Working Paper Series, N. 8, August 2014

Second-order accurate confidence regions based on members of the generalised power divergence family

Nicola Lunardon
Department of Statistical Sciences
University of Padova
Italy

Gianfranco Adimari
Department of Statistical Sciences
University of Padova
Italy

Abstract: Recently, a technique based on pseudo-observations has been proposed to tackle the so called convex hull problem for the empirical likelihood statistic. The resulting adjusted empirical likelihood also achieves the high-order precision of the Bartlett correction. Nevertheless, the technique induces an upper bound on the resulting statistic that may lead, in certain circumstances, to worthless confidence regions equal to the whole parameter space. In this paper we show that suitable pseudo-observations can be deployed to make each element of the generalised power divergence family Bartlett-correctable and released from the convex hull problem. Our approach is conceived to achieve this goal by means of two distinct sets of pseudo-observations with different tasks. An important effect of our formulation is to provide a solution that permits to overcome the problem of the upper bound. The proposal, whose effectiveness is confirmed by simulation results, gives back attractiveness to a broad class of statistics that potentially contains good alternatives to the empirical likelihood.

Keywords: Bartlett correction, High-order asymptotics, Maximum entropy, Empirical likelihood, Exponential empirical likelihood, Power divergence.

Contents

Department of Statistical Sciences
Via Cesare Battisti, 241
35121 Padova
Italy

tel: +39 049 8274168
fax: +39 049 8274170
http://www.stat.unipd.it

Corresponding author:
Nicola Lunardon
tel: +39 049 827 4124
lunardon@stat.unipd.it
http://homes.stat.unipd.it/nicolalunardon/

Second-order accurate confidence regions based on members of the generalised power divergence family

Nicola Lunardon
Department of Statistical Sciences
University of Padova
Italy

Gianfranco Adimari
Department of Statistical Sciences
University of Padova
Italy

Abstract: Recently, a technique based on pseudo-observations has been proposed to tackle the so called convex hull problem for the empirical likelihood statistic. The resulting adjusted empirical likelihood also achieves the high-order precision of the Bartlett correction. Nevertheless, the technique induces an upper bound on the resulting statistic that may lead, in certain circumstances, to worthless confidence regions equal to the whole parameter space. In this paper we show that suitable pseudo-observations can be deployed to make each element of the generalised power divergence family Bartlett-correctable and released from the convex hull problem. Our approach is conceived to achieve this goal by means of two distinct sets of pseudo-observations with different tasks. An important effect of our formulation is to provide a solution that permits to overcome the problem of the upper bound. The proposal, whose effectiveness is confirmed by simulation results, gives back attractiveness to a broad class of statistics that potentially contains good alternatives to the empirical likelihood.

Keywords: Bartlett correction, High-order asymptotics, Maximum entropy, Empirical likelihood, Exponential empirical likelihood, Power divergence.

1 Introduction

Let $x = x_1, \ldots, x_n$ be a random sample of size n from a random vector X having distribution function F_0 supported on $\mathcal{X} \subseteq \mathbb{R}^q$ and let $\theta = \theta(F) \in \Theta \subseteq \mathbb{R}^d$ be a functional of interest defined on arbitrary distributions F, $q, d \geq 1$. Define $\theta(\cdot)$ to be the solution of an unbiased estimating equation, i.e. θ satisfies

$$\mathbb{E}_F[h(X;\theta)] = 0, \tag{1}$$

where $h : \mathcal{X} \times \Theta \to \mathcal{C}$ is a known function and $\mathbb{E}_F[\cdot]$ indicates that expectation is taken with respect to F, $\mathcal{C} \subseteq \mathbb{R}^d$. Following Camponovo and Otsu (2014), the generalised power divergence family for θ is defined as the class $\{\Delta_{\gamma,\phi}(\theta)\}$ of functions

$$\Delta_{\gamma,\phi}(\theta) = \frac{2}{\gamma(\gamma+1)} \sum_{i=1}^{n} \left\{ [w_i(\theta)\, n]^{\gamma+1} - 1 \right\}, \tag{2}$$

where

$$w_i(\theta) = \begin{cases} n^{-1}\left\{1+\lambda(\theta)+\beta(\theta)^\top h(x_i;\theta)\right\}^{1/\phi} & \text{if } \phi \neq 0 \\ \lambda(\theta)\exp\left\{\beta(\theta)^\top h(x_i;\theta)\right\} & \text{if } \phi = 0 \end{cases} \tag{3}$$

and $\lambda(\theta) \in \mathbb{R}$, $\beta(\theta) \in \mathbb{R}^d$ are determined by the equations

$$\sum_{i=1}^n w_i(\theta) = 1,\ \sum_{i=1}^n w_i(\theta) h(x_i;\theta) = 0,$$

as long as θ ensures that the null vector belongs to the convex hull of $\{h(x_1;\theta),\dots,h(x_n;\theta)\}$. The real parameters γ, ϕ in (2) and (3) are the tuning parameters of the family. The special case $\gamma \in \{-1, 0\}$ is handled by taking the appropriate limit.

The generalised power divergence family extends the Read and Cressie (1988) power divergence family, which can be recovered by imposing the restriction $\gamma = \phi$, and includes several well known empirical discrepancy statistics as the Owen (1988, 1990) empirical likelihood statistic ($\gamma = \phi = -1$), the maximum entropy statistic ($\gamma = \phi = 0$), the Euclidean likelihood statistic ($\gamma = \phi = 1$), the Freeman-Tukey statistic ($\gamma = \phi = -1/2$), and the exponential empirical likelihood statistic ($\gamma = -1, \phi = 0$) (Jing and Wood, 1996; Corcoran, 1998, among others).

The family (2) can be deployed to draw inferences about θ by means of either approximate pivots or tests for which the usual chi-squared asymptotics apply. A straightforward extension of the limiting result provided by Baggerly (1998) for the Read-Cressie family shows that, for each pair (γ, ϕ), $\Delta_{\gamma,\phi}(\theta_0)$ is asymptotically chi-squared with d degrees of freedom, being $\theta_0 = \theta(F_0)$ the true parameter value. Typically, confidence sets obtained through the chi-squared calibration have coverage error of size $O(n^{-1})$. DiCiccio *et al.* (1991) show that this rate can be enhanced to $O(n^{-2})$ for empirical likelihood confidence sets through the Bartlett correction. The check of this result within both the Read-Cressie power divergence family and its generalisation reveals that the empirical likelihood is one of a kind since it is the only member admitting the Bartlett correction (Baggerly, 1998; Camponovo and Otsu, 2014). Despite the practical effectiveness of the Bartlett correction has been questioned (Corcoran *et al.*, 1995), the promise to achieve improved inference from the empirical likelihood has contributed to place the latter in a leading position with respect to the members in family (2).

A conceptual pitfall in the definition of the family is the so called convex hull problem: any member $\Delta_{\gamma,\phi}(\theta)$ is well defined in θ only if θ ensures that the null vector belongs to the convex hull of $\{h(x_1;\theta),\dots,h(x_n;\theta)\}$; conversely, the common practice is to set $\Delta_{\gamma,\phi}(\theta) = \infty$ (for the case of the empirical likelihood $\Delta_{-1,-1}(\theta)$, see e.g. Owen, 2001, Sect. 10.4). This convention, however, comes along with two drawbacks because the check of the convex hull condition is not straightforward in general, and because the relative plausibility of parameters not satisfying such condition can be stated at all. Recent advances provided by Chen *et al.* (2008) and Liu and Chen (2010) have demonstrated that for the empirical likelihood it is possible to link the resolution of the convex hull problem to high-order asymptotics. These authors demonstrate that an effective answer to the convex hull challenge is to compute an adjusted version of the empirical likelihood obtained by adding suitable pseudo-observations to $h(x_1;\theta),\dots,h(x_n;\theta)$. Surprisingly, if the pseudo-observations are well-tuned, then the resulting adjusted empirical likelihood is also automatically Bartlett-corrected. Li *et al.* (2011) show that the method applies also to the exponential empirical likelihood.

As soon as this solution was conceived, Emerson and Owen (2009) highlighted that, when inference focuses on the population mean, the adjusted empirical likelihood is always bounded above and that the magnitude of the bound depends on the added pseudo-observations. In some cases the bound can be tight enough to force the derived confidence regions to degenerate up to the trivial 100% confidence set.

In this paper we extend the approach by Liu and Chen (2010) and Li *et al.* (2011) to the entire generalised power divergence family. Specifically, we show that suitable pseudo-observations added to $h(x_1;\theta),\ldots,h(x_n;\theta)$ make each element of the family Bartlett-correctable and released from the convex hull constraint. Unlike the existing proposals, our approach is conceived to achieve this goal by means of two distinct sets of pseudo-observations with different roles and tasks. An important effect of our formulation is to provide a solution that, in practice, avoids the problem of the upper bound highlighted by Emerson and Owen (2009).

The rest of the paper is organised as follows. In Section 2, we set up the notation and provide an essential background about the Bartlett correction for the empirical likelihood and the methodology proposed by Liu and Chen (2010) that leads to the adjusted Bartlett-corrected empirical likelihood. In Section 3, we state our main results and give the proposed adjustment working for every element of the generalised power divergence family. In Section 4, we provide empirical evidence, through Monte Carlo simulations, about the effectiveness of our approach. In particular, we apply our technique to some members of the family and make a comparison, in terms of coverage accuracy of confidence regions, with the original Bartlett-corrected empirical likelihood and the adjusted Bartlett-corrected empirical likelihood considered in Liu and Chen (2010). Finally, a brief discussion is given in Section 5. Proofs are deferred to the Appendix.

2 Preliminaries

2.1 Notation setup and definitions

In the sequel we make use of index notation, whereby any set of positive integers $\{l_1,\ldots,l_m\}$, $m\in\{1,\ldots,d\}$, is intended to index components of an m-dimensional array rather than to denote powers, e.g. $h^{l_1\cdots l_m}(X;\theta)=h^{l_1}(X;\theta)\cdots h^{l_m}(X;\theta)$. The usual power notation is best avoided and continue to do so until further notice. The reader may refer to McCullagh (1987) for more details on index notation. Assume that the true parameter value θ_0 is the unique solution of $\mathbb{E}_{F_0}[h(X;\theta)]=0$ and let

$$\alpha^{l_1\cdots l_m}=\mathbb{E}_{F_0}\left[h^{l_1}(X;\theta_0)\cdots h^{l_m}(X;\theta_0)\right],$$

$$A^{l_1\cdots l_m}(\theta)=n^{-1}\sum_{i=1}^{n}h^{l_1}(x_i;\theta)\cdots h^{l_m}(x_i;\theta)-\alpha^{l_1\cdots l_m}.$$

Further, denote with

$$\kappa^{l_1,\cdots,l_m}(W)=\mathrm{cum}(W^{l_1},\cdots,W^{l_m}),$$

the m-th generalised joint cumulant of a d-dimensional random vector W.

When population moments need to be estimated, we consider $\hat\theta=\theta_0+O_p(n^{-1/2})$, where $\hat\theta$ solves the empirical version of (1), i.e. $n^{-1}\sum_{i=1}^{n}h(x_i;\hat\theta)=0$, and define the following root-n consistent estimates

$$\hat\alpha^{l_1\cdots l_m}=n^{-1}\sum_{i=1}^{n}h^{l_1}(x_i;\hat\theta)\cdots h^{l_m}(x_i;\hat\theta). \tag{4}$$

In the same guise, we define the empirical counterpart of the term $A^{l_1\cdots l_m}(\theta)$ that involves estimation of the embedded population moment, i.e.

$$\hat{A}^{l_1\cdots l_m}(\theta) = n^{-1}\sum_{i=1}^{n} h^{l_1}(x_i;\theta)\cdots h^{l_m}(x_i;\theta) - \hat{\alpha}^{l_1\cdots l_m}.$$

Note that for $m = 1$ we just index the components of $h(X;\theta)$, therefore $\hat{A}^{l_1}(\theta) = A^{l_1}(\theta)$ since $\alpha^{l_1} = 0$.

For the sake of clarity, we will adopt the following conventions throughout the review of the literature and the presentation of our developments. We denote by $\mathbb{E}[\cdot]$ expectation with respect to F_0 and reserve the symbols $A^{l_1\cdots l_m}$ and $\hat{A}^{l_1\cdots l_m}$ for the special cases $A^{l_1\cdots l_m}(\theta_0)$ and $\hat{A}^{l_1\cdots l_m}(\theta_0)$, respectively. The latter shorthand will be often recycled for some other functions of θ and its use will shine through the context. We follow the summation convention, that is whenever an index is repeated in an expression the sum with respect to such index is understood, and split free and dummy indexes into the disjoint sets $\{r,s,t,u,v,w\}$ and $\{a,b,\ldots,j\}$, respectively. To summarise, for instance, $A^{rsab}A^aA^b$ is equal to $\sum_{a=1}^{d}\sum_{b=1}^{d} A^{rsab}A^aA^b$, each $r,s \in \{1,\ldots,d\}$. Finally, without any loss of generality, we display simplified expansions by assuming that the covariance matrix of $h(X;\theta_0)$ satisfies $\alpha^{rs} = \delta^{rs}$, with δ^{rs} the Kronecker delta. Whenever the formulae involved will be of any practical relevance, we will supply directly their ready-to-compute versions for the general case $\alpha^{rs} \neq \delta^{rs}$.

2.2 Bartlett correction and empirical likelihood

The investigation about second-order properties of empirical discrepancy statistics, such as the empirical likelihood, is typically based on the study of the associated signed squared root. In our setting we consider the signed root of $\Delta_{\gamma,\phi}(\theta_0)$, denoted with $n^{1/2}R_{\gamma,\phi}(\theta_0) = n^{1/2}R_{\gamma,\phi}$, which is a d-dimensional random vector chosen to fulfil $n^{-1}\Delta_{\gamma,\phi}(\theta_0) = R^a_{\gamma,\phi}R^a_{\gamma,\phi} + O_p(n^{-5/2})$. By using the technique in DiCiccio *et al.* (1991) for the derivation of the signed root of $\Delta_{-1,-1}(\theta_0)$, it is possible to write $R^r_{\gamma,\phi} = R^r_{1;\gamma,\phi} + R^r_{2;\gamma,\phi} + R^r_{3;\gamma,\phi} + O_p(n^{-2})$, where

$$\begin{aligned}
R^r_{1;\gamma,\phi} &= A^r,\\
R^r_{2;\gamma,\phi} &= -\frac{1}{2}A^{ra}A^a + \frac{1}{6}(1-\gamma)\alpha^{rab}A^aA^b,\\
R^r_{3;\gamma,\phi} &= \frac{3}{8}A^{ra}A^{ab}A^b + \frac{1}{8}(1+\phi)(1-\phi+2\gamma)A^rA^aA^a + \frac{1}{6}(1-\gamma)A^{rab}A^aA^b +\\
&+ \frac{1}{72}[9(1-\phi)(1+\phi-2\gamma) - (1-\gamma)^2]\alpha^{rab}\alpha^{bcd}A^aA^cA^d - \frac{5}{24}(1-\gamma)\alpha^{rab}A^{bc}A^aA^c +\\
&- \frac{5}{24}(1-\gamma)\alpha^{abc}A^{rb}A^aA^c + \frac{1}{24}(3\phi^2+\gamma^2-6\gamma\phi+3\gamma-1)\alpha^{rabc}A^aA^bA^c, \qquad (5)
\end{aligned}$$

so that $R^r_{j;\gamma,\phi} = O_p(n^{-j/2}), j = 1,2,3$.

