Nonparametric inference by quasi-likelihood methods [1]

Richard H. Spady[2]

Abstract

Two techniques for nonparametric inference concerning statistical functionals are examined as quasi-likelihood methods. Empirical likelihood, introduced by Owen (1988), and the least-favorable family construction of DiCiccio and Romano (1990) are considered as defining test statistics through maximization problems. Using first-order theory only, a statistic derived from the latter approach outperforms the second-order (Bartlett) corrected version of the former in all of the cases considered in DiCiccio, Hall, and Romano (1991), and the fully iterated bootstrap results for variances reported in Lee and Young (1995). This suggests that coverage accuracy comparable to the computationally intensive iterated bootstrap can be achieved *without* recourse to simulation or higher–order asymptotic calculations.

Some keywords: Nonparametric inference; Empirical likelihood; Least favorable family; Bartlett correction.

This version 29 March 1996.

1. Introduction

Suppose the data x are characterized by distribution $F(x)$. It is desired to construct confidence intervals, hypothesis tests, etc. about the functional $\theta(F)$ while relying on as few assumptions concerning F as possible. In this paper we consider only cases where θ is defined by $\underset{F}{\mathrm{E}}\,(\psi(\theta)) = 0$ and consequently it is natural to estimate θ by the sample analog of this condition, namely the estimating equation $\sum \psi_i(\theta) = 0$. We further assume that the data are drawn independently from F.

We consider two well–known approaches to this problem. One is the method of empirical likelihood (EL) first proposed by Owen (1988). The other is the construction of least favorable families (LFF) as given e.g. in DiCiccio and Romano (1990). The methods are similar in that they test the hypothesis $\theta = \theta_0$ in a sample characterized by empirical distribution F_n (i.e. one with atoms of $(1/n)$ at each of sample points) by constructing a second distribution $F_n(\theta_0)$ with support only on the sample points and new atoms p_i such that $\underset{p}{\mathrm{E}}\,(\psi(\theta_0)) = 0$; the two methods differ in their choice of p. In

[1] I am grateful for discussions with Andrew Chesher, Stephen Corcoran, Anthony Davison, and Guido Imbens. This research was supported under the Analysis of Large and Complex Datasets initiative of the ESRC.

[2] Nuffield College, Oxford, OX1 1NF UK.

EL, the hypothesis test is based on the likelihood ratio (LR) statistic $2\ \{\sum \log(1/n) - \sum \log(p_i)\}$ and on refinements thereof. In LFF, the hypothesis can also be tested by the corresponding LR, though for theoretical and practical reasons, a test on the dual or tilting parameters that employs the robustified or "sandwich" estimate of their variance is preferable.

The main point of this paper is to demonstrate the surprisingly good properties of this latter test. The argument is based entirely on the first-order asymptotic theory of misspecified or quasi- likelihood functions, as found e.g. in White (1982). (In this context a quasi-likelihood function is an function that is aymptotically maximized at $\theta(F)$; the function is typically the likelihood function for some data generating process, though not necessarily the one in hand.) In a well–known paper, DiCiccio, Hall and Romano (1991) derived the Bartlett correction for the empirical likelihood ratio (ELR) test, thus producing a test with second-order properties that, as one might expect, improves upon the uncorrected ELR test in simple examples with small sample sizes. Corcoran (1995) has shown that empirical likelihood is the only construction based on a reweighting of the data subject to a moment condition that gives rise to a Bartlett–correctable LR–style test. This suggests a special role for empirical likelihood, but as demonstrated in Corcoran, Davison, and Spady (1995), the ELR test is consistently characterized by curved QQplots in even the simplest examples with fairly large samples, so that no multiplicative correction as suggested by a Bartlett correction theory is likely to be effective. This point is again demonstrated here. A further obvious drawback to any Bartlett correction procedure is that it requires complex computations and the "true" correction depends upon parameters which themselves need to be estimated, so that the feasible test, based on the estimated correction, will have a variable correction factor. Typically (though serendipity cannot be ruled out) this will tend to give somewhat poorer results than those that could be obtained with the theoretical correction. In contrast, the statistic proposed here produces remarkably straight QQplots, though conventional asymptotic theory is apparently again an unreliable guide to this behavior: the same methods that predict straight QQplots for the ELR test predict curved QQplots for our statistic, and in both cases the observed means are very different than those obtained from asymptotic approximations.

The procedure we advocate performs at least as well as either of the Bartlett-adjusted ELR tests in all of the examples given in DiCiccio, Hall and Romano (1991). We have further checked our results against those for the iterated bootstrap for the variance examples considered in Lee and Young (1995). We find that our statistic equals or exceeds the (two–sided) coverage accuracy of the iterated bootstrap, and thus exceeds the performance of the approximations to the fully–iterated bootstrap developed in Lee and Young (1995). Since all the methods considered in Lee and Young, like the Bartlett–corrected ELR test, have coverage error $O(n^{-2})$, we have an additional example of an $O(n^{-1})$ coverage error interval outperforming $O(n^{-2})$ methods.

The remainder of this paper is organized as follows. In the next section we provide a self-contained derivation of the two methods in the no-nuisance parameter case that emphasizes the quasi-likelihood interpretation. Section 3 derives the specific tests we employ, while Section 4 replicates the Monte-Carlo study of DiCiccio, Hall and Romano (1991), and compares some simulations with those of Lee and Young (1995). In Section 5 we provide an asymptotic analysis of the χ^2 test statistics and their cumulants, etc, while Section 6 provides a similar analysis of the behavior of the corresponding signed square root statistics. We find that the conventional aysmptotic analysis provides no apparent explanation for the superior performance of the statistic proposed here. Section 7 is a conclusion.

2. Derivation of the empirical likelihood and least favorable family approaches.

2.1 Empirical likelihood.

Given a sample with points x_i, $i = 1, ..., n$ empirical likelihood constructs corresponding probabilities p_i so as to maximize $\sum \log(p_i)$ subject to $\sum p_i \psi_i(\theta_0) = 0$, where θ_0 is the hypothesized value of θ. (Confidence intervals are constructed by inverting hypothesis tests.) DiCiccio and Romano (1990) point out that EL minimizes the discrepancy between F_n and $F_n(\theta_0)$ as measured by the Kullback-Leibler Information Criterion (KLIC), taken with resepect to F_n. That is, if F and G are distributions with corresponding densities f and g, then the KLIC of G from F, denoted $D(G; F)$, is given by:

$$\underset{F}{\mathrm{E}}\,(\log f - \log g) = \begin{cases} \int \log(f(x)/g(x)) f(x) dx & \textit{if continuous} \\ \\ \sum\limits_i \{\log f(x_i) - \log g(x_i)\} f(x_i) & \textit{if discrete} \end{cases} \tag{1}$$

Using $\hat{w}$ to denote $\{1/n, ..., 1/n\}$ and $p = \{p_1, ..., p_n\}$ to denote the probabilities (and by a slight abuse the corresponding discrete distributions) that we seek to define $F_n(\theta_0)$, the Lagrangean for the constrained optimization problem posed by EL is:

$$\begin{aligned} \mathcal{L} &= D(p; \hat{w}) + a(\sum p_i - 1) + b \cdot (\sum p_i \psi_i(\theta_0)) \qquad (2) \\ &= \sum \hat{w}_i \{\log \hat{w}_i - \log p_i\} + a(\sum p_i - 1) + b \cdot (\sum p_i \psi_i(\theta_0)) \end{aligned}$$

where b is a vector of $k = \dim(\theta)$ Lagrangean multipliers and $\cdot$ indicates a dot product. The first-order condition for p_i is given by:

$$\frac{\partial \mathcal{L}}{\partial p_i} = -\frac{\hat{w}_i}{p_i} + a + b \cdot \psi_i(\theta_0) = 0. \tag{3}$$

Multiplying each of the n equations in (3) by p_i and summing over i, one obtains (since $\sum p_i \psi_i(\theta_0) = 0$) $a = 1$. Substituting this back into (3) multiplied by p_i and using

$\hat{w}_i = (1/n)$, the solution is given by:

$$p_i = \frac{1}{n}\left(\frac{1}{1 + b \cdot \psi_i(\theta_0)}\right) \tag{4}$$

Examining (4) we see that the solution involves k unknown parameters. Using $\sum p_i \psi_i(\theta_0) = 0$ these are found by solving the k equations:

$$\sum \frac{1}{n}\left(\frac{\psi_i(\theta_0)}{1 + b \cdot \psi_i(\theta_0)}\right) = 0. \tag{5}$$

Notice that (5) serves to describe the first-order conditions of the (essentially unconstrained) optimization problem

$$\max_{\{b\}} \; Q(b) \equiv \sum \frac{1}{n} \log(1 + b \cdot \psi_i(\theta_0)) \tag{6}$$

One can confirm by differentiation of (6) twice that $Q(b)$ is globally concave in b for b meeting the restriction $1 + b \cdot \psi_i(\theta_0) > 0$ for all i, which is essential if p_i is to be positive. Thus solutions of (5) will be (at least local) maxima of (6). We return to a further discussion of the nature of (6) after the next section.

