property of the Hermite polynomials appearing in the Edgeworth expansion of W (Barndorff-Nielsen and Hall, 1988), gives the following Edgeworth expansion,

$$\sup_{c \in [c_0,\infty)} |\Pr(W \le c) - G_q(c) - B[G_{q+2}(c) - G_q(c)]/(2n)| = O(n^{-2}), \tag{A.25}$$

whose validity in the sense of Chandra and Ghosh (1979) follows, using the same arguments of Theorem 2, by the validity of (A.24). Let $\nabla G_q(\cdot) = G_{q+2}(\cdot) - G_q(\cdot)$; noting that $\nabla G_q(\cdot) = -2g_{q+2}(\cdot)$ and $g_{q+2}(c)/g_q(c) = c/q$, (12) follows immediately from (A.25). To prove (13) notice that by a Taylor expansion

$$G_q[c(1 + B/(nq))] = G_q(c) + Bcg_q(c)/(nq) + O(n^{-2}), \tag{A.26}$$

whence inserting (A.26) in (12) yields

$$\sup_{c \in [c_0,\infty)} |\Pr(W^B \le c) - G_q[c(1 + B/(nq))] + Bcg_q(c)/(nq)| = O(n^{-2}),$$

from which (13) follows.   ∎

Proof of Theorem 4. We first show that $\hat{B} \xrightarrow{P} B$ for any $\hat{\theta} \xrightarrow{P} \theta_0$. Let $f_{R_k}(z,\theta) = f_{r_1}(z,\theta)\ldots f_{r_k}(z,\theta)$ and notice that by GS1(ii) $f_{R_k}(z,\theta)$ is continuous at θ_0 with probability 1 (w.p.1) for $k = 2,3,4$; the consistency of $\hat{\theta}$ implies that there exists a $\delta_n \to 0$ such that $\|\hat{\theta} - \theta_0\| \le \delta_n$. Let $\Delta_f(\theta) = \sup_{\|\theta - \theta_0\| \le \delta_n} |f_{R_k}(z,\theta) - f_{R_k}(z,\theta_0)|$; continuity of $f_{R_k}(z,\theta)$ at θ_0 shows that $\Delta_f(\theta) \to 0$ with probability 1. The dominance condition GS2(i) implies that $\Delta_f(\theta) \le 2\sup_{\theta \in \mathcal{N}} |f_{R_k}(z,\theta)|$ whence by dominated convergence $E[\Delta_f(\theta)] \to 0$. By Markov inequality $\Pr(|\sum \Delta_f(\theta)/n| > \varepsilon) \le E[\Delta_f(\theta)]/\varepsilon \to 0$ for all $\varepsilon > 0$. By the weak law of large numbers $\sum f_{R_k}(z_i,\theta_0)/n \xrightarrow{P} E[f_{R_k}(z,\theta_0)]$; moreover

$$\left| \sum f_{R_k}(z_i,\hat{\theta})/n - \sum f_{R_k}(z_i,\theta_0)/n \right| \le \sum |f_{R_k}(z_i,\hat{\theta}) - f_{R_k}(z_i,\theta_0)|/n$$

$$\le \sum \Delta_{f_i}(\theta)/n \xrightarrow{P} 0.$$

Therefore, the triangle inequality implies that $|\sum f_{R_k}(z_i, \hat{\theta})/n - E[f_{R_k}(z, \theta_0)]| = o_p(1)$. The consistency of $\hat{B}$ follows then by the continuity of tensor multiplication and matrix inversion. To establish (17) we use the following mean value expansion:

$$\hat{B} = \hat{B}_0 + \partial\hat{B}_{\bar{\theta}}/\partial\theta^r (\widehat{\theta^r} - \theta_0^r), \tag{A.27}$$