The steps required to check if a member $\Delta_{\gamma,\phi}(\theta)$ of the family is Bartlett-correctable entail the computation of the cumulants of $n^{1/2}R_{\gamma,\phi}$ and the subsequent verification of the following conditions

$$\begin{array}{ll}
\kappa^{r,s,t}(n^{1/2}R_{\gamma,\phi}) = O(n^{-3/2}), & \kappa^{r,s,t,u}(n^{1/2}R_{\gamma,\phi}) = O(n^{-2}),\\
\kappa^{r,s,t,u,v}(n^{1/2}R_{\gamma,\phi}) = o(n^{-2}), & \kappa^{r,s,t,u,v,w}(n^{1/2}R_{\gamma,\phi}) = o(n^{-2}).
\end{array} \qquad (6)$$

If (6) is satisfied, then the first two cumulants determine an Edgeworth series for the distribution function of $n^{1/2}R_{\gamma,\phi}$ and the peculiar structure of such an approximation reveals that the signed root is distributed as a d-variate standard normal up to $O(n^{-3/2})$. Consequently, the k-th cumulant of $\Delta_{\gamma,\phi}(\theta_0)$ is

$$2^{k-1}(k-1)!d\left\{d^{-1}\mathbb{E}[\Delta_{\gamma,\phi}(\theta_0)]\right\}^k + O(n^{-3/2}), \tag{7}$$

where $2^{k-1}(k-1)d!$ is the k-th cumulant of a chi-squared random variable with d degrees of freedom (DiCiccio *et al.*, 1991, Sect. 4). Expression (7) encloses the essence of the Bartlett correction: division of $\Delta_{\gamma,\phi}(\theta)$ by $d^{-1}\mathbb{E}[\Delta_{\gamma,\phi}(\theta_0)]$ yields a statistic whose distribution is closer to the one of a chi-squared variate with d degrees of freedom.

Exceedingly tedious calculations, not reported, show that cumulants of $R_{\gamma,\phi}$ are of the correct size only for the pair $(\gamma,\phi)=(-1,-1)$, i.e the only member belonging to the family (2) that admits the Bartlett correction is the empirical likelihood function $\Delta_{-1,-1}(\theta)$. The Bartlett-corrected version of the empirical likelihood is

$$\Delta^B_{-1,-1}(\theta) = d\Delta_{-1,-1}(\theta)\mathbb{E}\left[\Delta_{-1,-1}(\theta_0)\right]^{-1} = \Delta_{-1,-1}(\theta)\left[1-(nd)^{-1}b\right],$$

and the second equality follows from the expression

$$\mathbb{E}\left[\Delta_{-1,-1}(\theta_0)\right] = \kappa^{a,a}(n^{1/2}R_{-1,-1}) + \kappa^a(n^{1/2}R_{-1,-1})\kappa^a(n^{1/2}R_{-1,-1}) = d + n^{-1}b + O(n^{-2}).$$

The term $b = b(\theta_0)$, known as the Bartlett factor, is

$$b = \frac{1}{2}\alpha^{abcd}\alpha_{ab}\alpha_{cd} - \frac{1}{3}\alpha^{abc}\alpha^{def}\alpha_{ad}\alpha_{be}\alpha_{cf}, \tag{8}$$

with α_{rs} the inverse of α^{rs}. The Bartlett correction has the effect of reducing the coverage error of empirical likelihood confidence sets by one order of magnitude, i.e.

$$\Pr\left\{\Delta^B_{-1,-1}(\theta_0) \leq c_\nu\right\} = \nu + O(n^{-2}), \tag{9}$$

where c_ν is the ν-quantile of a chi-squared random variable with d degrees of freedom.

It should be noted that $\Delta^B_{-1,-1}(\theta)$ depends on unknown population moments through the Bartlett factor b. In practice these moments can be replaced by their root-n consistent estimates in (4) without affecting the error in approximation (9), since the corresponding estimated Bartlett factor

$$\hat{b} = \frac{1}{2}\hat{\alpha}^{abcd}\hat{\alpha}_{ab}\hat{\alpha}_{cd} - \frac{1}{3}\hat{\alpha}^{abc}\hat{\alpha}^{def}\hat{\alpha}_{ad}\hat{\alpha}_{be}\hat{\alpha}_{cf} \tag{10}$$

is root-n consistent for b.

2.3 Adjusted Bartlett-corrected empirical likelihood

The adjusted empirical likelihood (Chen *et al.*, 2008) originally stemmed from the need to tackle the possible failure of the convex hull condition for $\Delta_{-1,-1}(\theta)$. Suppose we are willing to compute $\Delta_{-1,-1}$ at a generic value θ; then, $\Delta_{-1,-1}(\theta)$ is finite only if θ is such that the convex hull of $\{h(x_1;\theta),\ldots,h(x_n;\theta)\}$ contains the null vector. However, when n or the ratio n/d is moderate to small the convex hull may fail to contain the origin. The key idea to make $\Delta_{-1,-1}(\theta)$ well defined is to add a pseudo-observation to $h(x_1;\theta),\ldots,h(x_n;\theta)$, of the form $h^r(x_{n+1};\theta) = -a_nA^r(\theta)$, so that the null vector is always an interior point of the convex hull of $\{h(x_1;\theta),\ldots,h(x_{n+1};\theta)\}$.

This approach is successful as long as $a_n > 0$, because the quantities $-A^r(\theta)$ and $h^r(x_{n+1};\theta)$ are on opposite sides of the origin.

Chen *et al.* (2008) show that if $a_n = o_p(n^{2/3})$, then first-order asymptotic properties of the empirical likelihood are retained. This recipe has been then refined by Liu and Chen (2010) by setting

$$h^r(x_{n+1};\theta) = -bA^r(\theta)/2d, \tag{11}$$

which both overcomes the possible failure of convex hull condition and makes the resulting empirical likelihood Bartlett-corrected at a stroke. Actually, when $d > 1$ there is still a chance that the convex hull of $\{h(x_1;\theta),\dots,h(x_{n+1};\theta)\}$ will not contain the null vector, since b might be negative. The solution is to add a further pseudo-observation, $h(x_{n+2};\theta)$, and to suitably split the Bartlett factor into the sum of two non-negative terms satisfying $b = b_1 - b_2$. This formulation leads to the following allocation of the pseudo-observations

$$h^r(x_{n+1};\theta) = -b_1A^r(\theta)/2d \quad \text{and} \quad h^r(x_{n+2};\theta) = b_2A^r(\theta)/2d. \tag{12}$$

According to Liu and Chen (2010, Sect. 3.3) a possible representation of the components of the Bartlett factor is (summation convention does not apply)

$$\begin{aligned} b_1 &= \frac{1}{2}\sum_{a,b}\alpha^{aabb}\alpha_{ab}\alpha_{ab} - \frac{1}{3}\sum_{a,b,c}\alpha^{abb}\alpha^{bcc}\alpha_{ab}\alpha_{ac}\alpha_{bc} + \sum_{a,b,d,e}\alpha^{adbe}\alpha_{ad}\alpha_{be} - \sum_{a,b,c,d,e}\alpha^{abc}\alpha^{dee}\alpha_{ad}\alpha_{be}\alpha_{ce} \\ b_2 &= \sum_{a,b,c,d,e}\alpha^{abc}\alpha^{dee}\alpha_{ad}\alpha_{be}\alpha_{ce} + 2\sum_{a,b,c,d,e,f}\alpha^{abc}\alpha^{def}\alpha_{ad}\alpha_{be}\alpha_{cf}, \end{aligned}$$

where, only in this formula and whenever we deal with b_1 and b_2, we impose the restriction $d < e < f$.

If we denote with $\tilde{\Delta}^B_{-1,-1}(\theta_0)$ the adjusted Bartlett-corrected empirical likelihood based on $h(x_1;\theta_0),\dots,h(x_{n+1};\theta_0),h(x_{n+2};\theta_0)$, with $h(x_{n+1};\theta_0) = -b_1A^r/2d$ and $h(x_{n+2};\theta_0) = b_2A^r/2d$, then, as in (9), $\Pr\left\{\tilde{\Delta}^B_{-1,-1}(\theta_0) \le c_\nu\right\} = \nu + O(n^{-2})$. In analogy with the canonical Bartlett-corrected empirical likelihood $\Delta^B_{-1,-1}(\theta)$, the coverage error of the adjusted Bartlett-corrected empirical likelihood confidence sets is still $O(n^{-2})$ whenever the unknown Bartlett factor in (11) is estimated by (10) as well as b_1 and b_2 in (12) are replaced by

$$\begin{aligned} \hat{b}_1 &= \frac{1}{2}\sum_{a,b}\hat{\alpha}^{aabb}\hat{\alpha}_{ab}\hat{\alpha}_{ab} - \frac{1}{3}\sum_{a,b,c}\hat{\alpha}^{abb}\hat{\alpha}^{bcc}\hat{\alpha}_{ab}\hat{\alpha}_{ac}\hat{\alpha}_{bc} + \sum_{a,b,d,e}\hat{\alpha}^{adbe}\hat{\alpha}_{ad}\hat{\alpha}_{be} - \sum_{a,b,c,d,e}\hat{\alpha}^{abc}\hat{\alpha}^{dee}\hat{\alpha}_{ad}\hat{\alpha}_{be}\hat{\alpha}_{ce} \\ \hat{b}_2 &= \sum_{a,b,c,d,e}\hat{\alpha}^{abc}\hat{\alpha}^{dee}\hat{\alpha}_{ad}\hat{\alpha}_{be}\hat{\alpha}_{ce} + 2\sum_{a,b,c,d,e,f}\hat{\alpha}^{abc}\hat{\alpha}^{def}\hat{\alpha}_{ad}\hat{\alpha}_{be}\hat{\alpha}_{cf}, \end{aligned}$$

respectively. Li *et al.* (2011) extended this approach to the exponential empirical likelihood $\Delta_{-1,0}(\theta)$.

3 Adjusted Bartlett-corrected generalised power divergence family

3.1 Theoretical adjustment

By exploiting the idea that adding a perturbation to $h(x_1;\theta),\dots,h(x_n;\theta)$ can affect the high-order behaviour of the empirical likelihood, in Theorem 1 below we argue that it suffices to augment the sample with at most three pseudo-observations to make each member of family (2) both adjusted and Bartlett-correctable. Specifically, the term adjusted refers, in the same guise as for the adjusted empirical likelihood, to the fact that each member is released from the convex hull constraint.

Henceforth, quantities computed on an augmented pseudo-sample $h(x_1;\theta),\ldots,h(x_{n+k};\theta)$ of arbitrary size $n+k$ are denoted with an upper bar, $k\geq 1$. Therefore, in particular, for each pair (γ,ϕ), $\bar{\Delta}_{\gamma,\phi}(\theta_0)$ and $\bar{R}_{\gamma,\phi}$ stand for $\Delta_{\gamma,\phi}(\theta_0)$ and $R_{\gamma,\phi}$ based on $h(x_1;\theta_0),\ldots,h(x_{n+k};\theta_0)$, respectively.

To establish our results we will need the following assumptions:

A1 α^{rs} is finite and of full rank d;

A2 $\mathbb{E}[\|h(X;\theta_0)\|^{18}]<\infty$;

A3 $\limsup_{\|\zeta\|\to\infty}\mathbb{E}\left[\exp\left\{i\zeta^a h^a(X;\theta_0)\right\}\right]<1$, $i=-1^{1/2}$, $\zeta\in\mathbb{R}^d$.

Assumption **A1** is the essential requirement to establish Wilk's theorem for statistics derived from the generalised power divergence family. Assumption **A2** ensures that the 6-th moment of $n^{1/2}\bar{R}_{\gamma,\phi}$ is finite and together with Cramr's condition **A3** are needed to develop a valid Edgeworth series for the distribution function of $n^{1/2}\bar{R}_{\gamma,\phi}$ up to $O(n^{-3/2})$ (Bhattacharya and Ghosh, 1978; Skovgaard, 1981; Liu and Chen, 2010).

Before presenting our main theorem, in the following lemma we provide the size of the pseudo-observations and quantify their impact on the signed root $n^{1/2}\bar{R}_{\gamma,\phi}$.

Lemma 1 *Denote with $\Omega(\theta)=h(x_{n+1};\theta)$, $\Gamma(\theta)=h(x_{n+2};\theta)$, and $\Xi(\theta)=h(x_{n+3};\theta)$ pseudo-observations whose sizes are assumed to be $O_p(1)$, $O_p(n^{-1/2})$, and $O_p(n^{-1})$, respectively. Let $\Omega=\Omega(\theta_0)$, $\Gamma=\Gamma(\theta_0)$, and $\Xi=\Xi(\theta_0)$. Then, under assumptions* **A1** *and* **A2**,

$$n^{-1}\bar{\Delta}_{\gamma,\phi}(\theta_0)=\bar{R}^a_{1;\gamma,\phi}\bar{R}^a_{1;\gamma,\phi}+\bar{R}^a_{2;\gamma,\phi}\bar{R}^a_{2;\gamma,\phi}+2\bar{R}^a_{1;\gamma,\phi}\bar{R}^a_{2;\gamma,\phi}+2\bar{R}^a_{1;\gamma,\phi}\bar{R}^a_{3;\gamma,\phi}+O_p(n^{-5/2}),$$

where

$$\begin{aligned}\bar{R}^r_{1;\gamma,\phi}&=R^r_{1;\gamma,\phi},\\ \bar{R}^r_{2;\gamma,\phi}&=R^r_{2;\gamma,\phi}+\frac{1}{n}\Omega^r,\\ \bar{R}^r_{3;\gamma,\phi}&=R^r_{3;\gamma,\phi}+\frac{1}{n}\Gamma^r-\frac{1}{2n}A^{ra}\Omega^a+\frac{1}{3n}(1-\gamma)\alpha^{rab}A^a\Omega^b+\frac{1}{2n}\Omega^rA^a\Omega^a,\end{aligned}$$

with $R^r_{j;\gamma,\phi}$, $j=1,2,3$, defined in (5)*. Thus, $\bar{R}^r_{j;\gamma,\phi}=O_p(n^{-j/2})$.*

Lemma 1 states that only pseudo-observations of size $O_p(1)$ and $O_p(n^{-1/2})$ can affect the signed squared root of $\bar{\Delta}_{\gamma,\phi}(\theta_0)$ up to $O_p(n^{-3/2})$. Specifically, these observations enter in the expansion at second and third order, respectively. This behaviour is the key to turn each member of family (2) into an adjusted Bartlett-correctable one, as stated in the following theorem.

Theorem 1 *Under assumptions* **A1**, **A2**, *and* **A3**, *for every $\delta \geq 1/2$, if we set*

$$
\begin{aligned}
\Omega^r(\theta) &= \frac{n}{6}(1+\gamma)\alpha^{rab}\alpha_{ac}\alpha_{bd}A^c(\theta)A^d(\theta), && (13)\\
\Gamma^r(\theta) &= \frac{n}{6}\Big\{-\frac{6}{8}(1+\phi)(1-\phi+2\gamma)\alpha_{ab}A^r(\theta)A^a(\theta)A^b(\theta) - (1+\gamma)\alpha_{ac}\alpha_{bd}A^{rab}(\theta)A^c(\theta)A^d(\theta)+\\
&- (1+\gamma)\alpha^{rab}\alpha_{af}\alpha_{bc}\alpha_{de}A^{cd}(\theta)A^e(\theta)A^f(\theta) - (1+\gamma)\alpha^{abc}\alpha_{af}\alpha_{be}\alpha_{cd}A^{rd}(\theta)A^f(\theta)A^e(\theta)+\\
&+ \frac{1}{12}(9\phi^2+5\gamma^2-18\gamma\phi+16\gamma+20)\alpha^{rab}\alpha^{cde}\alpha_{af}\alpha_{bc}\alpha_{dg}\alpha_{eh}A^f(\theta)A^g(\theta)A^h(\theta)+\\
&- \frac{1}{4}(3\phi^2+\gamma^2-6\gamma\phi+3\gamma+5)\alpha^{rabc}\alpha_{ad}\alpha_{be}\alpha_{cf}A^d(\theta)A^e(\theta)A^f(\theta)+\\
&- \frac{1}{12}n(1+\gamma)^2\alpha^{rbc}\alpha^{ade}\alpha_{af}\alpha_{bg}\alpha_{ch}\alpha_{di}\alpha_{ej}A^f(\theta)A^g(\theta)A^h(\theta)A^i(\theta)A^j(\theta)\Big\}, && (14)\\
\Xi^r(\theta) &= -n^{-\delta}\left\{n(n+2)^{-1}[A^r(\theta)+n^{-1}\Omega^r(\theta)+n^{-1}\Gamma^r(\theta)]\right\}, && (15)
\end{aligned}
$$

then

$$\bar{R}_{\gamma,\phi} = R_{-1,-1} + O_p(n^{-2}),$$

which implies, for every pair (γ, ϕ),

$$
\begin{aligned}
&\kappa^{r,s,t}(n^{1/2}\bar{R}_{\gamma,\phi}) = O(n^{-3/2}), && \kappa^{r,s,t,u}(n^{1/2}\bar{R}_{\gamma,\phi}) = O(n^{-2}),\\
&\kappa^{r,s,t,u,v}(n^{1/2}\bar{R}_{\gamma,\phi}) = o(n^{-2}), && \kappa^{r,s,t,u,v,w}(n^{1/2}\bar{R}_{\gamma,\phi}) = o(n^{-2}).
\end{aligned}
$$

Therefore, each member of the family $\{\bar{\Delta}_{\gamma,\phi}(\theta)\}$ based on the pseudo-observations $\Omega(\theta)$, $\Gamma(\theta)$, and $\Xi(\theta)$ is adjusted and Bartlett-correctable.