2.2 The least favorable family construction.

An obvious alternative to choosing p to minimize $D(p; \hat{w})$ is to choose it to minimize $D(\hat{w}; p)$. This seems to be both conceptually and practically superior; see the discussion below. We will call this the least favorable family construction (LFF).

The Lagrangean function corresponding to (2) for LFF is thus:

$$\begin{aligned} \mathcal{L} &= D(\hat{w}; p) + a(\textstyle\sum p_i - 1) + t \cdot (\textstyle\sum p_i \psi_i(\theta_0)) \\ &= \textstyle\sum p_i \{\log p_i - \log \hat{w}_i\} + a(\textstyle\sum p_i - 1) + b \cdot (\textstyle\sum p_i \psi_i(\theta_0)) \end{aligned} \tag{7}$$

so that the first-order condition for p_i is now:

$$\frac{\partial \mathcal{L}}{\partial p_i} = 1 + \log p_i - \log \hat{w}_i + a + b \cdot \psi_i(\theta_0) = 0. \tag{8}$$

Equating (8) for two observations i and j, rearranging and exponentiating, one obtains:

$$\frac{p_i}{p_j} = \frac{\hat{w}_i}{\hat{w}_j} e^{-b \cdot (\psi_i(\theta_0) - \psi_j(\theta_0))}$$

Or, writing t for $-b$,

$$\frac{p_i}{p_j} = \frac{\hat{w}_i e^{t \cdot \psi_i(\theta_0)}}{\hat{w}_j e^{t \cdot \psi_j(\theta_0)}} \tag{9}$$

Thus exploiting the constraint that the probabilities sum to one, we have:

$$p_i = \frac{e^{t \cdot \psi_i(\theta_0)} \hat{w}_i}{\sum_{j=1}^{n} e^{t \cdot \psi_j(\theta_0)} \hat{w}_j}. \tag{10}$$

An objective function that produces $\sum p_i \psi_i(\theta_0) = 0$ (with p_i as defined in (10)) as its first-order condition is:

$$\max_{\{t\}} \; q^*(t) \equiv -\log\left(\sum_{i=1}^{n} e^{t \cdot \psi_i(\theta_0)} \hat{w}_i\right) \tag{11}$$

Notice that $q^*(t)$ is the negative of the cumulant generating function (cgf) of $\psi(\theta_0)/n \equiv \sum \psi_i(\theta_0)/n$ when the data is distributed as $\hat{w}_i$, i.e. as F_n. For $\hat{t}$ solving $\sum p_i \psi_i(\theta_0) = 0$, $-q^*(\hat{t}) \equiv K(\hat{t})$ equals $D(\hat{w}; p)$ as can be confirmed by direct calculation. Since (11) calls for the minimization of a cgf, the convexity of cgf's guarantees that $\hat{t}$ always exists and is unique.

For both theoretical and practical reasons it is easier to work with the moment generating function (mgf) $M(t)$ corresponding to $K(t)$, so that the objective function becomes

$$\max_{\{t\}} \; q(t) \equiv -\left(\sum_{i=1}^{n} e^{t \cdot \psi_i(\theta_0)} \hat{w}_i\right) \equiv -M(t) \tag{12}$$

Using (12) in place of (11) does not change the solution for t while allowing the objective function to be written as the sum of observation contributions and simplifying the expressions for derivatives.

2.3 Discussion.

The two methods restrict their attention to distributions that (1) have support only on the sample, and (2) have the property E $\psi(\theta_0) = 0$. Of this set, EL uses for its calculations the element that is closest, or is most likely to occur, *if* the true process is the sample replicated indefinitely. LFF uses that element on the basis of which the sample would be regarded as "most likely."

Since the property $\mathrm{E}\psi(\theta_0) = 0$ is by hypothesis true, it seems more sensible to calculate on the basis of some distribution for which it is indeed true. Both methods do this. However, in choosing the distribution itself it would seem this principle should also apply: EL is selecting an element on the basis of a probability distribution, namely the empirical distribution, that does not have this property.

As is often the case in quasi–maximum likelihood estimation, the objective function employed has an interpretation as the likelihood function for some data–generating process. For the LFF construction, the function $q(t)$ is the likelihood function for censored type-1 extreme value observations. Since our data is not assumed to have been generated by this process, it is apparent that estimation of the covariance matrix of $\sqrt{n}\hat{t}$ will require 'robustification' via White's (1982) procedure, i.e. the 'sandwich' estimator.

3. The Tests

The specific tests we discuss are the empirical likelihood ratio test (and its Bartlett correction) and the χ^2 test that the tilt parameter t estimated from (12) is zero that is formed from $\hat{t}'V^{-1}\hat{t}$, where V is the sandwich estimator of the variance of $\sqrt{n}\hat{t}$. Since $\hat{t}$ is an estimate of a Lagrange multiplier, it inherits a number of parameterization invariance properties, and can be viewed as an LM or score test; equally it can be interpreted as a Wald test in that it compares a parameter difference to an estimated variance. We will refer to it as a dual parameter test, following the nomenclature of Myklund (1995), who develops a 'dual likelihood' theory for martingales that corresponds to empirical likelihood in the independent case.

3.1 The empirical likelihood ratio test.

Our first test statistic, the empirical likelihood ratio test (ELRT), refers $-2(\sum \log w_i - \sum \log p_i)$ to the χ^2 distribution. Direct calculation from (4) yields:

$$\begin{aligned} \mathrm{S}_1 &= -2(-n \log n - (-n \log n - \sum \log(1 + b \cdot \psi_i(\theta_0))) \\ &= 2 \sum \log(1 + b \cdot \psi_i(\theta_0)) \end{aligned} \tag{13}$$

where b is estimated from (6).

3.2 The dual parameter test for the LFF construction.

Our second test statistic S_2 is a test of the dual or exponential tilt parameter in (12). Let $q_i \equiv -e^{t \cdot \psi_i(\theta_0)}\hat{w}_i$; then $\hat{t}'HJ^{-1}H\hat{t}$ is asymptotically χ^2 where

$$\begin{aligned} J &= \sum (\partial q_i/\partial t)(\partial q_i/\partial t)' \\ H &= -\sum \partial^2 q_i/\partial t \partial t' \end{aligned}$$

Here the variance V of $\sqrt{n}\hat{t}$ is estimated by $n\, H^{-1}JH^{-1}$. The typical elements of H and J can be written:

$$\begin{aligned} H_{jk} &= \sum_i e^{t \cdot \psi_i} \psi_{ij} \psi_{ik} \hat{w}_i \\ J_{jk} &= \sum_i (e^{t \cdot \psi_i})^2 \psi_{ij} \psi_{ik} \hat{w}_i^2 \end{aligned}$$

An examination of H and J reveals that these are, up to proportionality constants that cancel in the calculation of V, the relevant *expected* quantities with respect to p, so that the sample estimate of V coincides with the (estimated) population value for V. (This is of course not typically true in parametric contexts.)

