ON BARTLETT CORRECTABILITY OF EMPIRICAL LIKELIHOOD IN GENERALIZED POWER DIVERGENCE FAMILY

By

Lorenzo Camponovo and Taisuke Otsu

October 2011

COWLES FOUNDATION DISCUSSION PAPER NO. 1825

COWLES FOUNDATION FOR RESEARCH IN ECONOMICS
YALE UNIVERSITY
Box 208281
New Haven, Connecticut 06520-8281

http://cowles.econ.yale.edu/

On Bartlett correctability of empirical likelihood in generalized power divergence family*

Lorenzo Camponovo†
Department of Economics
University of Lugano

Taisuke Otsu‡
Cowles Foundation & Department of Economics
Yale University

October 2011

Abstract

Baggerly (1998) showed that empirical likelihood is the only member in the Cressie-Read power divergence family to be Bartlett correctable. This paper strengthens Baggerly's result by showing that in a generalized class of the power divergence family, which includes the Cressie-Read family and other nonparametric likelihood such as Schennach's (2005, 2007) exponentially tilted empirical likelihood, empirical likelihood is still the only member to be Bartlett correctable.

1 Introduction

Since Owen (1988), empirical likelihood has been used as a device to construct nonparametric likelihood for numerous statistical problems and models as surveyed by Owen (2001). In spite of its nonparametric construction based on observed data points, empirical likelihood shares similar properties to parametric likelihood. For example, the empirical likelihood ratio statistic obeys the chi-squared limiting distribution, so-called Wilks' phenomenon. Another distinguishing feature of empirical likelihood is that it admits Bartlett correction, a second-order refinement based on a mean adjustment. This point was first made by DiCiccio, Hall and Romano (1991) and extended to other contexts, such as quantiles (Chen and Hall, 1993), time series models (Kitamura, 1997; Monti, 1997), local linear smoothers (Chen and Qin, 2001), among others. Also Bartlett correctability has been studied for other constructions of nonparametric likelihood. Jing and Wood (1996) showed that exponential tilting (or empirical entropy) likelihood is not Bartlett correctable. Corcoran (1998) constructed some Bartlett correctable nonparametric likelihood based on a Taylor expansion of empirical likelihood. Baggerly (1998) strengthened

*The authors would like to thank Elvezio Ronchetti for helpful comments.
†E-mail: lorenzo.camponovo@usi.ch. Address: Via Giuseppe Buffi 13, 6900 Lugano, Ticino, Switzerland.
‡E-mail: taisuke.otsu@yale.edu. Website: http://cowles.econ.yale.edu/faculty/otsu.htm. Address: P.O. Box 208281, New Haven, CT 06520-8281, USA. Phone: +1-203-432-9771. Fax: +1-203-432-6167.

Jing and Wood's (1996) result by showing that empirical likelihood is the only member in the Cressie and Read (1984) power divergence family to be Bartlett correctable.

The Cressie-Read type nonparametric likelihood is computed by choosing a tuning constant to define both the shape of the criterion function and the form of weights allocated to data points. Schennach (2005, 2007) suggested to choose different tuning constants for the shape of the criterion and the form of weights, and proposed a more general class of nonparametric likelihood. In particular, Schennach (2005) showed that exponentially tilted empirical likelihood (where the criterion is log-likelihood but the weights are computed by exponential tilting) can emerge as a valid likelihood function for Bayesian inference by a limiting argument. Also Schennach (2007) argued that when generalized estimating equations are misspecified, the point estimator based exponentially tilted empirical likelihood shows some robustness compared to the one based on empirical likelihood. Given this background, it is of interest to extend Baggerly's (1998) analysis to accommodate such new likelihood constructions and to study their Bartlett correctability.

In this paper, we confirm that in a generalized class of the power divergence family containing two tuning constants, empirical likelihood is still the only member to be Bartlett correctable. This result not only includes Baggerly's (1998) result as a special case, but also implies Schennach's (2005, 2007) exponentially tilted empirical likelihood is not Bartlett correctable. Technically we follow a conventional approach based on the Edgeworth expansion (DiCiccio, Hall and Romano, 1991). We focus on characterizing the third and fourth order joint cumulants of the signed root of the test statistic based on the generalized power divergence family, and shows that those cumulants vanish at sufficiently fast rates only when we employ the empirical likelihood statistic.