The result in Theorem 1 yields an adjusted Bartlett-correctable generalised power divergence family by forcing the signed squared root of each element of the family to behave like the one of the empirical likelihood up to $O_p(n^{-3/2})$, and by further accounting for the convex hull constraint. To better understand the definition of $\bar{\Delta}_{\gamma,\phi}(\theta)$, we can conceptually split the pseudo-observations into two groups consistently with their intervention: $\Omega(\theta)$ and $\Gamma(\theta)$ enable each element of the family to be Bartlett-correctable, instead $\Xi(\theta)$ generates a slight perturbation, not appreciable up to $O_p(n^{-3/2})$, that allow the family to overcome the possibility that the convex hull of $\{h(x_1;\theta),\ldots,h(x_n;\theta),\Omega(\theta),\Gamma(\theta)\}$ does not contain the null vector.

Note that if one of the theoretical pseudo-observations $\Omega(\theta)$ or $\Gamma(\theta)$ is null, then it is unnecessary to include it in the pseudo-sample. Inspection of formulae (13), (14), and (15), reveals that γ may contribute to cancel $\Omega(\theta)$ and $\Gamma(\theta)$, whereas ϕ may act likewise on $\Gamma(\theta)$ only, i.e. γ is accountable for the matching of $R_{2;-1,-1}$ and $R_{3;-1,-1}$, whereas ϕ for the latter only. It turns out that for some specific pairs (γ,ϕ) we would need less than three artificial observations to compute $\bar{\Delta}_{\gamma,\phi}(\theta)$. In Table 1 we provide a comprehensive list of the possible combinations of tuning parameters, along with the resulting pseudo-sample size, the pseudo-observations needed, and the current expression of pseudo-observation $\Xi(\theta)$ in each case. For instance, when the empirical likelihood is considered, we need to accommodate for the convex hull constraint only, implying that $\bar{\Delta}_{-1,-1}(\theta)$ need to be computed on a pseudo-sample of size $n+1$, where the pseudo-observation $\Xi(\theta)$ reduces to $-n^{-\delta}A(\theta)$.

The guidelines for the allocation of the artificial observations in Theorem 1 tacitly imply that the Bartlett factor for each member $\bar{\Delta}_{\gamma,\phi}(\theta)$ of the new adjusted Bartlett-correctable generalised power divergence family is the one derived for the empirical likelihood. In the following corollary,

Table 1: Combination of tuning parameters and associated pseudo-observations for Theorem 1. In the fourth and fifth columns the symbol "-" indicates that the corresponding pseudo-observation is not needed. Whenever the pseudo-observations $\Omega(\theta)$ and $\Gamma(\theta)$ are present we report the corresponding reference in the text, whereas the expression for $\Xi(\theta)$ is indicated case by case.

Tuning parameters		Pseudo-sample size	$\Omega(\theta)$	$\Gamma(\theta)$	$\Xi(\theta)$
$\gamma \neq -1$	each ϕ	$n+3$	(13)	(14)	$-n^{-\delta}\left\{n(n+2)^{-1}[A(\theta)+n^{-1}\Omega(\theta)+n^{-1}\Gamma(\theta)]\right\}$
$\gamma = -1$	$\phi \neq -1$	$n+2$	-	(14)	$-n^{-\delta}\left\{n(n+1)^{-1}[A(\theta)+n^{-1}\Gamma(\theta)]\right\}$
$\gamma = -1$	$\phi = -1$	$n+1$	-	-	$-n^{-\delta}A(\theta)$

we state that Bartlett-corrected statistics are derived from $\left\{\bar{\Delta}_{\gamma,\phi}(\theta)\right\}$ in the traditional way, i.e. by affine transformation.

Corollary 1 *Under assumptions of Theorem 1, consider $\bar{\Delta}_{\gamma,\phi}(\theta)$ which is computed on the following pseudo-sample*

$$h(x_1;\theta),\ldots,h(x_n;\theta),\,\Omega(\theta),\,\Gamma(\theta),\,\Xi(\theta),$$

where $\Omega(\theta)$, $\Gamma(\theta)$, and $\Xi(\theta)$ are as in Table 1 according to the specific pair (γ,ϕ). Then

$$Pr\left\{\bar{\Delta}^B_{\gamma,\phi}(\theta_0)\leq c_\nu\right\}=\nu+O(n^{-2}),$$

with

$$\bar{\Delta}^B_{\gamma,\phi}(\theta_0)=\bar{\Delta}_{\gamma,\phi}(\theta_0)[1-(dn)^{-1}b],$$

The quantity b is the Bartlett factor provided in (8) *and c_ν is defined in* (9).

Empirical likelihood as well as other (standard) empirical discrepancy statistics are bounded above when n or the ratio n/d is moderate to small, implying that the derived confidence regions may attain only confidence levels restricted in the interval $(0,\upsilon)$, $\upsilon \ll 1$ (Tsao, 2004). The artificial observations $\Xi(\theta)$ and $-b_1A(\theta)/2d$, $b_2A(\theta)/2d$ may act likewise on $\bar{\Delta}_{\gamma,\phi}(\theta)$ and $\tilde{\Delta}^B_{-1,-1}(\theta)$, respectively, meaning that, in the worst case scenario, they may reinstate an upper bound which is relevant for practical purposes. Nevertheless, $\Xi(\theta)$ depends the positive scaling factor $n^{-\delta}$, and δ may be tuned to regulate the upper bound for the members of $\left\{\bar{\Delta}_{\gamma,\phi}(\theta)\right\}$.

By generalising the result in Emerson and Owen (2009, Sect. 3.1), it can be easily seen that the quantity

$$\frac{2}{\gamma(\gamma+1)}\left\{n\left[\left(\frac{a(n+j)}{n(a+j)}\right)^{\gamma+1}-1\right]+j\left[\left(\frac{(n+j)}{(a+j)}\right)^{\gamma+1}-1\right]\right\}$$

represents the upper bound for members of $\left\{\bar{\Delta}_{\gamma,\phi}(\theta)\right\}$ once we set $a=n^{-\delta}$ and $j=1$. The special cases $\gamma\in\{-1,0\}$ are elicited by taking limits. For $\gamma=-1$ we have

$$-2\left[n\log\left(\frac{a(n+j)}{n(a+j)}\right)+j\log\left(\frac{n+j}{a+j}\right)\right], \tag{16}$$

whereas for $\gamma = 0$

$$2\left[\frac{n+j}{a+j}\right]\left[a\log\left(\frac{a(n+j)}{n(a+j)}\right) + j\log\left(\frac{n+j}{a+j}\right)\right].$$

The bound for $\tilde{\Delta}^B_{-1,-1}(\theta)$ is recovered by plugging $a = b/2d = (b_1 - b_2)/2d$ and $j = 2$ in (16). As n diverges, bounds for $\bar{\Delta}_{\gamma,\phi}(\theta)$ and $\tilde{\Delta}^B_{-1,-1}(\theta)$ tend to infinity; however, for fixed n their value is shown to be inversely proportional in a. The bound on members of $\{\bar{\Delta}_{\gamma,\phi}(\theta)\}$ can be appreciably less narrow than that on $\tilde{\Delta}^B_{-1,-1}(\theta)$, because the former depends on $a = n^{-\delta} = o(1)$ rather than on $a = b/2d = O(1)$. As an example, we consider inference on the mean vector of a standard d-variate normal distribution when $n = 40$ and $q = d = 10$ (see Section 4.1). Then $b_1 = q^2/2, b_2 = 0$, whereby $a = q/4$, $j = 1$, and the upper bound for $\tilde{\Delta}^B_{-1,-1}(\theta)$ is about 20, i.e. confidence sets for the mean can attain at most the nominal level 0.97. On the other hand, if we consider the members $\bar{\Delta}_{-1,\phi}(\theta)$ and set $\delta = 1/2$, which is the minimum admissible value for δ imposed by Lemma 1, the upper bound is about 150, thus irrelevant for practical purposes.

The example above provides a clear-cut view of the effects of pseudo observations $\Xi(\theta)$ and $-b_1A(\theta)/2d$, $b_2A(\theta)/2d$ on the induced upper bounds for $\bar{\Delta}_{\gamma,\phi}(\theta)$ and $\tilde{\Delta}^B_{-1,-1}(\theta)$. Specifically, if we tie the resolution of the convex hull problem to high-order asymptotics as in Liu and Chen (2010), then the resulting pseudo-observations cannot be tuned to alleviate the side effects of the upper bound as their manipulation would invalidate the Bartlett correction for $\tilde{\Delta}^B_{-1,-1}(\theta)$. As opposed to $\tilde{\Delta}^B_{-1,-1}(\theta)$, the upper bound for $\bar{\Delta}_{\gamma,\phi}(\theta)$ can be readily calculated as it does not depend on unknown population moments in the Bartlett factor. Consequently, the user has the opportunity to select δ in order to be sure that the inference function is allowed to attain the desired confidence level.

In Section 4, we run our numerical investigations for the value $\delta = 1/2$ and the results highlight that this value is appropriate for our examples which deal with a fairly selection of values of n and d.

3.2 Empirical adjustment

The adjusted Bartlett-correctable generalised power divergence family $\{\bar{\Delta}_{\gamma,\phi}(\theta)\}$ depends on unknown population moments in the pseudo-observations $\Omega(\theta)$, $\Gamma(\theta)$, and $\Xi(\theta)$, and whenever they are replaced by their root-n consistent estimates (4), the result in Theorem 1 may be struck down. Recall that $\Omega = \Omega(\theta_0)$, $\Gamma = \Gamma(\theta_0)$, and $\Xi = \Xi(\theta_0)$. On the one hand, the empirical counterpart of Ω satisfies

$$\hat{\Omega} = \frac{n}{6}(1+\gamma)\hat{\alpha}^{rab}\hat{\alpha}_{ac}\hat{\alpha}_{bd}\hat{A}^c\hat{A}^d = \Omega + O_p(n^{-1/2}),$$

and once such expression is plugged back in $\bar{R}_{2;\gamma,\phi}$ the estimation error produces a disturbance of the same size of Γ that affects the actual expression of $\bar{R}_{3;\gamma,\phi}$. On the other hand, estimation of Γ involves estimation of population moments in the terms A^{rs} and A^{rst} and this process generates a reminder of size $O_p(n^{-1/2})$, again of the same magnitude of Γ, which in turn modifies the expression of $\bar{R}_{3;\gamma,\phi}$. As an aside, the pseudo-observation Ξ leaves the conclusions of Theorem 1 always unchanged as its estimate $\hat{\Xi}$ is $O_p(n^{-1/2-\delta})$, $\delta \geq 1/2$, and by Lemma 1 we are aware that terms of such size do not enter into the expression of $\bar{R}_{\gamma,\phi}$ up to $O_p(n^{-3/2})$.

Note that for the pairs $(-1, \phi)$, each ϕ, the adjusted Bartlett-correctable generalised power divergence family based on the empirical versions of the pseudo-observations in Theorem 1 is still Bartlett-correctable, because for $\gamma = -1$ we have $\Omega = 0$ and the terms A^{rs} and A^{rst} in Γ vanish. This happy chance leads to an estimation error for Γ whose size is $O_p(n^{-1})$, i.e. smaller than the size of Γ.

In the following we provide a revised version of Theorem 1 which is valid for every pair of (γ, ϕ) as it takes into an account the disturbance induced on the expression of $\bar{R}_{\gamma,\phi}$ by the estimation of population moments in Ω and Γ.

Theorem 2 *Under assumptions* **A1**, **A2**, **A3**, *for every* $\delta \geq 1/2$, *if*

(i) $\gamma = -1$ *and,*

$$\begin{aligned}
\hat{\Gamma}^r(\theta) &= \frac{n}{8}(1+\phi)^2 \Big\{ \hat{\alpha}_{ab}\hat{A}^r(\theta)\hat{A}^a(\theta)\hat{A}^b(\theta) - \hat{\alpha}^{rabc}\hat{\alpha}_{ad}\hat{\alpha}_{be}\hat{\alpha}_{cf}\hat{A}^d(\theta)\hat{A}^e(\theta)\hat{A}^f(\theta) + \\
&+ \hat{\alpha}^{rab}\hat{\alpha}^{cde}\hat{\alpha}_{bc}\hat{\alpha}_{af}\hat{\alpha}_{dg}\hat{\alpha}_{eh}\hat{A}^f(\theta)\hat{A}^g(\theta)\hat{A}^h(\theta) \Big\}, \qquad (17) \\
\hat{\Xi}^r(\theta) &= -n^{-\delta}[n(n+1)^{-1}][\hat{A}^r(\theta) + n^{-1}\hat{\Gamma}^r(\theta)];
\end{aligned}$$

