Submitted to the Annals of Statistics
arXiv: arXiv:0000.0000

EMPIRICAL LIKELIHOOD ON THE FULL PARAMETER SPACE

By Min Tsao* and Fan Wu

University of Victoria

We extend the empirical likelihood of Owen (1990) by partitioning its domain into the collection of its contours and mapping the contours through a continuous sequence of similarity transformations onto the full parameter space. The resulting extended empirical likelihood is a natural generalization of the original empirical likelihood to the full parameter space; it has the same asymptotic properties and identically shaped contours as the original empirical likelihood. It can also attain the second order accuracy of the Bartlett corrected empirical likelihood of DiCiccio, Hall and Romano (1991). A simple first order extended empirical likelihood is found to be substantially more accurate than the original empirical likelihood. It is also more accurate than available second order empirical likelihood methods in most small sample situations and competitive in accuracy in large sample situations. Importantly, in many one-dimensional applications this first order extended empirical likelihood is accurate for sample sizes as small as ten, making it a practical and reliable choice for small sample empirical likelihood inference.

1. Introduction. Since the seminal work of Owen (1988, 1990), there have been many advances in empirical likelihood method that have brought applications of the method to virtually every area of statistical research. It has been widely observed (*e.g.,* Hall and La Scala, 1990; Qin and Lawless, 1994; Corcoran, Davison and Spady, 1995; Owen, 2001; Liu and Chen, 2010) that empirical likelihood ratio confidence regions can have poor accuracy, especially in small sample and multidimensional situations. In particular, there is a persistent undercoverage problem in that coverage probabilities of empirical likelihood ratio confidence regions tend to be lower than the nominal levels. In this paper, we tackle a fundamental problem underlying the poor accuracy and undercoverage, that is, the empirical likelihood is defined on only a part of the parameter space. We call this the *mismatch problem* between the domain of the empirical likelihood and the parameter

*Principal and corresponding author; supported by an individual Discovery Grant from the National Science and Engineering Research Council of Canada.

AMS 2000 subject classifications: Primary 62G20; secondary 62E20

Keywords and phrases: Empirical likelihood, Bartlett correction, extended empirical likelihood, similarity transformation, composite similarity transformation.

space. We solve this problem through a geometric approach that expands the domain to the full parameter space. Our solution brings about substantial improvements in accuracy of the empirical likelihood inference and is particularly useful for small sample and multidimensional situations.

To see the mismatch problem, consider the example of empirical likelihood for the mean based on n observations $X_1, \ldots, X_n$ of a random vector $X \in \mathbb{R}^d$. The underlying parameter space Θ is $\mathbb{R}^d$ itself. But for a given $\theta \in \mathbb{R}^d$, when the convex hull of the $(X_i - \theta)$ does not contain 0, the empirical likelihood $L(\theta)$, the empirical likelihood ratio $R(\theta) = n^n L(\theta)$ and the empirical log-likelihood ratio $l(\theta) = -2 \log R(\theta)$ are all undefined. When this occurs, an established convention assigns $L(\theta) = 0$, and technically $L(\theta)$, $R(\theta)$ and $l(\theta)$ are now defined over the full parameter space. Nevertheless, to highlight the difference between the natural domain of the empirical likelihood and the parameter space, we define the common domain of $L(\theta)$, $R(\theta)$ and $l(\theta)$ as

$$\Theta_n = \{\theta : \theta \in \Theta \text{ and } l(\theta) < +\infty\}. \tag{1.1}$$

With this definition, the mismatch can now be expressed as $\Theta_n \subset \Theta$. For the mean, Θ_n is the interior of the convex hull of the X_i, which is indeed a proper subset of $\Theta = \mathbb{R}^d$. The mismatch $\Theta_n \subset \Theta$ holds for empirical likelihoods in general as the basic formulation common to all empirical likelihoods has a convex hull constraint on the origin, such as the one for the mean above, which may be violated by some θ values in the parameter space Θ.