2 Generalized power divergence family

We begin by introducing the generalized power divergence statistic. Consider a scalar random variable X from an unknown distribution F_0 with mean μ_0. Following Owen (1988), the log-empirical likelihood ratio statistic for the mean is written as

$$\ell_{EL}(\mu_0) = -2 \max_{p_1,\ldots,p_n} \sum_{i=1}^{n} \log(np_i), \quad \text{subject to } \sum_{i=1}^{n} p_i = 1, \ \sum_{i=1}^{n} p_i X_i = \mu_0.$$

It is known that under suitable regularity conditions the statistic $\ell_{EL}(\mu_0)$ converges in distribution to the χ_1^2 distribution (Owen, 1988) and admits Bartlett correction, which yields a confidence interval with coverage error of size n^{-2} (DiCiccio, Hall and Romano, 1991).

Baggerly (1998) adapted the Cressie and Read (1984) power divergence family for goodness-of-fit to the present context and considered the test statistic in the form of

$$\ell_\gamma(\mu_0) = \min_{p_1,\ldots,p_n} \frac{2}{\gamma(\gamma+1)} \sum_{i=1}^{n} \left\{ (np_i)^{\gamma+1} - 1 \right\}, \quad \text{subject to } \sum_{i=1}^{n} p_i = 1, \ \sum_{i=1}^{n} p_i X_i = \mu_0, \tag{1}$$

where $\gamma \in \mathbb{R}$ is a user-specified tuning constant. For the cases of $\gamma = -1$ and $\gamma = 0$, we use the continuous limits $\ell_\gamma(\mu_0) = \min_{p_1,\ldots,p_n} -2\sum_{i=1}^n \log(np_i)$ as $\gamma \to -1$ and $\ell_\gamma(\mu_0) = \min_{p_1,\ldots,p_n} 2n\sum_{i=1}^n p_i \log(np_i)$ as $\gamma \to 0$, respectively. The empirical likelihood ratio statistic $\ell_{EL}(\mu_0)$ corresponds to the case of $\gamma = -1$. The case of $\gamma = 0$ is often called the exponential tilting or empirical entropy statistic. Other popular choices for γ include the Neyman's modified χ^2 ($\gamma = 1$), Hellinger or Freeman-Tukey ($\gamma = -\frac{1}{2}$), and Pearson's χ^2 ($\gamma = -2$). Baggerly (1998) showed that the power divergence statistic $\ell_\gamma(\mu_0)$ converges in distribution to the χ_1^2 distribution for any γ, and that $\ell_\gamma(\mu_0)$ is Bartlett correctable only for the case of $\gamma = -1$, the empirical likelihood ratio statistic. As Baggerly (1998) argued, a key insight of (lack of) Bartlett correctability is that the third and fourth order cumulants of the signed root of $\ell_\gamma(\mu_0)$ do not vanish at sufficiently fast rates when $\gamma \neq -1$.

From different perspectives, Schennach (2005, 2007) introduced the exponentially tilted empirical likelihood statistic

$$\ell_{ETEL}(\mu_0) = -2\sum_{i=1}^n \log(np_{ET,i}),$$

i.e., the criterion function is defined by $\ell_\gamma(\mu_0)$ with $\gamma = -1$, where $p_{ET,1}, \ldots p_{ET,n}$ solve the minimization problem of $\ell_\gamma(\mu_0)$ with $\gamma = 0$,

$$\min_{p_1,\ldots,p_n} \sum_{i=1}^n p_i \log(np_i), \quad \text{subject to } \sum_{i=1}^n p_i = 1, \ \sum_{i=1}^n p_i X_i = \mu_0.$$

Schennach (2007) studied asymptotic properties of a point estimator based on this statistic for generalized estimating equations. Also Schennach (2005) argued that the function $\ell_{ETEL}(\mu)$ can be interpreted as a valid likelihood function for Bayesian inference. It should be noted that the statistic $\ell_{ETEL}(\mu_0)$ does not belong to the power divergence family (1). Therefore, Bartlett correctability of the statistic $\ell_{ETEL}(\mu_0)$ is an open question.