(ii) $\gamma \neq -1$ *and,*

$$\begin{aligned}
\hat{\Omega}^r(\theta) &= \frac{n}{6}(1+\gamma)\hat{\alpha}^{rab}\hat{\alpha}_{ac}\hat{\alpha}_{bd}\hat{A}^c(\theta)\hat{A}^d(\theta), \qquad (18) \\
\hat{\Gamma}^r(\theta) &= \frac{n}{6}\Big\{ \frac{6}{8}(\phi^2 - 2\gamma\phi + 2\gamma + 3)\hat{\alpha}_{ab}\hat{A}^r(\theta)\hat{A}^a(\theta)\hat{A}^b(\theta) + \\
&- \frac{1}{4}(3\phi^2 + \gamma^2 - 6\gamma\phi + 3\gamma + 5)\hat{\alpha}^{rabc}\hat{\alpha}_{ad}\hat{\alpha}_{be}\hat{\alpha}_{cf}\hat{A}^d(\theta)\hat{A}^e(\theta)\hat{A}^f(\theta) + \\
&+ \frac{1}{12}n(1+\gamma)^2\hat{\alpha}^{rbc}\hat{\alpha}^{ade}\hat{\alpha}_{af}\hat{\alpha}_{bg}\hat{\alpha}_{ch}\hat{\alpha}_{di}\hat{\alpha}_{el}\hat{A}^f(\theta)\hat{A}^g(\theta)\hat{A}^h(\theta)\hat{A}^i(\theta)\hat{A}^l(\theta) + \qquad (19) \\
&+ \frac{1}{12}(9\phi^2 + 5\gamma^2 - 18\gamma\phi + 16\gamma + 20)\hat{\alpha}^{rab}\hat{\alpha}^{cde}\hat{\alpha}_{bc}\hat{\alpha}_{af}\hat{\alpha}_{dg}\hat{\alpha}_{eh}\hat{A}^f(\theta)\hat{A}^g(\theta)\hat{A}^h(\theta) \Big\}, \\
\hat{\Xi}^r(\theta) &= -n^{-\delta}[n(n+2)^{-1}][\hat{A}^r(\theta) + n^{-1}\hat{\Omega}^r(\theta) + n^{-1}\hat{\Gamma}^r(\theta)],
\end{aligned}$$

by further assuming that

$$\hat{\alpha}^{rs} - n^{-1}\sum_{i=1}^{n}[h^r(x_i;\theta_0) - \hat{A}^r][h^s(x_i;\theta_0) - \hat{A}^s] = O_p(n^{-1})$$

$$\hat{\alpha}^{rst} - n^{-1}\sum_{i=1}^{n}[h^r(x_i;\theta_0) - \hat{A}^r][h^s(x_i;\theta_0) - \hat{A}^s][h^t(x_i;\theta_0) - \hat{A}^t] = O_p(n^{-1}),$$

then for each pair (γ, ϕ)

$$\bar{R}_{\gamma,\phi} = R_{-1,-1} + O_p(n^{-2}).$$

Therefore each member of the family $\{\bar{\Delta}_{\gamma,\phi}(\theta)\}$ *based on the pseudo-observations* $\hat{\Omega}(\theta)$, $\hat{\Gamma}(\theta)$, *and* $\hat{\Xi}(\theta)$ *given either in part (i) or (ii) is adjusted and Bartlett-correctable.*

Surprisingly, when $\gamma \neq -1$ a straightforward transition from the theoretical adjustment Ω to its empirical counterpart $\hat{\Omega}$ entails a reminder that modifies the expression of $\bar{R}_{3;\gamma,\phi}$ so that the difference $\bar{R}_{3;\gamma,\phi} - R_{3;-1,-1}$ is free of the terms involving A^{rs} and A^{rst}. This leads to a theoretical adjustment Γ whose estimation does not alter the expression of $\bar{R}_{3;\gamma,\phi}$ up to $O_p(n^{-3/2})$. In this connection, we claim that it is possible to replace $\hat{\Omega}(\theta)$ with alternatives root-n consistent estimates, say $\widehat{\Omega}(\theta)$, as long as $\hat{\Omega}(\theta) - \widehat{\Omega}(\theta) = O_p(n^{-1})$, because the aforesaid cancellations occur for the specific $O_p(n^{-1/2})$ error term induced by $\hat{\Omega}(\theta)$.

In Table 2 are summarised the possible outcomes of Theorem 2 according to each pair (γ, ϕ). The first row of the table, $\gamma \neq -1$, is relevant for part (ii), whereas the last ones, $\gamma = -1$, for part (i).

Table 2: Combination of tuning parameters and associated pseudo-observations for Theorem 2. In the fourth and fifth columns the symbol "-" indicates that the corresponding pseudo-observation is not needed. Whenever the pseudo-observations $\hat{\Omega}(\theta)$ and $\hat{\Gamma}(\theta)$ are present we report the reference in the text, whereas the expression for $\hat{\Xi}(\theta)$ is indicated case by case.

Tuning parameters		Pseudo-sample size	$\hat{\Omega}(\theta)$	$\hat{\Gamma}(\theta)$	$\hat{\Xi}(\theta)$
$\gamma \neq -1$	each ϕ	$n+3$	(18)	(19)	$-n^{-\delta}\left\{n(n+2)^{-1}[\hat{A}(\theta) + n^{-1}\hat{\Omega}(\theta) + n^{-1}\hat{\Gamma}(\theta)]\right\}$
$\gamma = -1$	$\phi \neq -1$	$n+2$	-	(17)	$-n^{-\delta}\left\{n(n+1)^{-1}[\hat{A}(\theta) + n^{-1}\hat{\Gamma}(\theta)]\right\}$
$\gamma = -1$	$\phi = -1$	$n+1$	-	-	$-n^{-\delta}\hat{A}(\theta)$

To close this section, we provide the revised version of Corollary 1.

Corollary 2 *Under assumptions of Theorem 2, consider $\bar{\Delta}_{\gamma,\phi}(\theta)$ computed with the estimated pseudo-observations $\hat{\Omega}(\theta)$, $\hat{\Gamma}(\theta)$, and $\hat{\Xi}(\theta)$ given in Table 2 according to the specific pair (γ, ϕ). Then the same result stated in Corollary 1 applies for*

$$\bar{\Delta}^B_{\gamma,\phi}(\theta_0) = \bar{\Delta}_{\gamma,\phi}(\theta_0)[1 - (dn)^{-1}\hat{b}],$$

where $\hat{b}$ is the estimated Bartlett factor given in (10).

4 Simulation study

In the present section we provide empirical evidence, through a simulation study, about the effectiveness of the Bartlett correction in reducing the coverage error of confidence sets derived from some elements of the generalised power divergence family. Hence, we will compare actual accuary of confidence sets based of some adjusted empirical discrepancies $\bar{\Delta}_{\gamma,\phi}(\theta)$ and their Bartlett-corrected versions $\bar{\Delta}^B_{\gamma,\phi}(\theta)$. We further aim to make a comparison to the accuracy of confidence sets based on the adjusted Bartlett-corrected empirical likelihood, $\tilde{\Delta}^B_{-1,-1}(\theta)$.

We consider four popular members within the generalised power divergence family, namely the empirical likelihood $\Delta_{-1,-1}(\theta)$, the exponential empirical likelihood $\Delta_{-1,0}(\theta)$, the maximum entropy $\Delta_{0,0}(\theta)$, and the Euclidean likelihood $\Delta_{1,1}(\theta)$ as well as two additional members, namely $\Delta_{-1,-2}(\theta)$ and $\Delta_{-1,-1/2}(\theta)$. This choice is meant to deal with a small yet representative number of instances in family (2): a member which is already Bartlett-correctable, $\Delta_{-1,-1}(\theta)$, and members $\Delta_{0,0}(\theta)$, $\Delta_{1,1}(\theta)$, $\Delta_{-1,0}(\theta)$, $\Delta_{-1,-1/2}(\theta)$, and $\Delta_{-1,-2}(\theta)$ whose distributional behaviour is respectively gradually closer to the one of the empirical likelihood. Effectively this means that $\Delta_{0,0}(\theta)$ and $\Delta_{1,1}(\theta)$ need both pseudo-observations $\Omega(\theta)$ and $\Gamma(\theta)$ to be Bartlett-correctable, whereas the remaining ones need $\Gamma(\theta)$ only.

For the sake of comparison with the existing literature, we firstly consider inference about both a scalar and vector-valued population mean. Later on, we broaden our investigation to encompass more general vector-valued parameters by considering examples in the regression and pairwise likelihood contexts. Without attempting a full discussion on pairwise likelihoods we defer the reader to Lindsay (1988) and Varin *et al.* (2011) and provide the essential details about the usefulness of empirical discrepancies in such setting in Section 4.3.

Adjusted Bartlett-corrected versions of the considered statistics are computed by resorting to the estimated adjustments in Table 2, with $\delta = 1/2$, and to the estimated Bartlett factor (10). For the sake of comparison, the Bartlett factor is estimated by replacing unknowns moments with the enhanced moment estimates suggested in Liu and Chen (2010, Sect. 3.3) rather than to the ones in (4); this is also the case for the estimation of the components of the Bartlett factor b_1 and b_2 for the computation of the adjusted Bartlett-corrected empirical likelihood $\tilde{\Delta}^B_{-1,-1}(\theta)$. Adjusted Bartlett-corrected maximum entropy and Euclidean likelihood have been considered only for the population mean examples because the function $h(x;\theta) = x - \theta$ satisfies the requirements in part (ii) of Theorem 2.

The number of Monte Carlo trials in our simulations is 100.000.

4.1 Population mean

Simulations for the scalar mean are conducted by generating samples of size 15 and 30. Univariate distributions are chosen with increasing values of both skewness and kurtosis: standard normal (d_1), exponential with unit mean (d_2), chi-squared with one degree of freedom (d_3), and standard log-normal (d_4). The results are summarised in Table 3. Empirical coverage probabilities for standard empirical likelihood $\Delta_{-1,-1}(\theta)$ are far away from the nominal levels and the situation worsens as the underlying distributions become markedly skewed and leptokurtic. The adjusted and Bartlett-corrected versions, $\tilde{\Delta}^B_{-1,-1}(\theta)$ and $\bar{\Delta}^B_{-1,-1}(\theta)$, are equivalent and provide a satisfactory improvement for normal data, whereas in the remaining cases there is still room for further improvement, even when the sample size is 30. The original versions of the remaining statistics mimic the results of the standard empirical likelihood. Their adjusted Bartlett-corrected versions, instead, uniformly outperform both $\tilde{\Delta}^B_{-1,-1}(\theta)$ and $\bar{\Delta}^B_{-1,-1}(\theta)$ and provide empirical coverages quite close to the nominal levels.

For the multivariate mean we consider five multivariate distributions, each of them having the following components

$d_5 : X_1 \sim \mathrm{N}(0, D^2)$, $X_2 \sim \mathrm{Gamma}(D^{-1}, 1)$, $X_3 \sim \chi^2_D$;

$d_6 : X_1 \sim \mathrm{Gamma}(D, 1)$, $X_2 \sim \mathrm{Gamma}(D^{-1}, 1)$, $X_3 \sim \mathrm{Gamma}(4 - D, 1)$;

$d_7 : X_1 \sim 0.2\mathrm{N}(5, D^2) + 0.8\mathrm{N}(-1.25, D^{-2})$, $X_2 \sim 0.2\mathrm{N}(5, D^{-2}) + 0.8\mathrm{N}(-1.25, D^2)$, $X_3 \sim \mathrm{N}(0, D^2)$;

$d_8 : X_j \sim \mathrm{N}(0, 1)$, $j = 1, \ldots, 10$;

$d_9 : X_j = \exp\{Y_j\}$ with $Y_j \sim \mathrm{N}(0, 1)$, $j = 1, \ldots, 10$;

where D is a given value generated from $\mathrm{U}(1,2)$ (see, Liu and Chen, 2010).

We generate samples of size $\{30,50\}$ for distributions (d_5), (d_6), (d_7) and $\{40,70\}$ for (d_8), (d_9) in order to accommodate for the higher dimension of the mean. The resulting empirical coverages are reported in Table 4. On the one hand, when lower dimensional distributions are considered, we have quite the same picture as for the scalar mean, with the standard versions of the statistics providing almost all the same outcomes, regardless of the underlying distribution. We observe again that the adjusted and Bartlett-corrected empirical likelihoods $\tilde{\Delta}^B_{-1,-1}(\theta)$ and $\bar{\Delta}^B_{-1,-1}(\theta)$ lead to fairly good improvements that however are not as outstanding as the ones exhibited by $\bar{\Delta}^B_{-1,-2}(\theta)$, $\bar{\Delta}^B_{-1,-1/2}(\theta)$, $\bar{\Delta}^B_{-1,0}(\theta)$, $\bar{\Delta}^B_{0,0}(\theta)$, and $\bar{\Delta}^B_{1,1}(\theta)$. On the other hand, when the focus moves on the higher dimensional distributions we have the opportunity to appreciate the effects of the boundedness of $\tilde{\Delta}^B_{-1,-1}(\theta)$ and how our proposed $\bar{\Delta}^B_{-1,-1}(\theta)$ is able to circumvent the side effects. For $n=40$ the empirical coverages for $\tilde{\Delta}^B_{-1,-1}(\theta)$ degenerate to 100%, whereas the ones for $\bar{\Delta}^B_{-1,-1}(\theta)$ are still meaningful. The upper bound effect for $\tilde{\Delta}^B_{-1,-1}(\theta)$ almost vanishes at $n=70$ for distribution (d_8) whereas is it still persistent for (d_9).

In Figure 1 we display the Q-Q plots of the results for scalar and vector-valued mean when the underlying distribution is the exponential with unit mean (d_2) and the multivariate normal (d_8). In both scenarios we picked the smallest sample sizes, i.e. 15 and 40, respectively. These plots give us a flavour of how the adjustment and the Bartlett correction act on $\bar{\Delta}^B_{-1,-2}(\theta)$, $\bar{\Delta}^B_{-1,-1/2}(\theta)$, $\bar{\Delta}^B_{-1,0}(\theta)$, $\bar{\Delta}^B_{0,0}(\theta)$, and $\bar{\Delta}^B_{1,1}(\theta)$ by improving the raw approximation. The plots for empirical likelihood are markedly curved and we highlight that for the multivariate normal distribution the behaviour of $\tilde{\Delta}_{-1,-1}(\theta)$ is dictated by its upper bound. Simulations results not reported, reveal a similar behaviour for the adjusted Bartlett-corrected exponential empirical likelihood considered in Li *et al.* (2011).

4.2 Linear and generalised linear models

Let X_i be independent realisations from $\mathrm{DE}_1(\mu_i,\sigma^2 V(\mu_i))$ which is a dispersion exponential family of order 1 with mean μ_i and variance which depends on the dispersion parameter $\sigma^2>0$ and on the variance function $V(\mu_i)$ (Jørgensen, 1983). Here μ_i is supposed to be a function of a d-dimensional vector of explanatory variables Z_i through the link function $g(\cdot)$, i.e. $\mathbb{E}[X_i|Z_i=z_i]=\mu_i=g(\theta^a z_i^a)$ where θ is a d-dimensional regression parameter. By adopting the canonical link function, inference on θ is based on the following estimating function

$$\frac{1}{\sigma^2}\sum_{i=1}^n (x_i-\mu_i)z_i^r, \tag{20}$$

$r=1,\dots,d$.

We consider examples when the elements picked from $\mathrm{DE}_1(\cdot,\cdot)$ are the normal, $\mathrm{N}(\mu_i,\sigma^2)$, and exponential distribution, $\mathrm{Exp}(\mu_i^{-1})$. In each model the triad $\{g(\mu_i),\sigma^2,V(\mu_i)\}$ is $\{\mu_i,1,1\}$ and $\{-\mu_i^{-1},1,\mu_i^2\}$, respectively, i.e. σ^2 is known. The true parameter value is $\theta_0=(1/2,1,1,1/2,3)$ and the explanatory variables z_i^r corresponding to each θ_0^r, but the for intercept θ_0^1, are randomly generated from $\mathrm{U}(1,2)$, $\mathrm{N}(5,1)$, $\mathrm{Gamma}(1,2)$, and $\mathrm{Bi}(n,1/2)$, respectively. The considered sample sizes are $\{50,70\}$.