A dual parameter test of S_2–type can also be defined for EL. In the simulation examples below, its performance is similar to that of the ELR test, and so these are not reported. Curiously, 'robustification' of the covariance matrix of EL's dual parameter b has no effect: using $Q_i \equiv \log(1 + b \cdot \psi_i(\theta))$,

$$\frac{\partial Q_i}{\partial b}\frac{\partial Q_i}{\partial b}' + \frac{\partial^2 Q_i}{\partial b \partial b'} \equiv 0,$$

so that the analogs of H and J are *always* the same. Thus, EL obeys the first Bartlett identity in a peculiar degenerate fashion, which may explain why it, among all such methods, is the only one that is Bartlett–correctable.

The cases considered below are all scalar, in which case S_2 can be written:

$$\mathsf{S}_2 = \frac{\hat{t}^2 \left(\sum_i e^{t\psi_i} \; \psi_i^2 \right)^2}{\sum_i (e^{t\psi_i} \; \psi_i)^2} \tag{14}$$

4. Monte Carlo results

DiCiccio, Hall, and Romano (1991) (DHR) presented a small Monte-Carlo study examining the effect of Bartlett correction on the coverage of the confidence regions generated by the empirical likelihood ratio (ELR) test. We have replicated their experiment (based on 5,000 replications) while adding to it the coverage levels of confidence regions generated by S_2. The three examples considered are the estimation of the mean in samples drawn from the normal, χ^2_1, and t with 5 degrees of freedom distributions at sample sizes 10 and 20, 20 and 40, and 15 and 30, respectively. The results are presented in Table 1, with DHR's results appearing in italics after the relevant entry; the S_2 results are labeled "Exponential Tilt (Wald)." The rows labeled "Empirical Likelihood" are based on S_1, while the rows labeled "Theoretical Bartlett" and "Estimated Bartlett" are Bartlett corrected versions of S_1, corresponding to $\mathsf{S}_1/(1 + c/n)$ and $\mathsf{S}_1/(1+\hat{c}/n)$, respectively, where $c = (k/2 - g^2/3)$, and g and k are the standardized third and fourth moments of the underlying distribution. For $\hat{c}$, the sample estimates of k and g are used. Thus the Theoretical Bartlett rows correspond to a statistic that is not available in practice. For the three cases at hand, $c = 3/2, 29/6$, and $9/2$.

In nearly every case, S_2 either generates the best result or very nearly does. In cases where there is a discernible difference between it and the best result, the competitor is the infeasible statistic based on the true value of c. This is rather surprising, since the argument used to develop S_2 has no second-order basis, unlike the Bartlett-corrected versions of S_1. On the basis of our results, internal consistency with the other two cases, and theoretical expectation, we believe that the two Bartlett corrected rows for the normal data experiments have been inadvertently exchanged in DHR's paper. Exchanging them results in the theoretical adjustment being superior in all cases in both papers; otherwise, the empirical adjustment outperforms the theoretical in DHR's normal case only, a result that we have not seen replicated.

A closer examination of the results is afforded by Figures 1-3 which are QQplots. (A QQplot plots the value of the statistic at each quantile of the simulation against the theoretical or reference value at that quantile; the vertical bars are at the .95 and .99 levels.) The left-hand column shows shows S_1 and its (theoretical) Bartlett correction at the two sample sizes, the dashed lines corresponding to the corrected statistic. The right hand column shows the corresponding QQplots of S_2 at the two sample sizes.

From the "A" figures it is apparent that the effect of the Bartlett adjustment is not dramatic, and perhaps more importantly, from Figures 1.A and 2.A it is apparent that the Bartlett adjustment *cannot* be effective at these sample sizes, since the QQplots are not linear. In contrast, the QQplots for S_2 are not only closer to the 45° line but they are clearly straighter.

A further indication of the accuracy of intervals is given in Table 2, where the coverage of intervals based on S_2 is compared with the iterated bootstrap and two asymptotic approximations thereto developed in Lee and Young (1995). Among a wide variety of procedures with $O(n^{-2})$ coverage accuracy, Lee and Young find that the computationally intensive iterated bootstrap (which requires two levels of resampling) consistently provides the best coverage, with their approximations I_1 and I_2 (requiring one and zero levels of resampling, respectively[3]), providing the next best performance. In these examples, S_2's performance is arguably the best and superior to I_1 and I_2, while requiring substantially less computation than any of the alternatives.

5. Asymptotic Analysis

Why does S_2 perform so much better than S_1 and generally better than its Bartlett-corrected variants which are based on higher-order theory? An inspection of Table 1 and the accompanying QQplots leads one to conjecture that S_2 might itself be second-order accurate, since its accuracy is at least as good as the *theoretical* Bartlett-corrected S_1 and that it might also be susceptible to further Bartlett-style correction since its QQplots are so straight—indeed much straighter than S_1 itself, which is Bartlett-correctable. To address these issues, we provide a tentative asymptotic analysis.

[3] Both I_1 and I_2 require substantial analytical calculations for which Lee and Young provide computer programs.

Our strategy is to expand S_1 and S_2 as Taylor-series around the null value $\hat{b} = 0$ (for S_1) and $\hat{t} = 0$ (for S_2), and to revert the estimating equations defining $\hat{b}$ and $\hat{t}$. The results of the "reversion" are approximations to $\hat{b}$ and $\hat{t}$ in terms of $\mu_h \equiv \sum_{i=1}^{n} [\psi_i(\theta_0)]^h$, for $h = 1, 2, 3, 4$. These approximations are then subsituted into the approximations for S_1 and S_2. The resulting expressions are functions of the μ_h's only.

5.1 The Bartlett correction for S_1.

We first confirm DiCiccio, Hall, and Romano's result that the Bartlett correction of S_1 is $\mu_4/2\mu_2^2 - \mu_3^2/3\mu_2^3$, where the "feasible" correction uses μ_h as defined above (i.e. without dividing by n) and the "theoretical" correction uses the corresponding expectations.

A Taylor series expansion of S_1 in b gives:

$$\mathsf{S}_1 = 2b\mu_1 - b^2\mu_2 + \frac{2b^3\mu_3}{3} - \frac{b^4\mu_4}{2} + O(b^5) \tag{15}$$

The value of b that is relevant for (15) is $\hat{b}$ satisfying (5) which is

$$\sum \frac{1}{n}\left(\frac{\psi_i(\theta_0)}{1 + b \cdot \psi_i(\theta_0)}\right) = 0. \tag{16}$$

Now in general if x is scalar and the function $h(x)$ admits an expansion of the form:

$$h(x) = h_0 + ax + bx^2 + cx^3 + dx^4 + O(x^5) \tag{17}$$

then x satisfying $h(x) = 0$ is approximately

$$x = -\frac{h_0}{a} - \frac{h_0^2 b}{a^3} + \left(\frac{-2b^3}{a^5} + \frac{c}{a^4}\right) h_0^3 + \left(\frac{-5b^3}{a^7} + \frac{5bc}{a^6} - \frac{d}{a^5}\right) h_0^4 + O(h_0^5) \tag{18}$$

so that x is approximately a polynomial in h_0. (A multivariate statement of (18), though to $O(h_0^4)$ only, can be found in Koenker et. al. (1993)). Applying the scalar reversion formula to the Taylor series expansion of (5) yields

$$\hat{b} = \frac{\mu_1}{\mu_2} + \frac{\mu_1^2\mu_3}{\mu_2^3} + \mu_1^3\left(\frac{2\mu_3^2}{\mu_2^5} - \frac{\mu_4}{\mu_2^4}\right) + \mu_1^4\left(\frac{5\mu_3^3}{\mu_2^7} - \frac{5\mu_3\mu_4}{\mu_2^6} + \frac{\mu_5}{\mu_2^5}\right) + O(\mu_1^5) \tag{19}$$

Substituting the value of $\hat{b}$ from (19) into (15) gives

$$\mathsf{S}_1 = \frac{\mu_1^2}{\mu_2} + \frac{2}{3}\frac{\mu_1^3\mu_3}{\mu_2^3} + \mu_1^4\left(\frac{\mu_3^2}{\mu_2^5} - \frac{\mu_4}{2\mu_2^4}\right) + O(\mu_1^5) \tag{20}$$

A Bartlett correction of S_1 is constructed by computing $\mathrm{E}(\mathsf{S}_1)$; in the scalar case this will be expressible as $(1 + c/n)$, where c is the Bartlett factor.