In order to address this issue, we generalize the power divergence statistic $\ell_\gamma(\mu_0)$ as follows: for $\gamma, \phi \in \mathbb{R}$,

$$\ell_{\gamma,\phi}(\mu_0) = \frac{2}{\gamma(\gamma+1)}\sum_{i=1}^n \left\{(np_{\phi,i})^{\gamma+1} - 1\right\}, \tag{2}$$

where $p_{\phi,1}, \ldots p_{\phi,n}$ solve

$$\min_{p_1,\ldots,p_n} \frac{2}{\phi(\phi+1)}\sum_{i=1}^n \left\{(np_i)^{\phi+1} - 1\right\}, \quad \text{subject to } \sum_{i=1}^n p_i = 1, \ \sum_{i=1}^n p_i X_i = \mu_0. \tag{3}$$

Note that the shape of the criterion function in (2) is given by $\ell_\gamma(\mu_0)$ but the probability weights $\{p_{\phi,1}, \ldots, p_{\phi,n}\}$ are computed by $\ell_\phi(\mu_0)$. If $\gamma = \phi$, this statistic reduces to the power divergence statistic $\ell_\gamma(\mu_0)$. Also this statistic covers the exponentially tilted empirical likelihood statistic $\ell_{ETEL}(\mu_0)$ when $\gamma = -1$ and $\phi = 0$ as a continuous limit.

By adapting the argument in Baggerly (1998) and Schennach (2005), we can see that the statistic $\ell_{\gamma,\phi}(\mu_0)$ converges in distribution to the χ_1^2 distribution under the same conditions in Baggerly (1998, Theorem 1). The goal of this paper is to study Bartlett correctability of the generalized statistic $\ell_{\gamma,\phi}(\mu_0)$.

3 Bartlett correctability

To investigate Bartlett correctability of the generalized power divergence statistic $\ell_{\gamma,\phi}(\mu_0)$, we follow the conventional recipe of DiCiccio, Hall and Romano (1991) and Baggerly (1998), among others. In particular, we first derive the signed root of the statistic $\ell_{\gamma,\phi}(\mu_0)$ based on a dual problem of the minimization in (3), and then evaluate the third and fourth order cumulants of the signed root. Based on these cumulant expressions, we verify for what values of γ and ϕ these cumulants vanish at sufficiently fast rates to admit Bartlett correction.

To simplify the presentation, we focus on the case where X_i is scalar and standardized as $\mu_0 = E[X_i] = 0$ and $Var(X_i) = 1$. Although the presentation and technical argument become more complicated, a similar result holds for the case where X_i is a vector and the parameter of interest is a smooth function of the mean of X_i. Hereafter we present only the main result. Technical details for the derivations are available from the authors upon request.

We use the following definitions

$$\alpha_j = E\left[X_i^j\right], \quad A_j = \frac{1}{n}\sum_{i=1}^n X_i^j - \alpha_j,$$

for $j = 1, 2, \ldots$. Note that $\alpha_1 = E[X_i] = 0$, $\alpha_2 = Var(X_i) = 1$, and $A_j = O_p\left(n^{-1/2}\right)$ for any j as far as sufficiently higher order moments exist.

First, we find the signed root of $\ell_{\gamma,\phi}(\mu_0)$ with $\mu_0 = 0$. By applying the Lagrange multiplier method, the solution of (3) is

$$p_{\phi,i} = \frac{1}{n}(1 + s + tX_i)^{\frac{1}{\phi}},$$

where s and t solve

$$\frac{1}{n}\sum_{i=1}^n (1 + s + tX_i)^{\frac{1}{\phi}} = 1, \quad \frac{1}{n}\sum_{i=1}^n (1 + s + tX_i)^{\frac{1}{\phi}} X_i = 0. \tag{4}$$