where $\hat{B}_0 = \hat{B}|_{\theta=\theta_0}$, $\hat{B}_{\bar{\theta}} = \hat{B}|_{\theta=\bar{\theta}}$ and $\|\theta_0\| \leq \|\bar{\theta}\| \leq \|\hat{\theta}\|$. Because by the weak law of large numbers $\hat{B}_0 \xrightarrow{p} B$, it suffices to show that $\partial\hat{B}_{\bar{\theta}}/\partial\theta^r \xrightarrow{p} E(\partial\hat{B}_0/\partial\theta^r)$. This follows by the same argument used to establish $\hat{B} \xrightarrow{p} B$. Because by GS2(ii) the derivative $\partial f_{R_k}(z, \theta)/\partial\theta^r$ is continuous at θ_0, it follows that $\Delta_{\partial f}(\theta) = \sup_{\|\theta-\theta_0\|\leq\delta_n}|\partial f_{R_k}(z,\theta)/\partial\theta^r - \partial f_{R_k}(z,\theta_0)/\partial\theta^r| \to 0$ (w.p.1), so that by GS2(iii), $\bar{\theta} \xrightarrow{p} \theta_0$, and the triangle inequality we have that

$$\left| n^{-1} \sum \partial f_{R_k}(z_i, \bar{\theta})/\partial\theta^r - E[\partial f_{R_k}(z, \theta_0)/\partial\theta^r] \right| \xrightarrow{p} 0,$$

i.e., $n^{-1} \sum \partial f_{R_k}(z_i, \bar{\theta})/\partial\theta = O_p(1)$. By the $n^{1/2}$ consistency of $\hat{\theta}$, the stochastic expansion (A.27) shows that $\hat{B} = B + O_p(n^{-1/2})$, whence by the delta method $\Pr(W^{\hat{B}} \leq c) = \Pr(W^B \leq c) + O(n^{-3/2})$. The conclusion of the theorem follows by the oddness/evenness property of the Hermite polynomials as in Corollary 3. ∎

Let χ_n denote the observed sample. In what follows we write $\Pr(\chi_n) = o(n^{-1})$ to denote the event "the observed sample χ_n is contained in a set of probability $o(n^{-1})$."

Proof of Theorem 5. We first show (18). Recall that $\partial^k W_{\lambda^*}^* = \sum f_{r_1}(z_i^*, \theta_0) \ldots f_{r_k}(z_i^*, \theta_0)/(1 + \lambda^{*r} f_r(z_i^*, \theta_0))^k$ and that the BDLE $\widehat{\lambda_n^*}$ solves $0 = \partial W_{\lambda^*}^*(\theta_0)$. By (A.7) and Markov inequality we have that for all $\varepsilon > 0$

$$\Pr^*(|\partial^k W_0^* - E^*[\partial^k W_0^*]| > \varepsilon n) = o(n^{-1}) \quad \text{for } 1 \leq |k| \leq 3,$$

$$\Pr^* \left(\max_{|k|=4} \sup_{\|\bar{\lambda}-\hat{\lambda}_n\|\leq\tau} |\partial^k W_{\bar{\lambda}}^*|/n > C \right) = o(n^{-1}), \tag{A.28}$$

except if $\Pr(\chi_n) = o(n^{-1})$. The triangle and Markov inequalities show that $|E^*(\partial^k W_0^*)| \leq |E^*(\partial^k W_0^*) - E(\partial^k W_0)| + |E(\partial^k W_0)|$ and $\Pr(|E^*(\partial^k W_0^*) - E(\partial^k W_0)| > \varepsilon) = o(n^{-1})$, implying that for $2 \leq |k| \leq 3$ except if $\Pr(\chi_n) = o(n^{-1})$

$$\Pr^*(|E^*\partial^k W_0^*| > C_\varepsilon) = o(n^{-1}), \qquad \Pr^*(|\partial^k W_0^*/n| > C_\varepsilon) = o(n^{-1}). \tag{A.29}$$