It will be convenient in what follows to divide S_1 into the (weighted) sum of four terms, namely,

$$Q = \left\{\frac{\mu_1^2}{\mu_2}, \frac{\mu_1^3\mu_3}{\mu_2^3}, \frac{\mu_1^4\mu_3^2}{\mu_2^5}, \frac{\mu_1^4\mu_4}{\mu_2^4}\right\} = \left\{z^2, \frac{z^3\hat{g}}{\sqrt{n}}, \frac{z^4\hat{g}^2}{n}, \frac{z^4\hat{k}}{n}\right\}, \tag{21}$$

where we have introduced the notation:

$$z = \frac{\mu_1}{\sqrt{\mu_2}}, \quad \hat{g} = \sqrt{n}\frac{\mu_3}{\mu_2^{3/2}}, \quad \hat{k} = n\frac{\mu_4}{\mu_2^2}, \tag{22}$$

where z is $O_p(1)$ and $\hat{g}$ and $\hat{k}$ are $O_p(n^{-1/2})$ estimates of standardized skewness and kurtosis. The weights generating (20) are of course $\left\{1, \frac{2}{3}, 1, -\frac{1}{2}\right\}$. (It will turn out that S_2 has weights $\{1, 1, \frac{9}{4}, -\frac{4}{3}\}$.)

The components of Q all have denominators that are powers of μ_2. To calculate the expectation of each component in turn we take Taylor series expansions, viewing μ_2 as $n(m_2 + n^{-1}\sum(\psi_i^2 - m_2)) = n(m_2 + d)$, using the approximation:

$$\frac{x}{[n(m_2+d)]^k} = \frac{x}{(nm_2)^k}\left[1 - \frac{d}{m_2}k + \frac{1}{2}\left(\frac{d}{m_2}\right)^2 k(1+k) - \frac{1}{6}\left(\frac{d}{m_2}\right)^3 k(1+k)(2+k) + O(d^4)\right] \tag{23}$$

Since the approximation (23) is valid for small d it is natural to take m_2 as $(1/n)\mathrm{E}(\mu_2) = (1/n)\sum\psi_i^2$ and so we introduce the notation $m_h = (1/n)\mathrm{E}(\mu_h)$.

Using (23) to analyze Q_1

$$\mathrm{E}\left(\frac{\mu_1^2}{\mu_2}\right) = \left(\frac{\mu_1^2}{nm_2}\right)\left(1 - \frac{d}{m_2} + \left(\frac{d}{m_2}\right)^2 - \left(\frac{d}{m_2}\right)^3 + \ldots\right).$$

Thus

$$\begin{aligned}
\mathrm{E}\left(\frac{\mu_1^2}{nm_2}\right) &= 1 \\
\mathrm{E}\left(\frac{-\mu_1^2 d}{nm_2^2}\right) &= \frac{1}{n}\left(1 - \frac{m_4}{m_2^2}\right) + O(n^{-2}) \\
\mathrm{E}\left(\frac{\mu_1^2 d^2}{nm_2^3}\right) &= \frac{1}{n}\left(-1 + \frac{2m_3^2}{m_2^3} + \frac{m_4}{m_2^2}\right) + O(n^{-2}) \\
\mathrm{E}\left(\frac{\mu_1^2 d^3}{nm_2^4}\right) &= O(n^{-2})
\end{aligned}$$

Notice that characteristically answers are expressed in terms of the standardized moments $M_h = m_h/(m_2)^{h/2}$; for M_3 and M_4 we introduce the notation g and k respectively. Adding the components, we obtain

$$\mathrm{E}\left(\frac{\mu_1^2}{\mu_2}\right) = 1 + \frac{2g^2}{n} + O(n^{-2}). \tag{24}$$

This can be confirmed from Hall[1993, p. 73], where the term $(3/n)$ vanishes because m_1 is in our context known to be zero.

We can continue in a like fashion to evaluate the other terms in Q. For Q_1, it is required to calculate up to d^2 in (23) to obtain terms $O\left(n^{-1}\right)$; for Q_2 the requirement turns out to be to include terms in d; for Q_3 and Q_4 the term "1" is sufficient. Thus

$$\begin{aligned}
\mathrm{E}\left(\frac{\mu_1^3\mu_3}{\mu_2^3}\right) &= \mathrm{E}\left(\frac{\mu_1^3\mu_3}{(nm_2)^3}\right) + \mathrm{E}\left(\frac{-3\mu_1^3\mu_3 d}{(nm_2)^3 m_2}\right) + O(n^{-2}) \\
\mathrm{E}\left(\frac{\mu_1^3\mu_3}{(nm_2)^3}\right) &= \frac{g^2}{n} + \frac{3k}{n} + O(n^{-2}) \\
\mathrm{E}\left(\frac{-3\mu_1^3\mu_3 d}{(nm_2)^3 m_2}\right) &= \frac{-9g^2}{n} + O(n^{-2})
\end{aligned}$$

so that

$$\mathrm{E}\left(\frac{\mu_1^3\mu_3}{\mu_2^3}\right) = \frac{-8g^2}{n} + \frac{3k}{n} + O(n^{-2}) \tag{25}$$

For Q_3 we have

$$\begin{aligned}
\mathrm{E}\left(\frac{\mu_1^4\mu_3^2}{\mu_2^5}\right) &= \mathrm{E}\left(\frac{\mu_1^4\mu_3^2}{(nm_2)^5}\right) + O(n^{-2}) \qquad (26)\\
\mathrm{E}\left(\frac{\mu_1^4\mu_3^2}{(nm_2)^5}\right) &= \frac{3g^2}{n} + O(n^{-2}),
\end{aligned}$$

while Q_4 gives

$$\begin{aligned}
\mathrm{E}\left(\frac{\mu_1^4\mu_4}{\mu_2^4}\right) &= \mathrm{E}\left(\frac{\mu_1^4\mu_4}{(nm_2)^4}\right) + O(n^{-2}) \qquad (27)\\
\mathrm{E}\left(\frac{\mu_1^4\mu_4}{(nm_2)^4}\right) &= \frac{3k}{n} + O(n^{-2})
\end{aligned}$$

(The computations for Q_3 and Q_4 can be easily confirmed by writing for Q_3

$$\begin{aligned}
E(Q_3) = E\left(\frac{\mu_1^4\mu_3^2}{\mu_2^5}\right) &= E\left(\frac{z^4\hat{g}^2}{n}\right) = E\left(\frac{z^4(g + O_p(n^{-1/2}))^2}{n}\right) \\
&= E\left(\frac{z^4(g^2 + 2gO_p(n^{-1/2}) + O_p(n^{-1}))}{n}\right)
\end{aligned}$$

$$= E\left(\frac{z^4g^2}{n}\right) + O(n^{-3/2}) + O(n^{-2}) = \frac{3g^2}{n} + O(n^{-3/2})$$

with a similar argument holding for Q_4 using $\hat{k} = k + O_p(n^{-1/2})$. Notice that the more refined calculations carried out above are apparently required for Q_2 and the smaller error bounds on Q_3 and Q_4.)

Using (20) so that S_1 is approximated by $Q_1 + (2/3)Q_2 + Q_3 - (1/2)Q_4$, we compute $\mathrm{E}(\mathsf{S}_1)$ as $1(1 + 2g^2/n) + (2/3)(-8g^2/n + 3k/n) + 1(3g^2/n) - (1/2)(3k/n)$, or

$$\mathrm{E}(\mathsf{S}_1) = 1 - \frac{g^2}{3n} + \frac{k}{2n} + O(n^{-2}), \tag{28}$$

which agrees with the result of DHR for the case of scalar means.