From Baggerly (1998), we can see that $s = O_p\left(n^{-1}\right)$ and $t = O_p\left(n^{-1/2}\right)$. Expansions of (4) and repeated substitutions yield expansions of s and t as follows

$$\begin{aligned} s &= \frac{1}{2}\phi(1+\phi)A_1^2 + \frac{1}{6}\phi(1+\phi)(1-\phi)\alpha_3 A_1^3 - \frac{1}{2}\phi(1+\phi)A_1^2 A_2 \\ &\quad + \frac{1}{2}\phi(1+\phi)A_1^2 A_2^2 - \frac{1}{2}\phi(1-\phi)(1+\phi)\alpha_3 A_1^3 A_2 + \frac{1}{6}\phi(1-\phi)(1+\phi)A_1^3 A_3 \\ &\quad + \frac{1}{8}\phi\left\{(1+\phi)^3 + (1-\phi)^2(1+\phi)\alpha_3^2 - \frac{1}{3}(1-\phi)(1+\phi)(1-2\phi)\alpha_4\right\}A_1^4 + O_p\left(n^{-5/2}\right). \end{aligned}$$

and

$$\begin{aligned} t &= -\phi A_1 - \frac{1}{2}\phi(1-\phi)\alpha_3 A_1^2 + \phi A_1 A_2 \\ &\quad - \phi A_1 A_2^2 + \frac{3}{2}\phi(1-\phi)\alpha_3 A_1^2 A_2 - \frac{1}{2}\phi(1-\phi)A_1^2 A_3 \\ &\quad - \frac{1}{2}\phi\left\{\phi(1+\phi) + (1-\phi)^2\alpha_3^2 - \frac{1}{3}(1-\phi)(1-2\phi)\alpha_4\right\}A_1^3 + O_p\left(n^{-2}\right). \end{aligned}$$

By inserting these expressions to an expansion of $\ell_{\gamma,\phi}(\mu_0)$,

$$\begin{aligned} & n^{-1}\ell_{\gamma,\phi}(\mu_0) \\ = \; & A_1^2 + \frac{1}{3}(1-\gamma)\alpha_3 A_1^3 - A_1^2 A_2 + A_1^2 A_2^2 - (1-\gamma)\alpha_3 A_1^3 A_2 + \frac{1}{3}(1-\gamma) A_1^3 A_3 \\ & + \left\{ \left(\frac{1}{4} + \frac{\gamma}{2} + \frac{\gamma\phi}{2} - \frac{\phi^2}{4} \right) + \left(\frac{1}{4} - \frac{\gamma}{2} + \frac{\gamma\phi}{2} - \frac{\phi^2}{4} \right)\alpha_3^2 + \left(-\frac{1}{12} + \frac{\gamma}{4} + \frac{\gamma^2}{12} - \frac{\gamma\phi}{2} + \frac{\phi^2}{4} \right)\alpha_4 \right\} A_1^4 \\ & + O_p\left(n^{-5/2}\right). \end{aligned}$$

Therefore, the signed root of the statistic $n^{-1}\ell_{\gamma,\phi}(\mu_0)$ is obtained as

$$n^{-1}\ell_{\gamma,\phi}(\mu_0) = \left(R_{\gamma,\phi}^{(1)} + R_{\gamma,\phi}^{(2)} + R_{\gamma,\phi}^{(3)} \right)^2 + O_p\left(n^{-5/2}\right),$$

where

$$\begin{aligned} R_{\gamma,\phi}^{(1)} &= A_1, \qquad R_{\gamma,\phi}^{(2)} = -\frac{1}{2}A_1 A_2 + \frac{1}{6}(1-\gamma)\alpha_3 A_1^2, \\ R_{\gamma,\phi}^{(3)} &= \frac{3}{8}A_1 A_2^2 - \frac{5}{12}(1-\gamma)\alpha_3 A_1^2 A_2 + \frac{1}{6}(1-\gamma) A_1^2 A_3 \\ & \quad + \left\{ \left(\frac{1}{8} + \frac{\gamma}{4} + \frac{\gamma\phi}{4} - \frac{\phi^2}{8} \right) + \left(\frac{1}{9} - \frac{2\gamma}{9} - \frac{\gamma^2}{72} + \frac{\gamma\phi}{4} - \frac{\phi^2}{8} \right)\alpha_3^2 + \left(-\frac{1}{24} + \frac{\gamma}{8} + \frac{\gamma^2}{24} - \frac{\gamma\phi}{4} + \frac{\phi^2}{8} \right)\alpha_4 \right\} A_1^3. \end{aligned}$$

Note that $R_{\gamma,\phi}^{(j)} = O_p\left(n^{-j/2}\right)$ for $j = 1, 2, 3$. We can confirm that for the empirical likelihood case (i.e., $\phi = -1$ and $\gamma = -1$), this signed root expression coincides with the one in DiCiccio, Hall and Romano (1991).