Simulation results are displayed in Table 5. Once again we have that the adjusted Bartlett-corrected $\bar{\Delta}^B_{-1,0}(\theta)$ and $\bar{\Delta}^B_{-1,-2}(\theta)$ are superior to $\bar{\Delta}^B_{-1,-1/2}(\theta)$, the original Bartlett corrected empirical likelihood $\Delta^B_{-1,-1}(\theta)$ as well as to its adjusted versions $\tilde{\Delta}^B_{-1,-1}(\theta)$ and $\bar{\Delta}^B_{-1,-1}(\theta)$. In

Table 3: Empirical coverage probabilities for scalar population mean. Monte Carlo standard errors for nominal levels $\nu = \{90, 95, 99\}$ are respectively 0.067, 0.048, and 0.022. Columns corresponding to the symbols "O", "B", "$\bar{B}$" and "$\tilde{B}$" report the values of the following statistics: original version, canonical Bartlett corrected version, adjusted and Bartlett corrected version, and $\tilde{\Delta}_{-1,-1}(\theta)$, respectively.

	n	ν	$\Delta_{-1,-1}$				$\Delta_{-1,-2}$		$\Delta_{-1,-1/2}$		$\Delta_{-1,0}$		$\Delta_{0,0}$		$\Delta_{1,1}$	
			O	B	$\tilde{B}$	$\bar{B}$	O	$\bar{B}$	O	$\bar{B}$	O	$\bar{B}$	O	$\bar{B}$	O	$\bar{B}$
d_1	15	90	87.0	88.6	88.8	89.1	86.4	89.5	86.8	88.8	86.1	89.6	86.6	89.7	86.9	89.2
		95	92.5	93.5	93.7	94.0	91.9	94.8	92.3	93.7	91.5	94.8	92.1	94.6	92.4	94.1
		99	97.5	97.9	98.0	98.0	97.1	99.1	97.3	98.0	96.7	99.1	97.2	98.4	97.5	98.2
	30	90	88.9	89.7	89.7	89.9	88.3	89.8	88.7	89.7	88.3	90.0	88.4	90.0	88.6	89.9
		95	94.0	94.6	94.7	94.8	93.4	94.8	93.9	94.7	93.4	94.9	93.7	95.0	93.7	94.9
		99	98.5	98.7	98.7	98.7	98.1	99.0	98.4	98.7	98.1	99.0	98.3	99.0	98.3	98.8
d_2	15	90	84.3	86.0	86.4	86.5	83.8	87.8	84.2	86.2	83.4	87.9	83.7	87.0	83.5	89.9
		95	89.7	91.0	91.3	91.3	89.3	93.3	89.6	91.2	88.9	93.4	89.1	92.5	88.5	95.0
		99	95.5	96.1	96.4	96.3	95.1	98.7	95.3	96.5	94.7	98.7	94.5	98.0	93.7	98.8
	30	90	87.3	88.4	88.5	88.5	86.9	88.8	87.2	88.4	86.8	88.7	86.6	88.6	86.2	89.9
		95	92.6	93.4	93.4	93.5	92.2	93.9	92.5	93.4	92.0	93.9	92.0	93.8	91.5	95.2
		99	97.8	98.1	98.1	98.1	97.4	98.7	97.7	98.1	97.3	98.7	97.0	98.6	96.3	99.3
d_3	15	90	81.4	83.5	84.1	83.9	81.3	86.2	81.7	83.8	80.7	86.2	81.2	85.4	80.7	92.0
		95	87.3	88.9	89.4	89.3	87.1	92.7	87.4	89.3	86.4	92.8	86.6	92.3	85.6	96.6
		99	93.8	94.4	94.8	94.6	93.7	98.4	93.8	95.1	93.1	98.4	92.4	98.4	90.9	99.5
	30	90	85.8	87.3	87.6	87.5	85.3	88.0	85.7	87.4	85.2	88.0	85.1	87.7	84.7	91.6
		95	91.4	92.5	92.7	92.6	91.0	93.6	91.3	92.6	90.8	93.5	90.5	93.4	89.6	96.9
		99	96.7	97.2	97.3	97.2	96.4	98.4	96.6	97.3	96.2	98.4	95.7	98.7	94.6	99.7
d_4	15	90	78.8	81.0	82.0	81.6	78.5	85.1	78.9	81.4	77.8	85.1	78.6	83.0	78.3	89.4
		95	84.9	86.6	87.3	87.0	84.7	91.7	85.1	87.1	83.9	91.6	84.0	89.7	82.8	94.8
		99	91.9	92.9	93.3	93.2	91.8	98.4	92.0	94.1	90.8	98.3	90.4	96.9	88.7	98.7
	30	90	82.7	84.5	85.1	84.6	82.1	85.5	82.5	84.5	81.9	85.4	82.3	85.0	82.2	89.3
		95	88.6	90.0	90.4	90.1	87.9	91.5	88.4	90.1	87.6	91.3	87.7	91.0	86.8	95.3
		99	95.3	95.9	96.0	95.9	94.9	97.8	95.1	96.1	94.5	97.7	93.8	97.6	92.3	99.4

the uppermost panels of Figure 2 we report the Q-Q plot of the results for the generalised linear model when $n = 50$. The inspection of this picture reveals an unusual downward curvature for $\tilde{\Delta}^B_{-1,-1}(\theta)$, not exhibited by $\bar{\Delta}^B_{-1,-1}(\theta)$, which can be likely ascribed to the upper bound effect.

4.3 Marginal pairwise likelihoods

Composite likelihoods are a likelihood-like class of functions that allow to carry out approximate inference for complex statistical models when the full likelihood is either impossible to specify or infeasible to compute. One drawback of this approach is that inference based on the composite log likelihood ratio is compromised since it obeys to the rules of misspecified likelihood ratios (Kent, 1982), i.e. its asymptotic null distribution is neither chi-squared nor asymptotically pivotal. Some strategies are available to modify the composite log likelihood ratio in order

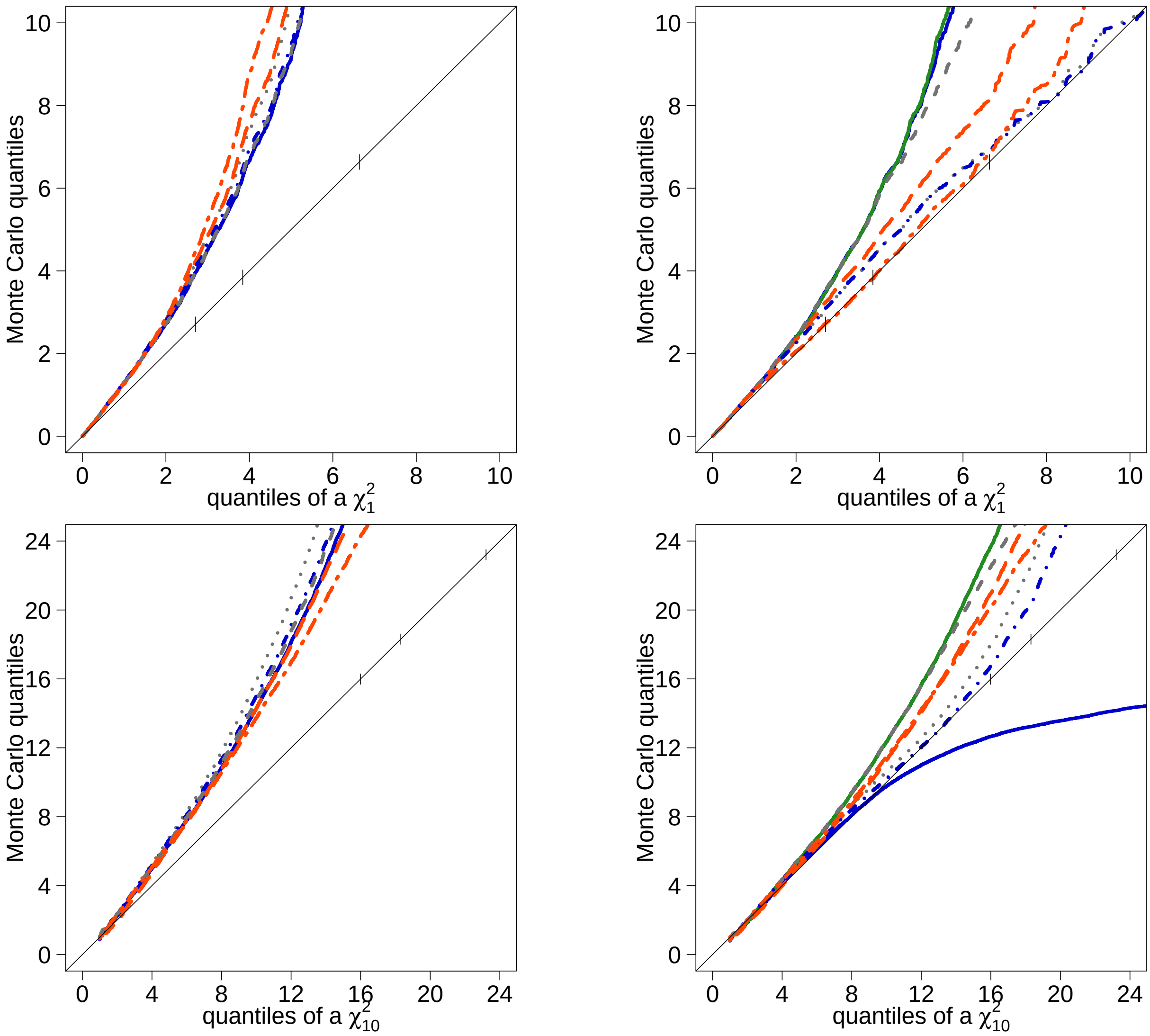

Figure 1: Q-Q plots for statistics when the underlying distribution is d_2 (uppermost panels) and d_8 (bottommost panels) for samples of size $n = 15$ and $n = 40$, respectively. On the left hand-side the plots are for the original statistics whereas on the right-hand side for the corresponding Bartlett-corrected/adjusted Bartlett-corrected versions. In particular: $\Delta_{-1,-1}$ and $\tilde{\Delta}^B_{-1,-1}$ (blue solid line), $\bar{\Delta}^B_{-1,-1}$ (green solid line), $\Delta_{-1,-2}$ and $\bar{\Delta}^B_{-1,-2}$ (blue dash-dotted line), $\Delta_{-1,-1/2}$ and $\bar{\Delta}^B_{-1,-1/2}$ (gray dashed line), $\Delta_{-1,0}$ and $\bar{\Delta}^B_{-1,0}$ (gray dotted line), $\Delta_{0,0}$ and $\bar{\Delta}^B_{0,0}$ (orange long-dashed line), $\Delta_{1,1}$ and $\bar{\Delta}^B_{1,1}$ (orange long/short-dashed line). Tick marks on the bisector indicate the 90%, 95%, and 99% quantiles of the reference distribution.

to recover the usual chi-squared reference (Chandler and Bate, 2007; Pace *et al.*, 2011). An alternative proposal is to use the empirical log likelihood ratio based on the composite score function (Lunardon *et al.*, 2013). Although this approach is appealing and theoretically sound its potential is dimmed as it inherits the well known and persistent undercoverage problem of the empirical likelihood-based confidence sets. In the sequel, we focus on a particular sub-

Table 4: Empirical coverage probabilities for vector-valued population mean. Monte Carlo standard errors for nominal levels $\nu = \{90, 95, 99\}$ are respectively 0.067, 0.048, and 0.022. Columns corresponding to the symbols "O", "B", "$\bar{B}$" and "$\tilde{B}$" report the values of the following statistics: original version, canonical Bartlett corrected version, adjusted and Bartlett corrected version, and $\tilde{\Delta}_{-1,-1}(\theta)$, respectively.

	n	ν	$\Delta_{-1,-1}$				$\Delta_{-1,-2}$		$\Delta_{-1,-1/2}$		$\Delta_{-1,0}$		$\Delta_{0,0}$		$\Delta_{1,1}$	
			O	B	$\tilde{B}$	$\bar{B}$	O	$\bar{B}$	O	$\bar{B}$	O	$\bar{B}$	O	$\bar{B}$	O	$\bar{B}$
d_5	30	90	81.2	84.8	86.6	85.2	79.7	89.8	80.7	85.6	79.0	89.3	79.6	88.0	78.9	89.8
		95	88.0	90.4	91.9	90.6	86.7	95.2	87.5	91.0	85.9	94.7	86.2	93.2	85.2	94.3
		99	94.7	96.0	97.1	96.2	94.1	99.1	94.4	96.6	93.3	98.9	93.5	97.9	92.5	98.3
	50	90	85.0	87.2	87.7	87.3	83.6	88.6	84.6	87.4	83.3	88.4	83.3	88.2	82.7	89.9
		95	91.1	92.7	93.2	92.8	90.1	94.3	90.9	92.9	89.7	94.2	89.6	93.8	88.7	95.2
		99	97.1	97.8	98.0	97.9	96.6	99.1	97.0	98.0	96.3	98.9	96.1	98.6	95.3	99.0
d_6	30	90	81.6	84.9	86.4	85.2	80.2	89.4	81.0	85.5	79.5	88.8	80.0	87.6	79.5	88.9
		95	87.7	90.2	91.6	90.4	86.6	95.1	87.3	90.7	85.8	94.6	86.1	92.8	85.3	93.7
		99	94.6	95.9	97.0	96.1	93.9	99.3	94.2	96.5	93.1	99.1	93.4	97.9	92.5	98.1
	50	90	84.8	87.2	87.7	87.4	83.5	88.6	84.4	87.5	83.2	88.4	83.3	88.3	82.5	89.9
		95	91.1	92.5	92.9	92.6	90.0	94.2	90.7	92.8	89.6	94.1	89.6	93.8	88.6	94.9
		99	97.1	97.7	97.9	97.8	96.5	99.1	96.9	97.9	96.2	99.0	96.0	98.6	95.1	98.9
d_7	30	90	84.8	87.5	88.5	87.9	83.6	90.8	84.8	88.4	82.9	90.4	83.3	90.2	82.6	90.5
		95	90.9	92.7	93.6	92.9	90.1	96.0	90.9	93.4	89.4	95.6	89.5	95.0	88.8	95.0
		99	96.9	97.6	98.3	97.7	96.5	99.4	96.8	98.1	96.0	99.2	96.0	98.7	95.2	98.5
	50	90	87.6	89.0	89.1	89.1	86.4	89.4	87.3	89.1	86.2	89.4	86.1	89.8	85.7	90.5
		95	93.2	94.0	94.1	94.1	92.1	94.7	92.9	94.1	91.9	94.6	91.9	94.9	91.4	95.3
		99	98.2	98.6	98.6	98.6	97.8	99.1	98.1	98.6	97.6	99.0	97.5	99.0	97.1	99.0
d_8	40	90	62.9	71.4	100.0	72.0	59.7	88.2	60.8	72.3	55.7	85.6	63.2	78.1	66.5	78.4
		95	71.2	79.1	100.0	79.5	68.3	92.3	69.3	80.3	64.0	90.4	71.9	84.9	75.6	85.9
		99	83.7	88.8	100.0	89.2	81.2	96.4	81.7	91.0	76.7	95.2	84.6	92.6	88.1	94.2
	70	90	80.0	84.7	88.8	84.9	76.7	90.8	78.4	84.6	74.5	89.2	78.7	87.8	78.4	86.3
		95	87.5	90.8	94.5	90.9	84.7	95.7	86.2	90.8	82.6	94.7	85.8	93.0	86.3	92.1
		99	95.3	96.8	99.4	96.9	93.9	98.8	94.6	96.9	92.3	98.3	94.3	97.8	95.0	97.6
d_9	40	90	41.2	54.2	100.0	54.6	39.8	80.0	40.6	56.7	35.0	76.2	42.3	64.0	39.4	66.0
		95	48.8	61.2	100.0	61.7	47.6	84.8	48.1	65.5	42.3	81.6	49.4	70.9	46.9	72.9
		99	61.5	72.2	100.0	72.5	60.3	90.5	60.7	80.0	54.1	87.9	61.8	81.1	60.1	83.1
	70	90	63.0	73.7	96.6	74.0	58.1	86.3	60.1	73.1	55.8	84.2	64.6	81.1	55.9	80.7
		95	72.0	80.8	99.2	81.0	67.7	92.3	69.4	80.7	64.8	90.6	71.7	87.1	64.2	86.8
		99	84.2	89.9	100.0	90.0	81.2	96.8	82.3	90.7	78.4	96.1	82.6	93.9	76.9	94.0

class of composite likelihoods, namely the marginal pairwise likelihoods (Cox and Reid, 2004; Varin *et al.*, 2011), and investigate whether the Bartlett calibration and in particular $\Delta^B_{-1,-1}(\theta)$, $\tilde{\Delta}^B_{-1,-1}(\theta)$, $\bar{\Delta}^B_{-1,-1}(\theta)$, $\bar{\Delta}^B_{-1,0}(\theta)$, $\bar{\Delta}^B_{-1,-2}(\theta)$, and $\bar{\Delta}^B_{-1,-1/2}(\theta)$ may prove useful in this context.