5.2 The expectation of S_2.

Using the methods and results of the preceeding section we can easily calculate the expectation of S_2. A Taylor-series expansion of (13) yields:

$$\mathsf{S}_2 = \hat{t}^2\mu_2 + \hat{t}^4\left(\frac{\mu_3^2}{\mu_2} - \mu_4\right) + O(\hat{t}^5) \tag{29}$$

and the reversion of the estimating equation defining $\hat{t}$ gives

$$\hat{t} = -\frac{\mu_1}{\mu_2} - \frac{\mu_1^2\mu_3}{2\mu_2^3} + \mu_1^3\left(\frac{-\mu_3^2}{2\mu_2^5} + \frac{\mu_4}{6\mu_2^4}\right) + \mu_1^4\left(\frac{-5\mu_3^3}{8\mu_2^7} + \frac{5\mu_3\mu_4}{12\mu_2^6} - \frac{\mu_5}{24\mu_2^5}\right) + O(\mu_1^5) \tag{30}$$

Continuing by substituting (30) into (29), the approximation for S_2 is:

$$\mathsf{S}_2 = \frac{\mu_1^2}{\mu_2} + \frac{\mu_1^3\mu_3}{\mu_2^3} + \mu_1^4\left(\frac{9\mu_3^2}{4\mu_2^5} - \frac{4\mu_4}{3\mu_2^4}\right) + O(\mu_1^5), \tag{31}$$

which, as mentioned above, consists of a sum of the elements of Q with weights $\{1, 1, \frac{9}{4}, -\frac{4}{3}\}$. It follows then that the expectation of S_2 can be expressed as:

$$\mathrm{E}(\mathsf{S}_2) = \left(1 + \frac{2g^2}{n}\right) + \left(\frac{-8g^2}{n} + \frac{3k}{n}\right) + \frac{9}{4}\left(\frac{3g^2}{n}\right) - \frac{4}{3}\left(\frac{3k}{n}\right) + O(n^{-2})$$

$$= 1 + \left(\frac{3g^2}{4n}\right) - \frac{k}{n} + O(n^{-2}) \tag{32}$$

Note that, had the weights on Q_3 and Q_4 been 2 and -1 instead of $9/4$ and $-4/3$, then $\mathrm{E}(\mathsf{S}_2)$ would have been $1 + O(n^{-2})$, a result that would mean, if S_2 is Bartlett-correctable, its Bartlett correction is zero. The gentle reader can be sure that the author has fiddled many stages of the argument in search of this result!

5.3 The accuracy of the $\mathrm{E}(\mathsf{S}_1)$ and $\mathrm{E}(\mathsf{S}_2)$ approximations.

The difficulty, equally with both (28) and (32), is that they don't actually capture the expectations of S_1 and S_2 very well. Among the examples of Table 1, the most interesting case is the χ^2 distribution, since it is the only example where $g \neq 0$. For the χ^2 distribution $g^2 = 8$ and $k = 15$ so the Bartlett factor for S_1 is 29/6, while the corresponding factor for S_2 (which we shall hereafter also call a Bartlett factor) is -9.

To confirm our assertion that the $O(n^{-1})$ asymptotic approximations for the expectations of S_1 and S_2 are not terribly accurate at small sample sizes and to investigate the possible reasons for this, we have extended the simulations of the previous section for the χ^2 case, adding 5,000 replications for $n = 80, 120, 160$ and 15,000 replications for $n = 200$ and $n = 400$. For each sample size, we have recorded the values of each of the components of Q as well as S_1 and S_2, and two other statistics introduced below. The mean values in the simulation (and an estimate of the standard error) are reported in Table 2, together with the values that would be expected on the basis of the $O(n^{-1})$ approximations.

The first column of Table 2 reports the results for $Q_1 = \mu_1^2/\mu_2$, the square of the Studentized normal test statistic for the mean. Until $n = 120$, it is apparent that the $O(n^{-1})$ approximation underpredicts the mean value of Q_1. It is not clear that the approximation is effective until $n = 160$, where, as for $n = 200$ and 400, the agreement is very close.

For the other components of Q, agreement is even more problematic; even at $n = 200$, agreement is not obtained in view of the standard errors; at $n = 400$, Q_2 has been brought into agreement, while the discrepancies of the remaining components is decreased.

The discrepancies for the last 3 components of Q up to $n = 200$ and to some extent beyond might lead one to suspect that the calculations of expected values are erroneous. We cannot rule that out, though it would be surprising to get the expectations of the components of Q wrong and still agree with DHR's result for S_1, which we have obtained as a weighted sum of Q components.

Partly to serve as an additional check on our calculations and partly because S_3 below is of independent interest, we formulate the statistics

$$\mathsf{S}_3 = Q_1 + Q_2 + 2Q_3 - Q_4, \tag{33}$$

$$\tilde{\mathsf{S}}_2 = Q_1 + Q_2 + \frac{9}{4}Q_3 - \frac{4}{3}Q_4,$$

As noted above, S_3 has the property $\mathrm{E}(\mathsf{S}_3) = 1 + O(n^{-2})$.

Table 2's penultimate column reports results for S_3; agreement is obtained at $n = 80$, even before Q_1.

To summarize, the following points emerge:

1. For no component of Q is there close agreement between sample means of simulations and the $O(n^{-1})$ approximations at sample sizes less than 160. For some components discernible discrepancies remain at $n = 400$. In most cases the discrepancies at small and moderate sample sizes are striking.
2. Expectations of sums of components of Q are more accurately approximated than the components themselves. This effect is particularly evident for S_3.
3. As for S_1 and S_2, the former has an expectation discernibly different from 1 until $n = 400$ while the latter has expectation 1 from about $n = 80$ onwards. Neither shows good agreement with its $O(n^{-1})$ approximation, though this is perhaps not so clear for S_1.

At least three explanations are consistent with these results. The first is that our calculations are in error, and that this error is decreasing in n. This is considered in Corcoran, Davison, and Spady (1995), and the agreements cited there between the current results and those obtained independently by others suggests that this is not probable. A second explanation is that $O(n^{-2})$ effects are important throughout. This is no doubt true, but the extension of the analysis to higher–orders in Corcoran *et. al.,* and an independent analysis kindly provided by Andrew Chesher, suggests that the higher–order terms are not particularly enlightening. For example, the expectation of S_2 for the $\chi^2(1)$ example is $1 - 9/n + 3273/n^2 + O(n^{-3})$, which is not very helpful in view of the results in Table 2.

We are forced rather uncomfortably to the conclusion that the terms in the conventional asymptotic expansions are so large as to be useless for any sample sizes that are liable to be of any interest.

6. Asymptotic analysis of the corresponding signed square-root statistics.

Corresponding to each of the S_i's is a signed square root statistic T_i. In the case of S_2, it is natural to use the t-statistic that corresponds to the Wald test; for reasons that will shortly become apparent, we change the sign of this statistic and use the resulting sign convention for the other signed sqaure root statistics. Thus the statistic T_2 is given by:

$$T_2 = \frac{-\hat{t}\left(\sum_i e^{t\psi_i}\ \psi_i^2\right)}{\sqrt{\sum_i (e^{t\psi_i}\ \psi_i)^2}} \tag{34}$$

A Taylor series expansion of T_2 in t around $t = 0$ gives:

$$T_2 = -t\sqrt{\mu_2} - t^3\left(\frac{\mu_3^2}{2\mu_2^{3/2}} - \frac{\mu_4}{2\sqrt{\mu_2}}\right) + O(t^4). \tag{35}$$

Substituting $\hat{t}$ as given in (30) into (35) yields:

$$\begin{aligned} T_2 &= \frac{\mu_1}{\sqrt{\mu_2}} + \frac{\mu_1^2\mu_3}{2\mu_2^{5/2}} + \mu_1^3\left(\frac{\mu_3^2}{\mu_2^{9/2}} - \frac{2\mu_4}{3\mu_2^{7/2}}\right) + O(\mu_1^4) \qquad (36)\\ &= z + \frac{z^2}{2n^{1/2}}\hat{g} + \frac{z^3}{n}\left(\hat{g}^2 - (2/3)\hat{k}\right) + O\left(n^{-3/2}\right) \end{aligned}$$

From (36) it is apparent that T_2 approaches the usual t-statistic in large samples, and that in smaller samples it is approximated by a polynomial transformation of z with weights determined by estimates of (standardized) skewness and kurtosis.