Furthermore the moderate deviations estimates of Bhattacharya and Rao (1976, Corollary 17.12) show that for $\alpha = 6$

$$\Pr^*(\|\partial W_0^*/n - E^*(\partial W_0^*)\| > Cn^{-1/2}(\log n)^{1/2}) = o(n^{-1}), \tag{A.30}$$

except if $\Pr(\chi_n) = o(n^{-1})$. Let S_n^* denote the union of sets where (A.28)–(A.30) hold, and consider a Taylor expansion of $0 = \partial W_{\lambda^*}^*(\theta_0)/n$ about the DLE $\hat{\lambda}_n$

$$0 = \partial W_{\hat{\lambda}_n}^*/n + \sum_{k=2}^{3} \partial^k W_{\hat{\lambda}_n}^*(\lambda_n^* - \hat{\lambda}_n)^{R_{k-1}}/(k-1)!n + R_n^*(\bar{\lambda}), \tag{A.31}$$

where $R_n^*(\bar{\lambda}) \leq C\|\lambda_n^* - \widehat{\lambda_n}\|^3 \max_{|k|=4} \sup_{\|\bar{\lambda}-\hat{\lambda}_n\|\leq\tau} |\partial^k W_{\bar{\lambda}}|/n$. A further mean value expansion of each of the $\partial^k W_{\hat{\lambda}_n}^*$'s in (A.31) about 0, the uniform moment condition **BDL**(i) and (ii), and (A.28)–(A.30) show that outside the set S_n^* with bootstrap proba-

bility $\Pr^*(S_n^*) = o(n^{-1})$, we can rewrite (A.31) as $\lambda_n^* - \hat{\lambda}_n = g_n(\lambda_n^* - \hat{\lambda}_n)$ where the continuous function $g_n(\cdot): \mathfrak{R}^q \to \mathfrak{R}^q$ satisfies $\|g_n(u)\| \le Cn^{-1/2}(\log n)^{1/2}$ for all $\|u\| \le Cn^{-1/2}(\log n)^{1/2}$. Thus as in the proof of Theorem 1, Brower's fixed point theorem implies that there exists a sequence of random vectors $\widehat{\lambda_n^*}$ such that

$$\Pr^*(\|\widehat{\lambda_n^*} - \hat{\lambda}_n\| \le Cn^{-1/2}(\log n)^{1/2}) = 1 - o(n^{-1}),$$

except if $\Pr(\chi_n) = o(n^{-1})$. To prove (19) we follow the same arguments of Theorem 1. First, by repeated application of (A.28)–(A.30) it is possible to show that, except if $\Pr(\chi_n) = o(n^{-1})$, $n^{1/2}(\widehat{\lambda_n^{*r}} - \widehat{\lambda_n^r})$ has the following stochastic expansion:

$$n^{1/2}(\widehat{\lambda_n^{*r}} - \widehat{\lambda_n^r}) = [-E^*(\partial^2 W_{\hat{\lambda}_n}^*)]^{-1}\partial W_{\hat{\lambda}_n}^*/n^{1/2} + \widehat{\Lambda_n^{*r}} + o_{p^*}(n^{-1}), \tag{A.32}$$

where $\widehat{\Lambda_n^{*r}}$ is an $\mathfrak{R}^q$-valued function of the BDL derivatives (and their expectations) evaluated at $\hat{\lambda}_n$ whose norm is $O_{p^*}(n^{-1}(\log n)^{3/2})$ except if $\Pr(\chi_n) = o(n^{-1})$ (i.e., $\widehat{\Lambda_n^{*r}}$ is the bootstrap analogue of Λ_n^r given in (A.11)). Next by Lemma 2 of Babu and Singh (1984) the characteristic function of $\widehat{\mathbf{Z}_{\bar{r}}^{3*}}$ satisfies with probability 1

$$\limsup_{n\to\infty} \sup_{\delta \le \|t^{\bar{r}}\| \le \exp(n\gamma)} E^*[\exp(\iota t^{\bar{r}} \widehat{\mathbf{Z}_{\bar{r}}^{3*}})] < 1 - \gamma'$$