We will define our simulation models by imposing suitable restrictions to the mean vector ξ and the covariance matrix Σ of a normally distributed q-dimensional random vector X.

The first model we consider (Model 1) has mean vector $\xi = (\mu, \ldots, \mu) \in \mathbb{R}^q$ and compound

Table 5: Linear and generalised linear model; empirical coverage probabilities for $\theta \in \mathbb{R}^5$. Monte Carlo standard errors for nominal levels 90, 95, and 99 are respectively 0.067, 0.048, and 0.022. Columns corresponding to the symbols "O", "B", "$\bar{B}$" and "$\tilde{B}$" report the values of the following statistics: original version, canonical Bartlett corrected version, adjusted and Bartlett corrected version, and $\tilde{\Delta}_{-1,-1}(\theta)$, respectively.

	n	ν	$\Delta_{-1,-1}$				$\Delta_{-1,-2}$		$\Delta_{-1,-1/2}$		$\Delta_{-1,0}$	
			O	B	$\tilde{B}$	$\bar{B}$	O	$\bar{B}$	O	$\bar{B}$	O	$\bar{B}$
LM	50	90	74.2	79.2	84.4	79.4	70.7	85.7	72.7	79.4	68.2	83.2
		95	81.9	86.0	91.9	86.2	78.7	93.2	80.5	86.3	76.0	90.9
		99	91.8	94.2	99.1	94.3	89.6	99.1	90.9	94.6	86.6	98.3
	70	90	79.4	83.4	85.8	83.5	76.4	87.3	78.3	83.5	74.7	85.7
		95	87.0	90.1	92.2	90.2	84.1	94.2	86.0	90.2	82.2	93.1
		99	95.2	96.6	98.2	96.7	93.6	99.3	94.6	96.8	91.9	98.8
GLM	50	90	65.3	71.9	81.5	72.2	73.1	91.2	64.3	80.2	59.1	88.4
		95	74.0	79.9	90.6	80.1	80.3	95.4	73.0	87.9	67.4	93.9
		99	86.3	90.1	98.6	90.3	89.8	98.9	85.3	96.3	79.7	98.2
	70	90	72.1	77.4	82.2	77.6	78.2	90.8	71.4	81.5	67.1	88.1
		95	80.2	84.5	89.7	84.7	84.6	95.4	79.5	88.6	75.1	94.1
		99	90.6	93.4	97.6	93.5	92.9	99.1	90.0	96.5	86.1	98.5

covariance matrix whose diagonal and off-diagonal elements are $\Sigma^{rr} = \sigma^2$ and $\Sigma^{rs} = \sigma^2\rho$, respectively, with $\sigma^2 > 0$ and $\rho \in (-(q-1)^{-1}, 1)$. This model has been widely studied in the composite likelihood framework (see, e.g., Pace *et al.*, 2011; Lunardon *et al.*, 2013) and inference focuses on the three-dimensional parameter $\theta = (\mu, \sigma^2, \rho)$. The pairwise log likelihood for θ is defined by summing all the log likelihood contributions arising from the $q(q-1)/2$ distinct pairs (X^r, X^s). For each pair it is specified a bivariate normal distribution with mean vector $(\mu, \mu)^\top$ and covariance matrix whose diagonal and off-diagonal elements are σ^2 and $\sigma^2\rho$, respectively. Under these assumptions the pairwise log likelihood for θ is

$$p\ell(\theta) = -\frac{nq(q-1)}{2}\left[\log\sigma^2 + \frac{\log(1-\rho^2)}{2}\right] - \frac{1}{2\sigma^2(1-\rho^2)}\sum_{i=1}^n \Lambda^{ab}(x_i-\mu)^a(x_i-\mu)^b,$$

$\Lambda^{rr} = (q-1)$ and $\Lambda^{rs} = -\rho$. The associated pairwise score function $pU(\theta) = \partial p\ell(\theta)/\partial\theta$ has the following components

$$\begin{aligned}
pU^1(x;\theta) &= -\frac{(q-1)\rho+\sigma^2}{\sigma^2(1-\rho^2)}\sum_{i=1}^n\sum_{j=1}^q (x_{ij}-\mu),\\
pU^2(x;\theta) &= -\frac{nq(q-1)}{2\sigma^2} + \frac{1}{2(1-\rho^2)(\sigma^2)^2}\sum_{i=1}^n \Lambda^{ab}(x_i-\mu)^a(x_i-\mu)^b,\\
pU^3(x;\theta) &= \frac{nq(q-1)\rho}{2(1-\rho^2)} - \frac{1}{2\sigma^2(1-\rho^2)^2}\sum_{i=1}^n \Upsilon^{ab}(x_i-\mu)^a(x_i-\mu)^b,
\end{aligned}$$

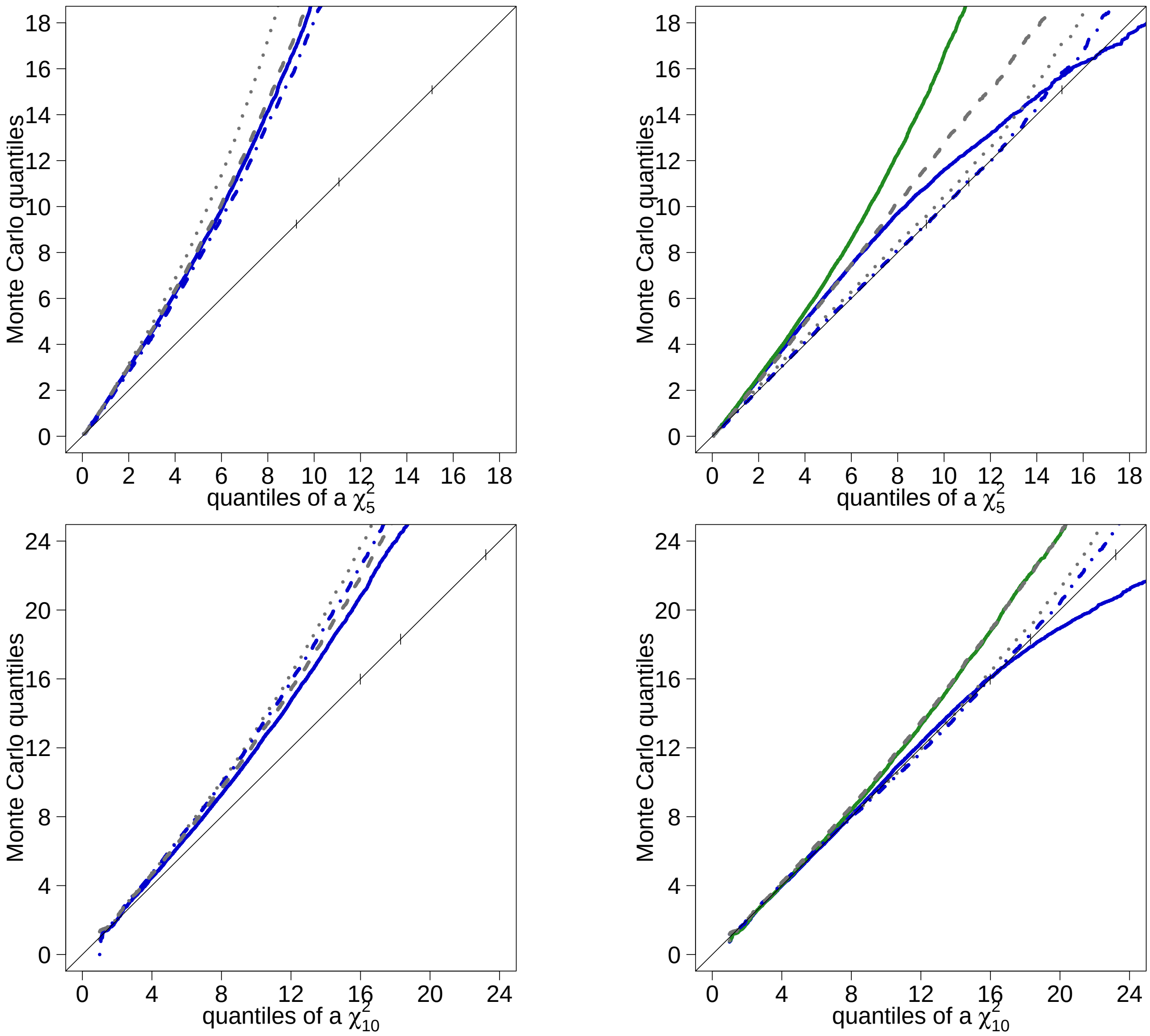

Figure 2: Q-Q plots for statistics for the generalized linear model (uppermost panels) and Model 2 (bottommost panels) for samples of size $n = 50$ and $n = 85$, respectively. On the left hand-side the plots are for the original statistics whereas on the right-hand side for the corresponding Bartlett-corrected/adjusted Bartlett-corrected versions. In particular: $\Delta_{-1,-1}$ and $\tilde{\Delta}^B_{-1,-1}$ (blue solid line), $\bar{\Delta}^B_{-1,-1}$ (green solid line), $\Delta_{-1,-2}$ and $\bar{\Delta}^B_{-1,-2}$ (blue dash-dotted line), $\Delta_{-1,-1/2}$ and $\bar{\Delta}^B_{-1,-1/2}$ (gray dashed line), $\Delta_{-1,0}$ and $\bar{\Delta}^B_{-1,0}$ (gray dotted line). Tick marks on the bisector indicate the 90%, 95%, and 99% quantiles of the reference distribution.

where the diagonal and off-diagonal elements of the matrix Υ are $\Upsilon^{rr} = 2\rho(q-1)$ and $\Upsilon^{rs} = -(1+\rho^2)$, respectively.

In the second model (Model 2), the mean vector ξ and the elements Σ^{rr} are treated as known and, without loss of generality, they are set equal to 0 and to 1, respectively. The parameter θ has dimension $q-1$ and its components are the correlations between pairs of X. We have only $q-1$ correlations since we work under the restriction $\theta^{|r-s|} = \text{cor}(X^r X^s) = \Sigma^{rs}$, $r \neq s$, i.e. all components at the same distance are equally correlated. By specifying for each pair (X^r, X^s) a standard bivariate normal distribution with correlation coefficient $\theta^{|r-s|}$ and by considering all the possible pairs, we have that the resulting pairwise score function is

$$pU^r(x;\theta) = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{q-r}\frac{1}{1-(\theta^r)^2}\left\{\theta^r - \frac{1}{1-(\theta^r)^2}\left[(x_i^j)^2 + (x_i^{j+r})^2 - 2(\theta^r)(x_i^j)(x_i^{j+r})\right]\right\},$$

$r = 1, \ldots, q-1$. To avoid any ambiguity we make a slight abuse of notation and indicate by $(\cdot)^s$ the s-th power of the quantity enclosed in brackets, e.g. $(\theta^r)^2$ is the square of the r-th component of θ and $(\theta^r)(x_i^j)(x_i^{j+r})$ is the product among the r-th component of θ and the j-th and $j+r$-th components of x, respectively. Summation convention does not apply for the indices r, j.

The definition of functions in classes $\{\Delta_{\gamma,\phi}(\theta)\}$ and $\{\bar{\Delta}_{\gamma,\phi}(\theta)\}$ is based on $h(x;\theta) = pU(x;\theta)$. Note that pairwise score functions satisfy (1) (see, e.g., Molenberghs and Verbeke, 2005).

The simulation settings are as follows. For Model 1, we consider samples of size $\{30, 50\}$, dimension $q = 50$, and assign to the true parameter $\theta_0 = (\mu_0, \sigma_0^2, \rho_0)$ the values $\mu_0 = 0$, $\sigma_0^2 = 1$, and $\rho_0 = \{0.2, 0.5, 0.9\}$. For Model 2, we explore high-dimensional settings, which in turn increase the dimension of θ, by setting $q = \{9, 11\}$. The sample sizes corresponding to each q are $\{70, 90\}$ and $\{85, 100\}$, respectively. The correlations in θ are considered stored in decreasing order. The true parameter value is set by fixing the highest value of correlation θ_0^1 and by subsequently determining the remaining components according to the relation $\theta_0^j = (\theta_0^1)^j$, i.e. the correlation between pairs decays exponentially with the distance, $j = 2, \ldots, q-1$. In all settings $\theta_0^1 = 0.9$.

For the two models, it is easy to check the validity of assumptions **A1** and **A3**, however this easiness does not apply to such an extent for **A2**. We avoid exceedingly lengthy calculations by resorting to the moment generating function of the random vector $h(X;\theta)$ in each model. These functions are available in closed form for both models and allow to verify that assumption **A2** holds for the considered θ_0.

Simulation results for Model 1 and Model 2 are reported in Table 6 and Table 7, respectively. On the one hand, empirical coverages for the empirical likelihood, $\Delta_{-1,-1}$, and exponential empirical likelihood, $\Delta_{-1,0}$, $\Delta_{-1,-2}$, and $\Delta_{-1,-1/2}$ are below the nominal levels regardless the sample size. In particular, if in Model 1 we almost double the sample size, we are still not able to get close to the nominal levels. On the other hand, the Bartlett and adjusted Bartlett corrected empirical likelihoods $\Delta^B_{-1,-1}(\theta)$ and $\bar{\Delta}^B_{-1,-1}$ provide little improvements and we note that at higher sample sizes their outcomes tend to be equivalent. Overall, the adjusted Bartlett corrected empirical and exponential empirical likelihood $\tilde{\Delta}^B_{-1,-1}$ and $\bar{\Delta}^B_{-1,0}$ as well as $\bar{\Delta}^B_{-1,-2}$ outperform $\Delta^B_{-1,-1}$, $\bar{\Delta}^B_{-1,-1}$, and $\bar{\Delta}^B_{-1,-1/2}$, and provide empirical coverages close to the nominal levels.

In the bottommost panels of Figure 2 we display the Q-Q plots of the results for Model 2 when $n = 85$ and $q = 11$. The adjusted Bartlett-corrected exponential empirical likelihood $\bar{\Delta}^B_{-1,0}$ and $\bar{\Delta}^B_{-1,-2}$ provide an excellent agreement with the χ^2_{10} reference. The downward curvature of $\tilde{\Delta}^B_{-1,-1}(\theta)$ indicates a unusual behaviour, still presumably due to the upper bound effect.