To derive an expression for the signed square root of S_1 we take a Taylor series expansion in μ_1 of the square root of its approximation as given by (20) around $\mu_1 = 0$ so that

$$T_1 = z + \frac{z^2}{3n^{1/2}}\hat{g} + \frac{z^3}{n}\left(\frac{4}{9}\hat{g}^2 - \frac{1}{4}\hat{k}\right) + O\left(n^{-3/2}\right), \tag{37}$$

where the sign of the square root has been chosen so as to agree with the limiting sign of (36). That (37) has been obtained by a valid argument can be confirmed by applying the argument to (31) in place of (20); this produces (36). It is convenient to analyze (36) and (37) as weighted sums of four terms:

$$R = \left\{z, \frac{z^2\hat{g}}{n^{1/2}}, \frac{z^3\hat{g}^2}{n}, \frac{z^3\hat{k}}{n}\right\} = Q/z \tag{38}$$

An analysis of S_3 similar to that for S_1 gives $T_3 = R_1 + R_2/2 + (7/8)R_3 - R_4/2 + O(n^{-3/2})$.

6.1 Moments and cumulants of the T-statistics.

The first four moments or cumulants of the T statistics can be obtained from the expectation of the first four powers of sums of R terms. To obtain the expectation of a T statistic we require the expectation of each component of R. These are:

$$\mathrm{E}(z) = \frac{-g}{2n^{1/2}} + O(n^{-3/2}) \tag{39}$$

$$\mathrm{E}\left(\frac{z^2\hat{g}}{n^{1/2}}\right) = \frac{g}{n^{1/2}} + O(n^{-3/2}) \tag{40}$$

$$\mathrm{E}\left(\frac{z^3\hat{g}^2}{n}\right) = O(n^{-3/2})$$

$$\mathrm{E}\left(\frac{z^3\hat{k}}{n}\right) = O(n^{-3/2})$$

From (37,38,39,40) we have:

$$\begin{aligned}
\mathrm{E}(T_2) &= \mathrm{E}(R_1) + \mathrm{E}(R_2)/2 + O(n^{-3/2}) &=& O(n^{-3/2}) \\
\mathrm{E}(T_1) &= \mathrm{E}(R_1) + \mathrm{E}(R_2)/3 + O(n^{-3/2}) &=& \frac{-g}{6n^{1/2}} + O(n^{-3/2})
\end{aligned}$$

Thus we see that compared to the usual t-statistic, i.e. z, empirical likelihood eliminates 2/3 of the $O(n^{-1/2})$ bias, while LFF eliminates this order of bias entirely.

$\mathrm{E}(T_1^2)$ and $\mathrm{E}(T_2^2)$ were calculated in the preceding section as:

$$\mathrm{E}(T_1^2) = 1 - \frac{g^2}{3n} + \frac{k}{2n} + O(n^{-2}),$$

$$\mathrm{E}(T_2^2) = 1 + \left(\frac{3g^2}{4n}\right) - \frac{k}{n} + O(n^{-2})$$

Now T_1^3 and T_2^3 are approximately:

$$T_1^3 = z^3 + \frac{z^4\hat{g}}{n^{1/2}} + \frac{5z^5\hat{g}^2}{3n} - \frac{3z^5\hat{k}}{4n} + O_p(n^{-3/2})$$

$$T_2^3 = z^3 + \frac{3z^4\hat{g}}{2n^{1/2}} + \frac{15z^5\hat{g}^2}{4n} - \frac{2z^5\hat{k}}{n} + O_p(n^{-3/2})$$

and in fact since $\mathrm{E}(z^5) = -55g/(2n^{1/2}) + O(n^{-3/2})$ the terms in z^5 have expectations $O(n^{-3/2})$ so the terms in $\mathrm{E}(T_1^3)$ and $\mathrm{E}(T_2^3)$ that are not $O(n^{-3/2})$ or smaller are derived from R_1^3 and $R_1^2 R_2$. These are:

$$\mathrm{E}(R_1^3) = -\frac{7g}{2n^{1/2}} + O(n^{-3/2})$$

$$\mathrm{E}(R_1^2 R_2) = \mathrm{E}\left(\frac{z^4\hat{g}}{n^{1/2}}\right) = \mathrm{E}\frac{z^4\left(g + O_p(n^{-1/2})\right)}{n^{1/2}} = \frac{3g}{n^{1/2}} + O(n^{-3/2})$$

(The first calculation can be confirmed from Hall, p. 73; the error order of the second requires a detailed calculation using (24).) Thus

$$E(T_1^3) = -\frac{g}{2n^{1/2}} + O(n^{-3/2})$$

$$E(T_2^3) = \frac{g}{n^{1/2}} + O(n^{-3/2})$$

To complete our calculation of the moments and cumulants of the T_i's we need finally to have $\mathrm{E}(T_i^4)$. The values of T_1^4 and T_2^4 are:

$$T_1^4 = z^4 + \frac{4}{3}\frac{z^5\hat{g}}{n^{1/2}} + \frac{22}{9}\frac{z^6\hat{g}^2}{n} - 1\frac{z^6\hat{k}}{n} + O(n^{-3/2})$$

$$T_2^4 = z^4 + 2\frac{z^5\hat{g}}{n^{1/2}} + \frac{11}{2}\frac{z^6\hat{g}^2}{n} - \frac{8}{3}\frac{z^6\hat{k}}{n} + O(n^{-3/2})$$

The relevant combinations of R terms are:

$$E(R_1^4) = 3 + n^{-1}(28g^2 - 2k) + O(n^{-2})$$

$$E(R_1^2R_2^2) = E(R_1^3R_3) = E(n^{-1}z^6\hat{g}^2) = \frac{15g^2}{n} + O(n^{-2})$$

$$E(R_1^3R_3) = E(n^{-1}z^6\hat{k}) = \frac{15k}{n} + O(n^{-2})$$

$$E(R_1^3R_2) = E(n^{-1/2}z^5\hat{g}) = E(n^{-1/2}z^5(g + O_p(n^{-1/2})) = \frac{-50g^2}{n} + \frac{15k}{n} + O(n^{-3/2}),$$

where the last result requires detailed calculations. Thus we have:

$$E(T_1^4) = 3 - \frac{2g^2}{n} + \frac{3k}{n} + O(n^{-2})$$

$$E(T_2^4) = 3 + \frac{21}{2}\frac{g^2}{n} - \frac{12k}{n} + O(n^{-2})$$

We are now in a position to compute the cumulants of z, T_1, and T_2, using e.g. the standard formulas in Hall (p. 73) that express cumulants as functions of expectations of powers. These results are displayed in Table 3.

The notable aspects of the results of Table 3 are:

1. The skewness and kurtosis of T_1 are zero to $O(n^{-3/2})$ and $O(n^{-2})$, respectively; the bias is (1/3) that of z.
2. The bias of T_2 is zero to $O(n^{-3/2})$; the skewness is of the opposite direction of z, and one-half as large.
3. The constants and linear coefficients of all three polynomial transformations to normality (see below) are the same.
4. The polynomial transformation of T_1 to normality, is, to this order, linear.
5. All of the forgoing are more consistent with an expectation that T_1 would be the superior statistic.

6.2 The T-statistics as polynomial transformations to normality.

The usual t-statistic z converges to a standard normal distribution and from (37,38) we can see that T_1 and T_2 can be interpreted as polynomial transformations of z, and if they are to be effective, these transformations should ideally result in a statistic that is nearly normal. From arguments given in Stuart and Ord (Kendall and Stuart, vol. 1, 1988), we can calculate the polynomial transformation of z that will bring it into conformity with normality; the formulae in Stuart and Ord bring this agreement to $O(n^{-2})$. Using the terms up to $O(n^{-1})$ and the calculations of the cumulants of z from the preceding section we obtain

$$\tau(z) = T_4 = \frac{g}{6\sqrt{n}} + \left(1 + \frac{13g^2}{72n} - \frac{k}{4n}\right) z + \frac{g}{3\sqrt{n}} z^2 + \left(\frac{-g^2}{18n} + \frac{k}{12n}\right) z^3 \tag{41}$$

Notice that T_4 is defined in terms of standardized moments, not estimates of standardized moments, as the other T-statistics are.