Next, to investigate Bartlett correctability of the generalized power divergence statistic, we evaluate the third and fourth order cumulants (denoted by $\kappa_3(\gamma,\phi)$ and $\kappa_4(\gamma,\phi)$, respectively) of the signed root $R_{\gamma,\phi} = R_{\gamma,\phi}^{(1)} + R_{\gamma,\phi}^{(2)} + R_{\gamma,\phi}^{(3)}$. In particular, if the cumulants satisfy

$$\kappa_3(\gamma,\phi) = O\left(n^{-3}\right), \quad \kappa_4(\gamma,\phi) = O\left(n^{-4}\right), \tag{5}$$

then the conventional argument based on the Edgeworth expansion guarantees Bartlett correctability of the statistic $n^{-1}\ell_{\gamma,\phi}(\mu_0)$. After lengthy calculations, the third order cumulant is obtained as

$$\begin{aligned} \kappa_3(\gamma,\phi) &= E\left[R_{\gamma,\phi}^{(1)}\right] + 3E\left[\left(R_{\gamma,\phi}^{(1)}\right)^2 R_{\gamma,\phi}^{(2)}\right] - 3E\left[\left(R_{\gamma,\phi}^{(1)}\right)^2\right] E\left[R_{\gamma,\phi}^{(2)}\right] + O\left(n^{-3}\right) \\ &= -\frac{1}{2}(1+\gamma)\alpha_3 \left\{ E\left[A_1^4\right] - \left(E\left[A_1^2\right]\right)^2 \right\} + O\left(n^{-3}\right) \\ &= . \end{aligned}$$

It is interesting to note that the third order cumulant $\kappa_3(\gamma,\phi)$ does not depend on the tuning constant ϕ. From this expression, we can see that the first requirement in (5) is satisfied only when $\gamma = -1$. Therefore, to evaluate the fourth order cumulant, we focus on the case of $\gamma = -1$. To this end, it is useful to note that for $\gamma = -1$, it holds

$$R_{-1,\phi}^{(1)} = R_{-1,-1}^{(1)}, \qquad R_{-1,\phi}^{(2)} = R_{-1,-1}^{(2)}, \qquad R_{-1,\phi}^{(3)} = R_{-1,-1}^{(3)} - \frac{1}{8}(\phi+1)^2\left(1+\alpha_3^2-\alpha_4\right)A_1^3.$$

By utilizing these relationships, the fourth order cumulant $\kappa_4(\gamma,\phi)$ with $\gamma=-1$ is obtained as

$$\begin{aligned}\kappa_4(-1,\phi) &= 4E\left[\left(R^{(1)}_{-1,-1}\right)^3 R^{(3)}_{-1,\phi}\right] - 12E\left[\left(R^{(1)}_{-1,-1}\right)^2\right] E\left[R^{(1)}_{-1,-1}R^{(3)}_{-1,\phi}\right] \\ &\quad +(\text{terms involving moments of } R^{(1)}_{-1,-1}\text{and } R^{(2)}_{-1,-1}\text{only}) + O\left(n^{-4}\right) \\ &= -\frac{1}{8}(1+\phi)^2\left(1+\alpha_3^2-\alpha_4\right)\left(4E\left[A_1^6\right]-12E\left[A_1^2\right]E\left[A_1^4\right]\right)+O\left(n^{-4}\right) \\ &= .\end{aligned}$$

From this expression, we can see that the second requirement in (5) is satisfied only when $\phi=-1$. Therefore, in the class of the generalized power divergence statistic $\ell_{\gamma,\phi}(\mu_0)$ considered in this paper, only empirical likelihood (i.e., the case of $\phi=-1$ and $\gamma=-1$) is Bartlett correctable. This result also implies that the exponentially tilted empirical likelihood statistic (i.e., the case of $\gamma=-1$ and $\phi=0$) considered by Schennach (2005, 2007) is not Bartlett correctable.