for some positive constants δ, γ, γ', where $\widehat{\mathbf{Z}_{\bar{r}}^{3*}}$ is the $\mathfrak{R}^{\bar{q}_3}$-valued vector containing all the different BDL derivatives (up to third order) evaluated at $\hat{\lambda}_n$. Thus, as in Theorem 1, it is possible to show that except if $\Pr(\chi_n) = o(n^{-1})$ for every class $\mathcal{B}$ of Borel subsets in $\mathfrak{R}^q$ satisfying $\sup_{B\in\mathcal{B}} \int_{(\partial B)^\epsilon} \phi_{q,\kappa_{rs}}(x)\, dx = O(\epsilon)$ as $\epsilon \downarrow 0$

$$\sup_{B\in\mathcal{B}} \left| \Pr^*(n^{1/2}(\widehat{\lambda_n^*} - \hat{\lambda}_n) \in B) - \int_B \widehat{s_n^*}(x)\, dx \right| = o(n^{-1}),$$

where $\widehat{s_n^*}(\cdot) = \phi_{q,\kappa_{rs}^*}(\cdot)[1 + \sum_{j=1}^2 \widehat{p_j^*}(\cdot)/n^{j/2}]$ is the density of the empirical Edgeworth expansion obtained by replacing the population moments with their bootstrap analogue. Thus (19) follows if for all $\varepsilon > 0$

$$\Pr(|\widehat{p_j^*}(\cdot) - p_j(\cdot)| > \varepsilon) = o(n^{-1}) \qquad j = 1,2. \tag{A.33}$$

Because each bootstrap moment can be expressed as $E^*[\prod_{l=1}^k (\widehat{Z_{R_k}^*})]$ for $1 \le |k| \le 4$, and

$$\limsup_{n\to\infty} E^*\left[\prod_{l=1}^k (\widehat{Z_{R_l}^*}) \right] \le \limsup_{n\to\infty} \sum_{\lambda\in\Gamma_\tau} \prod_{l=1}^k (Z_{iR_l})/n,$$

where Z_{iR_l} is the ith component of the $\mathfrak{R}^{q^l}$-valued array of DL derivatives evaluated at an arbitrary $\lambda \in \Gamma_\tau$, (A.33) follows by verifying that for $1 \le |k| \le 4$

$$\Pr\left(\left| \sum \prod_{l=1}^k (\hat{Z}_{iR_k}) - \sum \prod_{l=1}^k (Z_{iR_l}) \right| > n\varepsilon \right) = o(n^{-1}),$$

$$\Pr\left(\left| \sum \prod_{l=1}^k [Z_{iR_k} - E(Z_{R_l})] \right| > n\varepsilon \right) = o(n^{-1}). \tag{A.34}$$

The first limit follows by a mean value expansion about 0. To see this notice that by the second equality in (A.7), we have

$$\Pr\left(\left|n^{-1}\sum \partial\left[\prod_{l=1}^{k} Z_{iR_l} - E\left(\prod_{l=1}^{k} Z_{R_l}\right)\right]/\partial\lambda\right| > \varepsilon\right) = o(n^{-1}),$$

so that by the triangle inequality $\Pr(|n^{-1}\sum \partial(\prod_{l=1}^{k}\bar{Z}_{iR_k})/\partial\lambda^r| > C_\varepsilon) = o(n^{-1})$ for some $C_\varepsilon < \infty$, where $\bar{Z}_{iR_k} = Z_{r_1\dots r_k}|_{\lambda=\bar{\lambda}}$ and $\|\bar{\lambda}\| \le \|\hat{\lambda}_n\|$. Moreover by (6) and Markov inequality $\Pr(\|\hat{\lambda}_n\| > \varepsilon) = o(n^{-1})$ whence

$$\Pr\left(\left|n^{-1}\sum_i \left(\partial\prod_{l=1}^{k}\bar{Z}_{iR_l}/\partial\lambda^r\right)\widehat{\lambda_n^s}\right| > \varepsilon\right) = o(n^{-1}).$$