T_4 is surprisingly effective, though its computation requires *knowledge* of g and k which is practice is not available. Moreover, its effectiveness relies on this knowledge; estimates of g and k, if substituted in (41), produce very different results.

This is illustrated in Figure 4. In panels A and D we plot chi-squared QQplots for $n = 20$ and 80 respectively, showing z^2 (which corresponds closely to the usual Studentized test-statistic) as a solid line; the dashed line, as indicated in the legend, that corresponds closely to the 45° line in both figures is T_4^2. The divergent dashed line is T_4 with $\hat{g}$ and $\hat{k}$ used in place of the true values; this statistic is denoted T_5 and is feasible, unlike T_4. *Notice that substituting estimates in $\tau(\cdot)$ is worse than substituting zero, i.e., than using z untransformed.*

The conclusion we draw from this exercise is that polynomial transformations to normality can be effective for these problems, but that an interpretation of the effectiveness of S_2 as being the square of such a transformation is elusive. While T_2 has a special bias property, it is hard to see why this should explain S_2's superior performance.

7. Conclusion.

We have demonstrated the superior performance of nonparametric confidence intervals based on a test of a dual or tilting parameter with a robustified or sandwich estimator of its variance. The statistic not only provides superior coverage accuracy, but also has remarkably straight QQplots, suggesting that a mean correction, if calculable, would be extremely effective.

Conventional higher–order asymptotic theory does not suggest that our statistic should be particularly effective, and indeed suggests that another statistic, the empirical likelihood ratio test, should provide better coverage after Bartlett correction. However, the ELR test does not display the straight QQplots that one would expect on the basis of its apparent Bartlett correctability. The only explanation that is credible for the apparent shortcomings of the ELR test and strengths of S_2 in practice is that the approximations of conventional asymptotic theory are not effective here because the terms in the approximating series are so large. This remains to be confirmed.

S_2 and analogous statistics have been discussed in papers by Corcoran, Davison, and Spady (1995) and Imbens, Johnson, and Spady (1995). The former considers the shortcomings of asymptotic theory identified here, carrying calculations out to the next higher order terms and confirming that these do not reoslve the apparent anomalies. The latter paper discusses the use of S_2 in overidentified systems of estimating equations, i.e. in situations where the number of estimating equations or moment conditions exceeds the number of unknow parameters. In this context, S_2 again proves to be surprisingly effective.

REFERENCES

[1] Corcoran, S., Personal communication, August, 1995.

[2] Corcoran, S., Davison, A., and Spady, R., "Reliable inference from empirical likelihoods," Department of Statistics, Oxford University, March, 1995.

[3] DiCiccio, T. and Romano, J., "Nonparametric confidence limits by resampling methods and least favorable families," *International Statistical Review*, 1990, **58**, 59-76.

[4] DiCiccio, T., Hall, P. and Romano, J., "Empirical likelihood is Bartlett-correctable," *Annals of Statistics*, 1991, **19**, 1053-1061.

[5] Hall, Peter, *The Bootstrap and Edgeworth Expansion,* Springer-Verlag, 1992.

[6] Imbens, G., Johnson, P., and Spady, R. "Information theoretic approaches to inference in moment condition models," Harvard Institute of Economic Research, Discussion Paper 1736, September, 1995.

[7] Koenker, R., Machado, J., Skeels, C. and Welsh, A. "Momentary lapses: moment expansions and the robustness of minimum distance estimation," Feb., 1993, University of Illinois & Australian National University working paper.

[8] Lee, S.M., and Young, G.A., "Asymptotic iterated bootstrap confidence intervals," *Annals of Statistics,* 1995, **23,** 1301-1330.

[9] Owen, A. "Empirical likelihood ratio confidence intervals for a single functional," *Biometrika*, 1988, **75**, 237-250.

[10] Stuart, A. and Ord, J., *Kendall's Advanced Theory of Statistics,* Volume 1, Fifth Edition, Oxford University Press, 1987.

[11] White, H. "Maximum likelihood estimation of misspecified models," *Econometrica,*

1982, **50**, 1-25.

Table 1

Normal data	80%	90%	95%	99%
$n = 10$				
Empirical Likelihood	0.7434 (*0.7470*)	0.8446 (*0.8446*)	0.9014 (*0.8964*)	0.9570 (*0.9542*)
Theoretical Bartlett	0.7788 (*0.7796*)	0.8740 (*0.8706*)	0.9208 (*0.9182*)	0.9682 (*0.9650*)
Estimated Bartlett	0.7694 (*0.7938*)	0.8672 (*0.8802*)	0.9160 (*0.9246*)	0.9650 (*0.9696*)
Exponential Tilt (Wald)	0.7858	0.8914	0.9474	0.9932
$n = 20$				
Empirical Likelihood	0.7752 (*0.7834*)	0.8688 (*0.8830*)	0.9304 (*0.9342*)	0.9800 (*0.9804*)
Theoretical Bartlett	0.7900 (*0.8006*)	0.8820 (*0.8962*)	0.9430 (*0.9418*)	0.9830 (*0.9844*)
Estimated Bartlett	0.7886 (*0.8034*)	0.8784 (*0.8980*)	0.9402 (*0.9424*)	0.9824 (*0.9848*)
Exponential Tilt (Wald)	0.7956	0.8950	0.9582	0.9924

χ^2_1 data	80%	90%	95%	99%
$n = 20$				
Empirical Likelihood	0.7390 (*0.7314*)	0.8354 (*0.8288*)	0.8866 (*0.8878*)	0.9492 (*0.9536*)
Theoretical Bartlett	0.7938 (*0.7872*)	0.8790 (*0.8772*)	0.9246 (*0.9262*)	0.9696 (*0.9706*)
Estimated Bartlett	0.7566 (*0.7634*)	0.8482 (*0.8546*)	0.9018 (*0.9034*)	0.9560 (*0.9616*)
Exponential Tilt (Wald)	0.7570	0.8646	0.9216	0.9786
$n = 40$				
Empirical Likelihood	0.7648 (*0.7644*)	0.8664 (*0.8680*)	0.9198 (*0.9236*)	0.9740 (*0.9740*)
Theoretical Bartlett	0.7930 (*0.7910*)	0.8856 (*0.8896*)	0.9344 (*0.9418*)	0.9790 (*0.9800*)
Estimated Bartlett	0.7794 (*0.7804*)	0.8748 (*0.8788*)	0.9264 (*0.9334*)	0.9774 (*0.9774*)
Exponential Tilt (Wald)	0.7808	0.8830	0.9426	0.9886

t_5 data	80%	90%	95%	99%
$n = 15$				
Empirical Likelihood	0.7496 (*0.7516*)	0.8540 (*0.8502*)	0.9126 (*0.9094*)	0.9732 (*0.9692*)
Theoretical Bartlett	0.8254 (*0.8266*)	0.9140 (*0.9106*)	0.9570 (*0.9544*)	0.9874 (*0.9862*)
Estimated Bartlett	0.7706 (*0.7898*)	0.8736 (*0.8884*)	0.9284 (*0.9348*)	0.9772 (*0.9794*)
Exponential Tilt (Wald)	0.7734	0.8936	0.9532	0.9942
$n = 30$				
Empirical Likelihood	0.7770 (*0.7768*)	0.8770 (*0.8784*)	0.9284 (*0.9308*)	0.9784 (*0.9780*)
Theoretical Bartlett	0.8090 (*0.8114*)	0.9008 (*0.9042*)	0.9466 (*0.9496*)	0.9850 (*0.9866*)
Estimated Bartlett	0.7902 (*0.7954*)	0.8856 (*0.8928*)	0.9366 (*0.9422*)	0.9816 (*0.9832*)
Exponential Tilt (Wald)	0.7926	0.8950	0.9492	0.9898

Table 2. Coverage Levels for Variances,
S_2 vs. Iterated Bootstrap and Approximations

	IB	S_2	I_1	I_2
Normal				
$n = 20$	.848	.853	.833	.832
$n = 35$	.853	.876	.854	.853
Folded Normal (high skew, low kurtosis)				
$n = 20$	.815	.799	.803	.800
$n = 35$	.834	.853	.821	.819
Log Normal (high skew, high kurtosis)				
$n = 20$	.544	.546	.526	.526
$n = 35$	.630	.603	.602	.602

Table 3. Expectations of Q and related statistics.