4 Discussions

It is interesting to investigate other properties of the generalized statistic $\ell_{\gamma,\phi}(\mu_0)$ for different values of γ and ϕ. For example, Baggerly (1998) argued that the weight $p_{\phi,i}$ given by (3) can be negative when $\phi>0$, and Schennach (2007, p. 641) conjectured that lack of $\sqrt{n}$-consistency of the point estimator under misspecified generalized estimating equations can occur for any $\phi\le 0$. Also, if some weight $p_{\phi,i}$ takes zero, the statistic $\ell_{\gamma,\phi}(\mu_0)$ diverges when $\gamma\ge -1$ but is finite when $\gamma\le 0$. From a practical point of view, it should be noted that the minimization in (3) has an explicit solution when $\phi=1$.

This paper shows that in the class of generalized power divergence family, only empirical likelihood ($\phi=\gamma=-1$) admits Bartlett correction, which yields a confidence interval with coverage error of size $O_p\left(n^{-2}\right)$. This result also confirms the finding of Jing and Wood (1996), exponential tilting likelihood ($\phi=\gamma=0$) is not Bartlett correctable. However, recent research has shown an attractive multiplicative feature of the coverage error of the exponential tilting-based confidence interval. In particular, Ma and Ronchetti (2011) show that in measurement error models, the coverage error of the exponential tilting-based confidence interval takes the form of $\left\{1-F_{\chi^2}(\cdot)\right\}O_p\left(n^{-1}\right)$, where $F_{\chi^2}(\cdot)$ is a distribution function of some χ^2 distribution. As pointed out in Ma and Ronchetti (2011), since the term $\left\{1-F_{\chi^2}(\cdot)\right\}$ is often very small in hypothesis testing, the relative error $\left\{1-F_{\chi^2}(\cdot)\right\}O_p\left(n^{-1}\right)$ may be potentially more meaningful than the absolute error. It is interesting to extend this study to the generalized power divergence family considered in this paper, and determine the values of γ and ϕ to admit the multiplicative form of the coverage error. This extension is currently under investigation by the authors.

References

[1] Baggerly, K. A. (1998) Empirical likelihood as a goodness-of-fit measure, *Biometrika*, 85, 535-547.

[2] Chen, S. X. and P. Hall (1993) Smoothed empirical likelihood confidence intervals for quantiles, *Annals of Statistics*, 21, 1166-1181.

[3] Chen, S. X. and Y. S. Qin (2001) Empirical likelihood confidence intervals for local linear smoothers, *Biometrika*, 87, 946-953.

[4] Corcoran, S. A. (1998) Bartlett adjustment of empirical discrepancy statistics, *Biometrika*, 85, 967-972.

[5] Cressie, N. and T. R. C. Read (1984) Multinomial goodness-of-fit tests, *Journal of the Royal Statistical Society*, B, 46, 440-464.

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

[7] Jing, B. Y. and A. T. A. Wood (1996) Exponential empirical likelihood is not Bartlett correctable, *Annals of Statistics*, 24, 365-369.

[8] Kitamura, Y. (1997) Empirical likelihood methods with weakly dependent process, *Annals of Statistics*, 25, 2084-2102.

[9] Ma, Y. and E. Ronchetti (2011) Saddlepoint test in measurement error models, *Journal of the American Statistical Association*, 106, 147-156.

[10] Monti, A. C. (1997) Empirical likelihood confidence regions in time series models, *Biometrika*, 84, 395-405.

[11] Owen, A. B. (1988) Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, 75, 237-249.

[12] Owen, A. B. (2001) *Empirical Likelihood*, Chapman Hall/CRC.

[13] Schennach, S. M. (2005) Bayesian exponentially tilted empirical likelihood, *Biometrika*, 92, 31-46.

[14] Schennach, S. M. (2007) Point estimation with exponentially tilted empirical likelihood, *Annals of Statistics*, 35, 634-672.