The second limit in (A.34) follows by Rosenthal and Markov inequality. ∎

Proof of Theorem 6. Let $Z^*_{r_1\dots r_k} = \sum f_{r_1}(z^*_i, \theta_0)\dots f_{r_1}(z^*_i, \theta_0)$ and $\kappa^*_{r_1\dots r_k} = E^*(Z^*_{r_1\dots r_k}/n)$ and notice that $\partial^k W^*_0 = (-1)^{k-1}(k-1)(Z^*_{r_1\dots r_k} - \kappa^*_{r_1\dots r_k})$, where $W^*_0 = W^*_{\lambda^*}|_{\lambda^*=0}$ is the "restricted" BDL evaluated at 0. As in Theorem 5 we have that, except if $\Pr(\chi_n) = o(n^{-1})$, for $1 \le |k| \le 3$

$$\Pr^*(|Z^*_{r_1\dots r_k}/n - \kappa^*_{r_1\dots r_k}| > Cn^{-1/2}(\log n)^{1/2}) = o(n^{-1}),$$

whereas by Markov inequality $|\kappa^*_{r_1\dots r_k} - \kappa_{r_1\dots r_k}| = o_p(n^{-1})$, so that

$$\Pr^*(|Z^*_{r_1\dots r_k}/n - \kappa_{r_1\dots r_k}| > Cn^{-1/2}(\log n)^{1/2}) = o(n^{-1}), \tag{A.35}$$

except if $\Pr(\chi_n) = o(n^{-1})$. A Taylor expansion of W^* about 0 gives

$$W^* = \sum_{k=1}^{3} (-1)^{k-1}(Z^*_{r_1\dots r_k} - \kappa^*_{r_1\dots r_k})\widehat{\lambda_n^{*r_1}}\dots\widehat{\lambda_n^{*r_k}}/k + R^*_n(\lambda^*), \tag{A.36}$$

where $R^*_n(\lambda^*) \le C\|\widehat{\lambda_n^*}\|^4 \max_{|k|=4}\sup_{\|\lambda^*\|\le\tau}|\partial^k W^*_{\lambda^*}|/n$. As in Theorem 5 it is possible to show that $n^{1/2}\widehat{\lambda_n^{*r}}$ admits a stochastic expansion of the form

$$n^{1/2}\widehat{\lambda_n^{*r}} = \kappa^{*rs}(Z^*_s - \kappa^*_s)/n^{1/2} + \Lambda^{*r}_n + o_{p^*}(n^{-1}), \tag{A.37}$$

where κ^{*rs} is the matrix inverse of κ^*_{rs} and Λ^{*r}_n is as in (A.32). Repeated applications of (A.35) show that $\Pr^*(\|n^{1/2}\widehat{\lambda_n^*}\| > C\log n^{1/2}) = o(n^{-1})$ except if $\Pr(\chi_n) = o(n^{-1})$, and thus by triangle inequality it follows that

$$\Pr^*(|R^*_n(\lambda^*)| > \|\widehat{\lambda_n^*}\|^4 C(1 + n^{-1/2}(\log n)^{1/2})) = o(n^{-1}).$$

Substituting (A.37) into (A.36) shows that, except if $\Pr(\chi_n) = o(n^{-1})$, the BELR admits the following stochastic expansion:

$$W^* = \kappa^{*rs}(Z^*_r - \kappa^*_r)(Z^*_s - \kappa^*_s)p_n(Z^*_{r_1\dots r_k})/n + o_{p^*}(n^{-1}),$$

where $p_n(\cdot)$ is a polynomial in $Z^*_{r_1\dots r_k}$ with coefficients depending on $\kappa^*_{r_1\dots r_k}$. Let $Z^{*c}_r = Z^*_r - \kappa^*_r$; as in Theorem 2, we can obtain the signed square root decomposition of $W^* = \kappa^{*rs}W^*_r W^*_s + o_{p^*}(n^{-1})$ with

$$W_r^* = Z_r^{*c}/n^{1/2} + [-Z_{rs}^*(\kappa^{*ss'}Z_{s'}^{*c})/2 + \kappa_{rst}^*(\kappa^{*ss'}Z_{s'}^{*c})(\kappa^{*tt'}Z_{t'}^{*c})/3]/n$$