Table 6: Model 1, empirical coverage probabilities for $\theta = (\mu, \sigma^2, \rho)$. Monte Carlo standard errors for nominal levels $\nu = \{90, 95, 99\}$ are respectively 0.067, 0.048, and 0.022. Columns corresponding to the symbols "O", "B", "$\bar{B}$" and "$\tilde{B}$" report the values of the following statistics: original version, canonical Bartlett corrected version, adjusted and Bartlett corrected version, and $\tilde{\Delta}_{-1,-1}(\theta)$, respectively.

n	ρ	ν	$\Delta_{-1,-1}$				$\Delta_{-1,-2}$		$\Delta_{-1,-1/2}$		$\Delta_{-1,0}$	
			O	B	$\tilde{B}$	$\bar{B}$	O	$\bar{B}$	O	$\bar{B}$	O	$\bar{B}$
30	0.2	90	80.2	83.7	80.0	84.0	82.6	90.4	79.9	84.5	78.5	88.5
		95	86.3	88.9	86.2	89.1	88.2	95.0	86.1	89.6	84.8	93.6
		99	93.5	95.0	93.4	95.1	94.5	98.8	93.2	95.5	92.3	98.3
	0.5	90	80.4	83.8	80.2	84.1	83.5	90.3	80.1	85.0	78.6	88.0
		95	86.4	89.2	86.2	89.5	88.7	94.9	86.1	90.1	84.8	92.5
		99	93.6	94.9	93.5	95.0	94.8	98.1	93.2	95.3	92.4	97.1
	0.9	90	80.5	84.1	80.3	84.4	86.6	90.2	80.2	85.4	78.9	88.3
		95	86.7	89.3	86.5	89.5	91.0	94.9	86.3	90.2	85.0	92.4
		99	93.1	94.6	93.1	94.8	95.5	98.7	92.9	95.2	92.0	96.5
50	0.2	90	85.0	87.2	84.9	87.3	85.9	90.2	84.7	87.5	83.7	88.5
		95	90.6	92.2	90.6	92.4	91.3	94.8	90.5	92.6	89.6	93.8
		99	96.8	97.5	96.8	97.5	96.9	98.8	96.6	97.6	96.0	98.4
	0.5	90	84.9	87.3	84.9	87.5	86.7	90.2	84.8	87.8	83.7	89.0
		95	90.9	92.5	90.8	92.6	92.1	95.4	90.8	92.9	89.8	94.0
		99	96.6	97.3	96.6	97.3	97.2	98.7	96.4	97.5	95.9	98.0
	0.9	90	85.0	87.1	84.9	87.2	89.5	90.5	84.7	87.6	83.6	89.3
		95	90.8	92.4	90.8	92.5	93.9	95.4	90.6	92.9	89.7	94.1
		99	96.7	97.4	96.7	97.5	98.0	98.9	96.5	97.6	96.0	97.9

5 Final remarks

The methodology, originally introduced by Liu and Chen (2010), based on pseudo-observations, has been shown to be effective in solving the convex hull problem and in getting the Bartlett correction for the empirical likelihood and the exponential empirical likelihood (Li *et al.*, 2011). In this paper we have extended the method to the entire generalised power divergence family, identifying an appropriate adjustment for every member of the family. The proposed refinement gives back attractiveness to a broad class of statistics that potentially contains good alternatives to the empirical likelihood. Moreover, unlike the original proposal, our formulation allows, in practice, to overcome the problem of the upper bound highlighted by Emerson and Owen (2009).

Simulation results seem to confirm the effectiveness of our approach. For the considered statistics, the effect of the adjustment (and the Bartlett correction) in reducing the coverage error of confidence regions is not homogeneous and the better results are obtained for items other than empirical likelihood. This in turn gives importance to members of the family other than empirical likelihood. Further, our solution to the "upper bound problem" via the tuning parameter δ seems to be effective and reasonable because, even for $\delta = 1/2$, in all the considered cases, $\bar{\Delta}^B_{-1,-1}(\theta)$ resembles the results of the original Bartlett-corrected empirical likelihood $\Delta^B_{-1,-1}(\theta)$. This lead

Table 7: Model 2, empirical coverage probabilities for vector-valued θ. Monte Carlo standard errors for nominal levels $\nu = \{90, 95, 99\}$ are respectively 0.067, 0.048, and 0.022. Columns corresponding to the symbols "O", "B", "$\bar{B}$" and "$\tilde{B}$" report the values of the following statistics: original version, canonical Bartlett corrected version, adjusted and Bartlett corrected version, and $\tilde{\Delta}_{-1,-1}(\theta)$, respectively.

q	n	ν	$\Delta_{-1,-1}$				$\Delta_{-1,-2}$		$\Delta_{-1,-1/2}$		$\Delta_{-1,0}$	
			O	B	$\tilde{B}$	$\bar{B}$	O	$\bar{B}$	O	$\bar{B}$	O	$\bar{B}$
9	70	90	76.7	82.1	87.2	82.2	73.2	89.5	74.9	82.9	71.4	88.5
		95	84.4	88.4	93.6	88.5	81.4	93.7	82.7	89.2	79.7	92.9
		99	93.3	95.2	99.1	95.3	91.4	98.0	92.2	95.7	89.9	97.5
	90	90	80.6	84.6	86.9	84.7	77.3	90.0	79.2	85.0	76.3	89.5
		95	87.5	90.5	92.7	90.6	84.9	94.4	86.2	91.0	83.9	93.9
		99	95.2	96.7	98.4	96.8	93.8	98.4	94.5	97.0	92.9	98.1
11	85	90	76.2	82.2	89.5	82.4	71.2	89.8	73.2	82.0	69.4	88.8
		95	84.2	88.7	95.7	88.9	79.9	94.4	81.7	88.7	78.0	93.5
		99	93.4	95.9	99.8	96.0	90.9	98.4	91.9	95.9	89.2	97.9
	100	90	79.3	84.5	88.6	84.5	74.6	90.2	76.7	84.0	73.3	89.4
		95	87.1	90.6	94.5	90.6	83.2	94.9	85.1	90.5	81.7	94.2
		99	95.3	96.8	99.2	96.9	93.1	98.7	94.1	96.9	91.9	98.4

us to conclude that the intervention of pseudo-observation $\Xi(\theta)$ on $\bar{\Delta}^B_{-1,-1}(\theta)$, as well as on any other member in $\left\{\bar{\Delta}^B_{\gamma,\phi}(\theta)\right\}$, is negligible when the convex hull condition is satisfied and, at the same time, is effective when the latter need to be accounted for.

Appendix

Basic expansions

As outlined in Section 2.1, recall that the covariance matrix of $h(X;\theta_0)$ is assumed to satisfy $\alpha^{rs} = \delta^{rs}$.

The Lagrange multipliers $\lambda(\theta) \in \mathbb{R}$ and $\beta(\theta) \in \mathbb{R}^d$ appearing in (3) solve the equations $\sum_{i=1}^n w_i(\theta) = 1$ and $\sum_{i=1}^n w_i(\theta)h(x_i;\theta) = 0$. For our developments it is convenient to set $z(\lambda(\theta);\beta(\theta)) = \sum_{i=1}^n w_i(\theta) - 1 = 0$ and $f(\lambda(\theta);\beta(\theta)) = \sum_{i=1}^n w_i(\theta)h(x_i;\theta) = 0$. If we evaluate these expressions at $\theta = \theta_0$, then $w_i = w_i(\theta_0)$, $\lambda = \lambda(\theta_0)$, and $\beta = \beta(\theta_0)$. Note that expression for λ and β are developed for the case $\phi \neq 0$, only, and analogue calculations can be carried out to obtain the corresponding expressions for $\phi = 0$. Nevertheless, the expansion for $\Delta_{\gamma,\phi}$ given in (21) is valid for each pair (λ, ϕ).

After plugging-in the expression of the $w_i's$ in functions $z(\cdot;\cdot)$ and $f(\cdot;\cdot)$ we obtain the following McLaurin series expansions

$$\begin{aligned}
z(\lambda;\beta) &= n^{-1}\sum_{i=1}^{n}[1+\lambda+\beta^a h^a(x_i;\theta_0)]^{1/\phi}-1=0\\
&= \frac{\lambda}{\phi}+\frac{1}{\phi}\beta^a A^a+\frac{1}{2\phi^2}(1-\phi)\left[\lambda\lambda+2\lambda\beta^a A^a+\beta^a\beta^b(A^{ab}+\delta^{ab})\right]+\\
&+ \frac{1}{6\phi^3}(1-\phi)(1-2\phi)\left[3\lambda\beta^a\beta^b\delta^{ab}+\beta^a\beta^b\beta^c(A^{abc}+\alpha^{abc})\right]+\\
&+ \frac{1}{24\phi^4}(1-\phi)(1-2\phi)(1-3\phi)\alpha^{abcd}\beta^a\beta^b\beta^c\beta^d+O_p(n^{-5/2})=0,
\end{aligned}$$

and

$$\begin{aligned}
f^r(\lambda;\beta) &= n^{-1}\sum_{i=1}^{n}[1+\lambda+\beta^a h^a(x_i;\theta_0)]^{1/\phi}h^r(x_i;\theta_0)=0\\
&= A^r+\frac{1}{\phi}\lambda A^r+\frac{1}{\phi}\beta^a\left(A^{ra}+\delta^{ra}\right)+\frac{1}{2\phi^2}(1-\phi)\left[2\lambda\beta^a(A^{ra}+\delta^{ra})+\beta^a\beta^b(A^{rab}+\alpha^{rab})\right]+\\
&+ \frac{1}{6\phi^3}(1-\phi)(1-2\phi)\alpha^{rabc}\beta^a\beta^b\beta^c+O_p(n^{-2})=0.
\end{aligned}$$

Inversion of these series leads to the following expression for λ (in the first line $O_p(n^{-1})$ terms, in the second line $O_p(n^{-3/2})$ terms, and in the subsequent lines $O_p(n^{-2})$ terms)

$$\begin{aligned}
\lambda &= \frac{1}{2}\phi(1+\phi)A^aA^a+\\
&- \frac{1}{2}\phi(1+\phi)A^{ab}A^aA^b+\frac{1}{6}\phi(1-\phi^2)\alpha^{abc}A^aA^bA^c+\\
&+ \frac{1}{2}\phi(1+\phi)A^{ab}A^{bc}A^aA^c+\frac{1}{8}\phi(1+\phi)^3A^aA^aA^bA^b+\frac{1}{6}\phi(1-\phi^2)A^{abc}A^aA^bA^c+\\
&- -\frac{1}{2}\phi(1-\phi^2)\alpha^{abc}A^{cd}A^aA^bA^d+\frac{1}{8}\phi(1+\phi)(1-\phi)^2\alpha^{abc}\alpha^{cde}A^aA^bA^dA^e+\\
&+ \frac{1}{24}(1-\phi^2)(1-2\phi)\alpha^{abcd}A^aA^bA^cA^d+O_p(n^{-5/2}),
\end{aligned}$$

and for β (in the first line $O_p(n^{-1/2})$ terms, in the second line $O_p(n^{-1})$ terms, and in the subsequent lines $O_p(n^{-3/2})$ terms)

$$\begin{aligned}
\beta^r &= -\phi A^r+\\
&+ \phi A^{ra}A^a-\frac{1}{2}\phi(1-\phi)\alpha^{rab}A^aA^b+\\
&- \phi A^{ra}A^{ab}A^b-\frac{1}{2}\phi^2(1+\phi)A^rA^aA^a+\frac{1}{2}\phi(1-\phi)\alpha^{abc}A^{ra}A^bA^c+\phi(1-\phi)\alpha^{rab}A^{bc}A^aA^c+\\
&- \frac{1}{2}\phi(1-\phi)A^{rab}A^aA^b-\frac{1}{2}\phi(1-\phi)^2\alpha^{rab}\alpha^{bcd}A^aA^cA^d+\frac{1}{6}\phi(1-\phi)(1-2\phi)\alpha^{rabc}A^aA^bA^c+\\
&+ O_p(n^{-2}).
\end{aligned}$$

Once the $w_i's$ are substituted in the expression of $\Delta_{\gamma\phi}(\theta_0)$, a McLaurin series expansion gives

$$\begin{aligned}
n^{-1}\Delta_{\gamma,\phi} &= \frac{2}{n\gamma(\gamma+1)}\sum_{i=1}^{n}\left\{[1+\lambda+\beta^a h^a(x_i;\theta_0)]^{(\gamma+1)/\phi}-1\right\} = \\
&= \frac{2}{\gamma}\left\{\frac{\lambda}{\phi}+\frac{1}{\phi}\beta^a A^a+\frac{1}{2\phi^2}(\gamma+1-\phi)\left[\lambda\lambda+2\lambda\beta^a A^a+\beta^a\beta^b(A^{ab}+\delta^{ab})\right]+\right.\\
&+ \frac{1}{6\phi^3}(\gamma+1-\phi)(\gamma+1-2\phi)\left[\beta^a\beta^b\beta^c(A^{abc}+\alpha^{abc})+3\lambda\beta^a\beta^b\delta^{ab}\right]+\\
&+ \left.\frac{1}{24\phi^4}(\gamma+1-\phi)(\gamma+1-2\phi)(\gamma+1-3\phi)\alpha^{abcd}\beta^a\beta^b\beta^c\beta^d\right\}+O_p(n^{-5/2}).
\end{aligned}$$

Finally, plugging in the above expansion the expressions of λ and β leads to (in the first line $O_p(n^{-1})$ terms, in the second line $O_p(n^{-3/2})$ terms, and in the subsequent lines $O_p(n^{-2})$ terms)

$$\begin{aligned}
n^{-1}\Delta_{\gamma,\phi} &= A^aA^a+ \qquad (21)\\
&- A^{ab}A^aA^b+\frac{1}{3}(1-\gamma)\alpha^{abc}A^aA^bA^c+\\
&+ A^{ab}A^{bc}A^aA^c+\frac{1}{4}(1+\phi)(1-\phi+2\gamma)A^aA^aA^bA^b+\frac{1}{3}(1-\gamma)A^{abc}A^aA^bA^c+\\
&- (1-\gamma)\alpha^{abc}A^{cd}A^aA^bA^d+\frac{1}{4}(1-\phi)(1+\phi-2\gamma)\alpha^{abc}\alpha^{cde}A^aA^bA^dA^e+\\
&+ \frac{1}{12}(-1+3\gamma+\gamma^2-6\gamma\phi+3\phi^2)\alpha^{abcd}A^aA^bA^cA^d+O_p(n^{-5/2}).
\end{aligned}$$

The signed square root $n^{1/2}R_{\gamma,\phi}=n^{1/2}R_{\gamma,\phi}(\theta_0)$ presented in Section 2.2 is developed by matching the expansion of $n^{-1}\Delta_{\gamma,\phi}(\theta_0)$ order by order, keeping in mind both that the size of each term in $R^r_{\gamma,\phi}$ is $R^r_{j;\gamma,\phi}=O_p(n^{-j/2})$, $j=1,2,3$, and that

$$n^{-1}\Delta_{\gamma,\phi}(\theta_0)=R^a_{1;\gamma,\phi}R^a_{1;\gamma,\phi}+R^a_{2;\gamma,\phi}R^a_{2;\gamma,\phi}+2R^a_{1;\gamma,\phi}R^a_{2;\gamma,\phi}+2R^a_{1;\gamma,\phi}R^a_{3;\gamma,\phi}+O_p(n^{-5/2}).$$

More details about the derivation of the signed root are given in the proof of Lemma 1.