	$\frac{\mu_1^2}{\mu_2}$	$\frac{\mu_1^3\mu_3}{\mu_2^3}$	$\frac{\mu_1^4\mu_3^2}{\mu_2^5}$	$\frac{\mu_1^4\mu_4}{\mu_2^4}$	S_1	S_2	S_3	$\tilde{\mathsf{S}}_2$
$n = 20$								
theoretical	1.80000	-0.95000	1.20000	2.25000	1.24167	0.55000	1.00000	0.55000
simulated	1.44766	0.46228	0.39931	0.78020	1.75984	1.23473	1.92837	1.76813
est. s.d.	0.03421	0.02556	0.02219	0.02828	0.06730	0.02649	0.07161	0.06833
$n = 40$								
theoretical	1.40000	-0.47500	0.60000	1.12500	1.12083	0.77500	1.00000	0.77500
simulated	1.32681	0.05749	0.17823	0.55104	1.28826	1.11546	1.18971	1.05059
est. s.d	0.03140	0.01393	0.00850	0.02127	0.02949	0.02194	0.03511	0.03279
$n = 80$								
theoretical	1.20000	-0.23750	0.30000	0.56250	1.06042	0.88750	1.00000	0.88750
simulated	1.16000	-0.05530	0.12847	0.34489	1.10846	1.02226	1.01676	0.93392
est. s.d.	0.02724	0.00989	0.00465	0.01551	0.02339	0.01992	0.02083	0.01932
$n = 120$								
theoretical	1.13333	-0.15833	0.20000	0.37500	1.04028	0.92500	1.00000	0.92500
simulated	1.10265	-0.07924	0.12038	0.27833	1.05327	0.98888	0.98584	0.92316
est. s.d.	0.02526	0.01065	0.00511	0.01393	0.02227	0.02002	0.02069	0.01947
$n = 160$								
theoretical	1.10000	-0.11875	0.15000	0.28125	1.03021	0.94375	1.00000	0.94375
simulated	1.11235	-0.07590	0.10711	0.22520	1.07044	1.02061	1.02546	0.97717
est. s.d.	0.02347	0.01005	0.00461	0.01115	0.02096	0.01946	0.01987	0.01888
$n = 200$								
theoretical	1.08000	-0.09500	0.12000	0.22500	1.02417	0.95500	1.00000	0.95500
simulated	1.08446	-0.07255	0.09036	0.19166	1.04208	0.99663	1.00097	0.95968
est. s.d.	0.01393	0.00563	0.00231	0.00593	0.01230	0.01137	0.01148	0.01086
$n = 400$								
theoretical	1.04000	-0.04750	0.06000	0.11250	1.01208	0.97750	1.00000	0.97750
simulated	1.04076	-0.04779	0.05167	0.10172	1.01292	0.98559	0.99460	0.97361
est. s.d.	0.01279	0.00424	0.00138	0.00312	0.01176	0.01120	0.01136	0.01095

TABLE 4.

Expectation of z^j and cumulants of z.

$$\begin{aligned} E(z) &= \frac{-g}{2\sqrt{n}} + O(n^{-3/2}) & \kappa_1 &= \frac{-g}{2\sqrt{n}} + O(n^{-3/2}) \\ E(z^2) &= 1 + \frac{2g^2}{n} + O(n^{-2}) & \kappa_2 &= 1 + \frac{7g^2}{4n} + O(n^{-2}) \\ E(z^3) &= \frac{-7g}{2\sqrt{n}} + O(n^{-3/2}) & \kappa_3 &= \frac{-2g}{\sqrt{n}} + O(n^{-3/2}) \\ E(z^4) &= 3 + n^{-1}(28g^2 - 2k) + O(n^{-2}) & \kappa_4 &= n^{-1}(12g^2 - 2k) + O(n^{-2}) \end{aligned}$$

Transformation to normality:

$$\tau(z) = \frac{g}{6\sqrt{n}} + \left(1 + \frac{13g^2}{72n} - \frac{k}{4n}\right) z + \frac{g}{3\sqrt{n}} z^2 + \left(\frac{-g^2}{18n} + \frac{k}{12n}\right) z^3$$

Edgeworth polynomials (cf. Hall, p.48):

$$\begin{aligned} p_1(z) &= \frac{g}{6} + \frac{gz^2}{3} \\ p_2(z) &= \left(\frac{-g^2}{6} - \frac{k}{4}\right) z + \left(\frac{-g^2}{9} + \frac{k}{12}\right) z^3 - \frac{g^2 z^5}{18} \end{aligned}$$

Expectation of T_1^j and cumulants of T_1.

$$\begin{aligned} E(T_1) &= \frac{-g}{6\sqrt{n}} + O(n^{-3/2}) & \kappa_1 &= \frac{-g}{6\sqrt{n}} + O(n^{-3/2}) \\ E(T_1^2) &= 1 - \frac{g^2}{3n} + \frac{k}{2n} + O(n^{-2}) & \kappa_2 &= 1 - \frac{13g^2}{36n} + \frac{k}{2n} + O(n^{-2}) \\ E(T_1^3) &= \frac{-g}{2\sqrt{n}} + O(n^{-3/2}) & \kappa_3 &= O(n^{-3/2}) \\ E(T_1^4) &= 3 + n^{-1}(-2g^2 + 3k) + O(n^{-2}) & \kappa_4 &= O(n^{-2}) \end{aligned}$$

Transformation to normality:

$$\tau(T_1) = \frac{g}{6\sqrt{n}} + \left(1 + \frac{13g^2}{72n} - \frac{k}{4n}\right) T_1$$

Edgeworth polynomials:

$$\begin{aligned} p_1(T_1) &= \frac{g}{6} \\ p_2(T_1) &= \left(\frac{g^2}{6} - \frac{k}{4}\right) T_1 \end{aligned}$$

Expectation of T_2^j and cumulants of T_2.

$$\begin{aligned} E(T_2) &= O(n^{-3/2}) & \kappa_1 &= O(n^{-3/2}) \\ E(T_2^2) &= 1 + \frac{3g^2}{4n} - \frac{k}{n} + O(n^{-2}) & \kappa_2 &= 1 + \frac{3g^2}{4n} - \frac{k}{n} + O(n^{-2}) \\ E(T_2^3) &= \frac{g}{\sqrt{n}} + O(n^{-3/2}) & \kappa_3 &= \frac{g}{\sqrt{n}} + O(n^{-3/2}) \\ E(T_2^4) &= 3 + n^{-1}(\tfrac{21}{2}g^2 - 12k) + O(n^{-2}) & \kappa_4 &= n^{-1}(6g^2 - 6k) + O(n^{-2}) \end{aligned}$$

Transformation to normality:

$$\tau(T_2) = \frac{g}{6\sqrt{n}} + \left(1 + \frac{13g^2}{72n} - \frac{k}{4n}\right) T_2 - \frac{g}{6\sqrt{n}} T_2^2 + \left(\frac{-5g^2}{36n} + \frac{k}{4n}\right) T_2^3$$

Edgeworth polynomials:

$$\begin{aligned} p_1(T_2) &= \frac{g}{6} - \frac{gT_2^2}{6} \\ p_2(T_2) &= \left(\frac{g^2}{6} - \frac{k}{4}\right) T_2 + \left(\frac{-g^2}{9} + \frac{k}{4}\right) T_2^3 - \frac{g^2 T_2^5}{72} \end{aligned}$$