$$+ [Z_{rst}^*(\kappa^{*ss'}Z_{s'}^{*c})(\kappa^{*tt'}Z_{t'}^{*c})/3 + 3Z_{rs}^*Z_{st}^*(\kappa^{*tt'}Z_{t'}^{*c})/8$$

$$- 5Z_{stu}^*Z_{rs}^*(\kappa^{*tt'}Z_{t'}^{*c})(\kappa^{*uu'}Z_{u'}^{*c})/6 - (\kappa_{rstu}^*/4 - 4\kappa_{rst'}^*\kappa_{t'tu}^*/9)$$

$$\times (\kappa^{*ss'}Z_{s'}^{*c})(\kappa^{*tt'}Z_{t'}^{*c})(\kappa^{*uu'}Z_{u'}^{*c})]/n,$$

which shows that, except if $\Pr(\chi_n) = o(n^{-1})$, W_r^* is a function of the first three BDL derivatives outside a set S_n^* with bootstrap probability $\Pr^*(S_n^*) = o(n^{-1})$. Thus, by the same arguments of Theorem 5 it can be shown that W_r^* admits a valid Edgeworth expansion in the sense of Bhattacharya and Ghosh (1978) except if $\Pr(\chi_n) = o(n^{-1})$. Therefore following the same steps of Theorem 2 it can be shown that, except if $\Pr(\chi_n) = o(n^{-1})$, W^* admits the following valid (in the sense of Chandra and Ghosh, 1979) Edgeworth expansion

$$\sup_{c \in [c_0, \infty)} |\Pr^*(W^* \le c) - G_q(c) - B^*[G_{q+2}(c) - G_q(c)]/(2n)| = o(n^{-1}), \qquad \textbf{(A.38)}$$

where $B^* = \kappa_{rstu}^*\kappa^{*rs}\kappa^{*tu}/2 - \kappa_{rtv}^*\kappa_{suw}^*\kappa^{*rs}\kappa^{*tu}\kappa^{*vw}/3$ is the bootstrap Bartlett correction, i.e., (16) at $\theta = \theta_0$. By (A.34) we have

$$\Pr(|B^* - B| > \varepsilon) = o(n^{-1}), \qquad \textbf{(A.39)}$$

and thus combining (A.38), (A.39), and (A.25) yields

$$\sup_{c \in [c_0, \infty)} |\Pr^*(W^* \le c) - \Pr(W \le c) - (B^* - B)[G_{q+2}(c) - G_q(c)]/(2n)| = o(n^{-1}),$$

$$\sup_{c \in [c_0, \infty)} |\Pr^*(W^* \le c) - \Pr(W \le c) - o_p(n^{-2})| = o(n^{-1}),$$

except if $\Pr(\chi_n) = o(n^{-1})$. ■

Proof of Corollary 7. As in the proof of Corollary 3 it is possible to show that the BELR admits a valid Edgeworth expansion up to $o(n^{-2})$ except if $\Pr(\chi_n) = o(n^{-2})$. Thus by the results of Corollary 3, Theorem 6, and the generalized Cornish–Fisher expansion formula of Hill and Davies (1968) it follows that the asymptotic expansions for the α point of W and W^* are

$$c_{\alpha,n} = c_\alpha(1 + B/(nq)) + O(n^{-2}),$$

$$c_{\alpha,n}^* = c_\alpha(1 + B^*/(nq)) + O_p(n^{-2}),$$

respectively, where $c_\alpha = \Pr(\chi_q^2 \ge c_\alpha) = \alpha$. Therefore using the delta method, a Taylor expansion, and the Edgeworth expansion of W, we have

$$\Pr(W > c_{\alpha,n}^*) = \Pr(W - (c_{\alpha,n}^* - c_{\alpha,n}) > c_{\alpha,n})$$

$$= \Pr(W - O_p(n^{-2}) > c_\alpha + c_\alpha B/(nq))$$

$$= 1 - G_q[c_\alpha(1 + B/(nq))] + c_\alpha Bg(c_\alpha)/(nq) + O(n^{-2})$$

$$= \alpha + O(n^{-2}),$$

which shows that effectively the BELR accomplishes the Bartlett correction automatically. ■