Proof of Lemma 1

When the function defined in (2) is computed on a pseudo-sample $h(x_1;\theta_0),\ldots,h(x_{n+3};\theta_0)$, with $h(x_{n+1};\theta_0)=\Omega=O_p(1)$, $h(x_{n+2};\theta_0)=\Gamma=O_p(n^{-1/2})$, and $h(x_{n+3};\theta_0)=\Xi=O_p(n^{-1/2-\delta})$, $\delta\geq 1/2$, we need to develop new expressions for the Lagrange multipliers λ and β in order to accommodate for the pseudo-observations. Recall that quantities computed on an augmented pseudo-sample are denoted with an upper bar and that we provide computations for $\phi\neq 0$. Let $\bar{z}(\bar{\lambda};\bar{\beta})$ and $\bar{f}^r(\bar{\lambda};\bar{\beta})$ be the analogues of $z(\lambda;\beta)$ and $f^r(\lambda;\beta)$ computed on $h(x_1;\theta_0),\ldots,h(x_n;\theta_0),\Omega,\Gamma,\Xi$, then

$$\begin{aligned}
\bar{\lambda} &= \frac{n}{n+3}\left[\lambda+\frac{1}{n}\phi(1+\phi)A^a\Omega^a+\frac{1}{n}\phi(1+\phi)A^a\Gamma^a-\frac{1}{n}\phi(1+\phi)A^{ab}A^a\Omega^b+\right.\\
&- \left.\frac{1}{2n}\phi(1+\phi)A^a\Omega^aA^b\Omega^b+\frac{1}{2n}\phi(1-\phi^2)\alpha^{abc}A^aA^b\Omega^c+\frac{1}{2n^2}\phi(1+\phi)\Omega^a\Omega^a\right]+O_p(n^{-5/2}),
\end{aligned}$$

and

$$\bar{\beta}^r = \beta^r - \frac{1}{n}\phi\left[\Omega^r + \Gamma^r + A^{ra}\Omega^a + \Omega^r A^a \Omega^a - (1-\phi)\alpha^{rab} A^a \Omega^b\right] + O_p(n^{-2}).$$

These expressions do not contain the pseudo-observation Ξ as a straightforward check of $\bar{z}(\bar{\lambda};\bar{\beta})$ and $\bar{f}^r(\bar{\lambda};\bar{\beta})$ highlights that the size of the terms involving Ξ are far beyond $O_p(n^{-5/2})$ and $O_p(n^{-2})$, i.e. the size of the remainder terms considered in the expansions of $\bar{z}(\cdot;\cdot)$ and $\bar{f}(\cdot;\cdot)$, respectively. The expansion of $n^{-1}\bar{\Delta}_{\gamma,\phi}$ for each pair (γ,ϕ) is then

$$\begin{aligned} n^{-1}\bar{\Delta}_{\gamma,\phi} &= n^{-1}\Delta_{\gamma,\phi}(\theta_0) + \frac{2}{n}A^a\Omega^a + \frac{2}{n}A^a\Gamma^a - \frac{2}{n}A^{ab}A^a\Omega^b + \\ &+ \frac{1}{n}(1-\gamma)\alpha^{abc}A^aA^b\Omega^c + \frac{1}{n}A^a\Omega^aA^b\Omega^b + \frac{1}{n^2}\Omega^a\Omega^a + O_p(n^{-5/2}). \end{aligned}$$

Since the difference $n^{-1}\bar{\Delta}_{\gamma,\phi} - n^{-1}\Delta_{\gamma,\phi}$ is of size $O_p(n^{-3/2})$, we must have $\bar{R}^r_{1;\gamma,\phi} = R^r_{1;\gamma,\phi} = A^r$. The expression of $\bar{R}^r_{2;\gamma,\phi}$ is obtained by matching $2\bar{R}^a_{1;\gamma,\phi}\bar{R}^a_{2;\gamma,\phi}$ with the terms of size $O_p(n^{-3/2})$ appearing in $n^{-1}\bar{\Delta}_{\gamma,\phi}$. Therefore the solution to the equation

$$2\bar{R}^a_{1;\gamma,\phi}\bar{R}^a_{2;\gamma,\phi} = -A^{ab}A^aA^b + \frac{1}{3}(1-\gamma)\alpha^{abc}A^aA^bA^c + \frac{2}{n}A^a\Omega^a,$$

leads to

$$\bar{R}^r_{2;\gamma,\phi} = R^r_{2;\gamma,\phi} + \frac{1}{n}\Omega^r.$$

Finally, by an analogous argument, the last element of $\bar{R}^r_{\gamma,\phi}$, $\bar{R}^r_{3;\gamma,\phi}$, is determined according to the following relation

$$\begin{aligned} 2\bar{R}^a_{1;\gamma,\phi}\bar{R}^a_{3;\gamma,\phi} &= A^{ab}A^{bc}A^aA^c + \frac{1}{4}(1+\phi)(1-\phi+2\gamma)A^aA^aA^bA^b + \frac{1}{3}(1-\gamma)A^{abc}A^aA^bA^c + \\ &- (1-\gamma)\alpha^{abc}A^{cd}A^aA^bA^d + \frac{1}{4}(1-\phi)(1+\phi-2\gamma)\alpha^{abc}\alpha^{cde}A^aA^bA^dA^e + \\ &+ \frac{1}{12}(-1+3\gamma+\gamma^2-6\gamma\phi+3\phi^2)\alpha^{abcd}A^aA^bA^cA^d + \frac{2}{n}A^a\Gamma^a - \frac{2}{n}A^{ab}A^a\Omega^b \\ &+ \frac{1}{n}(1-\gamma)\alpha^{abc}A^aA^b\Omega^c + \frac{1}{n}A^a\Omega^aA^b\Omega^b + \frac{1}{n^2}\Omega^a\Omega^a - \bar{R}^a_{2;\gamma,\phi}\bar{R}^a_{2;\gamma,\phi}. \end{aligned}$$

After some painstaking work we get

$$\bar{R}^r_{3;\gamma,\phi} = R^r_{3;\gamma,\phi} + \frac{1}{n}\Gamma^r - \frac{1}{2n}A^{ra}\Omega^a + \frac{1}{3n}(1-\gamma)\alpha^{rab}A^a\Omega^b + \frac{1}{2n}\Omega^rA^a\Omega^a.$$

This completes the proof.

Proof of Theorem 1

To enable Bartlett correctability of $\bar{\Delta}_{\gamma,\phi}$ it suffices to match order by order the expansions of $\bar{R}_{\gamma,\phi}$ and $\bar{R}_{-1,-1}$. Since the differences of these expansions arise at $O_p(n^{-1})$ and $O_p(n^{-3/2})$, we firstly need to determine Ω^r according to

$$\bar{R}^r_{2;\gamma,\phi} - R^r_{2;-1,-1} = \frac{1}{6}(1-\gamma)\alpha^{rab}A^aA^b + \frac{1}{n}\Omega^r - \frac{1}{3}\alpha^{rab}A^aA^b = 0,$$

which leads to

$$\Omega^r = \frac{n}{6}(1+\gamma)\alpha^{rab}A^aA^b.$$

Once the expression of Ω is plugged into $\bar{R}_{3;\gamma,\phi}$, we get

$$\begin{aligned}\bar{R}^r_{3;\gamma,\phi} &= R^r_{3;\gamma,\phi} - \frac{1}{24}(1+\gamma)\alpha^{abc}A^{ra}A^bA^c - \frac{1}{24}(1+\gamma)\alpha^{rab}A^{ca}A^bA^c + \\ &+ \frac{1}{18}(1-\gamma^2)\alpha^{rab}\alpha^{bcd}A^aA^cA^d + \frac{1}{72}n(1+\gamma)^2\alpha^{rab}\alpha^{cde}A^aA^bA^cA^dA^e + \frac{1}{n}\Gamma^r.\end{aligned}$$

Finally, we proceed to match the expression of $R^r_{3;-1,-1}$ by solving for Γ^r the equation

$$\bar{R}^r_{3;\gamma,\phi} - R^r_{3;-1,-1} = 0.$$

After some lengthy but routine algebraic work we obtain

$$\begin{aligned}\Gamma^r &= \frac{n}{6}\Big\{-\frac{6}{8}(1+\phi)(1-\phi+2\gamma)A^rA^aA^a - (1+\gamma)A^{rab}A^aA^b - (1+\gamma)\alpha^{rab}A^{bc}A^aA^c + \\ &- (1+\gamma)\alpha^{abc}A^{rb}A^aA^c + \frac{1}{12}(9\phi^2+5\gamma^2-18\gamma\phi+16\gamma+20)\alpha^{rab}\alpha^{bcd}A^aA^cA^d + \\ &- \frac{1}{4}(3\phi^2+\gamma^2-6\gamma\phi+3\gamma+5)\alpha^{rabc}A^aA^bA^c - \frac{1}{12}n(1+\gamma)^2\alpha^{rab}\alpha^{cde}A^aA^bA^cA^dA^e\Big\}.\end{aligned}$$

The expressions for Ω and Γ provided above are suitable for the case $\alpha^{rs} = \delta^{rs}$. To restore the scale we need to apply the following substitutions: $\alpha^{rst} \to \eta_{ra}\eta_{sb}\eta_{tc}\alpha^{abc}$, $\alpha^{rstu} \to \eta_{ra}\eta_{sb}\eta_{tc}\eta_{ud}\alpha^{abcd}$, $A^r \to \eta_{ra}A^a$, $A^{rs} \to \eta_{ra}\eta_{sb}A^{ab}$, and $A^{rst} \to \eta_{ra}\eta_{sb}\eta_{tc}A^{abc}$, where η_{rs} is a matrix satisfying $\alpha_{rs} = \eta_{ar}\eta_{as}$. Once we evaluate the adjustments at a generic value of θ we obtain the expressions provided in Theorem 1. This completes the proof.

Proof of Theorem 2

In order to prove the part (i) of the theorem we only need to observe that the empirical counterpart of theoretical adjustment $\Gamma^r = \Gamma^r(\theta_0)$ satisfies $\hat{\Gamma}^r = \Gamma^r + O_p(n^{-1})$ (see the discussion at the beginning of Section 3.2).

The part (ii) of the theorem is simple to deal with once $h(\cdot;\theta)$ satisfies the conditions required in the statement since

$$\begin{aligned}\hat{\alpha}^{rs} &= n^{-1}\sum_{i=1}^n[h^r(x_i;\theta_0) - \bar{h}^r(x;\theta_0)][h^s(x_i;\theta_0) - \bar{h}^s(x;\theta_0)] + O_p(n^{-1}) = \\ &= \alpha^{rs} + A^{rs} + O_p(n^{-1}) \\ \hat{\alpha}^{rst} &= n^{-1}\sum_{i=1}^n[h^r(x_i;\theta_0) - \bar{h}^r(x;\theta_0)][h^s(x_i;\theta_0) - \bar{h}^s(x;\theta_0)][h^t(x_i;\theta_0) - \bar{h}^t(x;\theta_0)] + O_p(n^{-1}) = \\ &= \alpha^{rst} + A^{rst} - \delta^{rs}A^t - \delta^{rt}A^s - \delta^{st}A^r + O_p(n^{-1}),\end{aligned}$$

where $\bar{h}^r(x;\theta_0) = A^r$, and with these relations it is easy to verify the following

$$\hat{\Omega}^r = \Omega^r + \frac{n}{6}(1+\gamma)(A^{rab}A^aA^b - 3A^rA^aA^a - \alpha^{rab}A^{bc}A^aA^c - \alpha^{abc}A^{rb}A^aA^c) + O_p(n^{-1}). \quad (22)$$

When $\hat{\Omega}^r$ is plugged in $\bar{R}_{2;\gamma,\phi}$, the second term in the right hand side of (22) enters in the expression $\bar{R}_{3;\gamma,\phi}$ so that the difference $\bar{R}_{3;\gamma,\phi} - R_{3;-1,-1}$ is free of terms involving A^{rs} and A^{rst}.

Finally, as remarked in the proof of Theorem 1, the expressions for $\hat{\Omega}$ and $\hat{\Gamma}$ suitable for the case $\alpha^{rs} \neq \delta^{rs}$ are obtained by applying the following substitutions: $\hat{\alpha}^{rst} \to \hat{\eta}_{ra}\hat{\eta}_{sb}\hat{\eta}_{tc}\hat{\alpha}^{abc}$, $\hat{\alpha}^{rstu} \to \hat{\eta}_{ra}\hat{\eta}_{sb}\hat{\eta}_{tc}\hat{\eta}_{ud}\hat{\alpha}^{abcd}$, $\hat{A}^r \to \hat{\eta}_{ra}\hat{A}^a$, where $\hat{\eta}_{rs}$ satisfies $\hat{\alpha}_{rs} = \hat{\eta}_{ar}\hat{\eta}_{as}$. This proves the theorem.

Proof of Corollary 1 and Corollary 2

The proof of Theorem 2 in Chen and Cui (2007), when the situation of just-identified moment restrictions is considered, can be reproduced verbatim for $\bar{\Delta}^B_{\gamma,\phi}$ either in the case of Corollary 1 or Corollary 2. It suffices to observe that after the adjustment $\bar{\Delta}_{\gamma,\phi} = \Delta_{-1,-1} + O_p(n^{-3/2})$ each pair (γ, ϕ). Therefore, the proofs of our corollaries are omitted.

References

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

Bhattacharya, R. and Ghosh, J. (1978). On the validity of the formal edgeworth expansion. *Ann. Statist.*, **6**, 434–451.

Camponovo, L. and Otsu, T. (2014). On bartlett correctability of empirical likelihood in generalized power divergence family. *Statist. Prob. Lett.*, **86**, 38 –43.

Chandler, R. and Bate, S. (2007). Inference for clustered data using the independence loglikelihood. *Biometrika*, **94**, 167–183.

Chen, J., Variyath, A., and Abraham, B. (2008). Adjusted empirical likelihood and its properties. *J. Comput. Graph. Statist.*, **17**, 426–443.

Chen, S. and Cui, H. (2007). On the second-order properties of empirical likelihood with moment restrictions. *J. Econometrics*, **141**, 492–516.

Corcoran, S. (1998). Bartlett adjustment of empirical discrepancy statistics. *Biometrika*, **85**, 967–972.

Corcoran, S., Davison, A., and Spady, R. (1995). *Reliable inference from empirical likelihoods*. Nuffield College Oxford.

Cox, D. and Reid, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika*, **91**, 729–737.

DiCiccio, T., Hall, P., and Romano, J. (1991). Empirical likelihood is Bartlett-correctable. *Ann. Statist.*, **19**, 1053–1061.

Emerson, S. and Owen, A. (2009). Calibration of the empirical likelihood method for a vector mean. *Electronic J. Statist.*, **3**, 1161–1192.

Jing, B. and Wood, A. (1996). Exponential empirical likelihood is not Bartlett-correctable. *Ann. Statist.*, **24**, 365–369.

Jørgensen, B. (1983). Maximum likelihood estimation and large-sample inference for generalized linear and nonlinear regression models. *Biometrika*, **70**, 19–28.

Kent, J. (1982). Robust properties of likelihood ratio tests. *Biometrika*, **69**, 19–27.

Li, X., Chen, J., Wu, Y., and Tu, D. (2011). Constructing nonparametric likelihood confidence regions with high order precisions. *Statist. Sinica*, **21**, 1767.

Lindsay, B. (1988). Composite likelihood methods. In *Statistical inference from stochastic processes (Ithaca, NY, 1987)*, volume 80 of *Contemp. Math.*, pages 221–239. Amer. Math. Soc., Providence, RI.

Liu, Y. and Chen, J. (2010). Adjusted empirical likelihood with high-order precision. *Ann. Statist.*, **38**, 1341–1362.

Lunardon, N., Pauli, F., and Ventura, L. (2013). A note on empirical likelihoods derived from pairwise score functions. *J. Statist. Comput. Simul.*, **83**, 1405–1414.

McCullagh, P. (1987). *Tensor methods in statistics*. Chapman and Hall, London.

Molenberghs, G. and Verbeke, G. (2005). *Models for discrete longitudinal data*. Springer, New York.

Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, **75**, 237–249.

Owen, A. (1990). Empirical likelihood ratio confidence regions. *Ann. Statist.*, **18**, 90–120.

Owen, A. (2001). *Empirical likelihood*. Chapman and Hall/CRC Press, New York.

Pace, L., Salvan, A., and Sartori, N. (2011). Adjusting composite likelihood ratio statistics. *Statist. Sinica*, **21**, 129–148.

Read, T. and Cressie, N. (1988). *Goodness-of-fit statistics for discrete multivariate data*. Springer-Verlag, New York.

Skovgaard, I. (1981). Transformation of an edgeworth expansion by a sequence of smooth functions. *Scandinavian J. Statist.*, **8**, 207–217.

Tsao, M. (2004). Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. *Ann. Statist.*, **32**, 1215–1221.

Varin, C., Reid, N., and Firth, D. (2011). An overview of composite likelihood methods. *Statist. Sinica*, **21**, 5–42.

Acknowledgements

The Authors are grateful to L. Pace for reading the early version of the manuscript and for his valuable suggestions.

Working Paper Series
Department of Statistical Sciences, University of Padua

You may order paper copies of the working papers by emailing wp@stat.unipd.it

Most of the working papers can also be found at the following url: http://wp.stat.unipd.it

UNIVERSITA DEGLI STUDI DI PADOVA
DIPARTIMENTO DI SCIENZE STATISTICHE

Department of Statistical Sciences
University of Padua
Italy