The convex hull constraint violation underlying the mismatch is well-known in the empirical likelihood literature. It was first noted in Owen (1990) for the case of the mean. See also Owen (2001). To assess its impact on coverage probabilities of empirical likelihood ratio confidence regions, Tsao (2004) investigated bounds on coverage probabilities resulting from the convex hull constraint. To bypass this constraint, Bartolucci (2007) introduced a *penalized empirical likelihood* (PEL) for the mean which removes the convex hull constraint in the formulation of the original empirical likelihood (OEL) of Owen (1990, 2001) and replaces it with a penalizing term based on the Mahalanobis distance. For parameters defined by general estimating equations, Chen, Variyath and Abraham (2008) introduced an *adjusted empirical likelihood* (AEL) which retains the formulation of the OEL but adds a pseudo-observation to the sample. The AEL is just the OEL defined on the augmented sample. But due to the clever construction of the pseudo-observation, the convex hull constraint will never be violated by the AEL. Emerson and Owen (2009) showed that the AEL statistic has a boundedness problem which may lead to trivial 100% confidence regions. They proposed an extension of the AEL involving adding two pseudo-observations to

the sample to address the boundedness problem. Chen and Huang (2012) also addressed the boundedness problem by modifying the adjustment factor in the pseudo-observation. Liu and Chen (2010) proved a surprising result that under a certain level of adjustment, the AEL confidence region achieves the second order accuracy of the Bartlett corrected empirical likelihood (BEL) region by DiCiccio, Hall and Romano (1991). Recently, Lahiri and Mukhopadhyay (2012) showed that under certain dependence structures, a modified PEL for the mean works in the extremely difficult case of large dimension and small sample size. The PEL and the AEL are both defined on $\mathbb{R}^d$, and are thus free from the mismatch problem.

In this paper, we propose a new *extended empirical likelihood* (EEL) that is also free from the mismatch problem. We derive this EEL through the domain expansion method of Tsao (2013) which expands the domain of the OEL but retains its important geometric characteristics. This EEL makes effective use of the dimension information in the data and can attain the second order accuracy of the BEL. The most important aspect of this EEL, however, is that there is an easy-to-use first order version which is substantially more accurate than the OEL. This first order EEL is also more accurate than available second order empirical likelihood methods in most small sample and multidimensional applications and is comparable in accuracy to the latter when the sample size is large. The focus of the present paper is on the construction of EEL for the mean through which we introduce the basic idea of and important tools for expanding the OLE domain to the full parameter space. Under certain conditions, EEL for other parameters may also be constructed but this will be discussed elsewhere.

For brevity, we will use "OEL $l(\theta)$" and "EEL $l^*(\theta)$" to refer to the original and extended empirical log-likelihood ratios for the mean, respectively. Throughout this paper, we assume that the parameter space Θ is $\mathbb{R}^d$. The case where Θ is a known subset of $\mathbb{R}^d$ can be handled by finding EEL $l^*(\theta)$ defined on $\mathbb{R}^d$ first and then, for $\theta \notin \Theta$ only, redefine it as $l^*(\theta) = +\infty$.

2. Preliminaries. We review several key results and assumptions for developing the EEL defined on the full parameter space.

2.1. *Empirical likelihood for the mean.* Let $X \in \mathbb{R}^d$ be a random vector with mean θ_0 and covariance matrix Σ_0. Two assumptions we will need are:

(A_1) Σ_0 is a finite covariance matrix with full rank d; and
(A_2) $\limsup_{\|t\|\to\infty} |E[\exp\{it^T X\}]| < 1$ and $E\|X\|^{15} < +\infty$.

These are also assumptions under which the OEL for the mean is Bartlett correctable (DiCiccio, Hall and Romano, 1991; Chen and Cui, 2007).

Let $X_1, \ldots, X_n$ be independent copies of X where $n > d$. Let Θ_n be the collection of points in the interior of the convex hull of the X_i. For a $\theta \in \mathbb{R}^d$, Owen (1990) defined the empirical likelihood ratio $R(\theta)$ as

$$R(\theta) = \sup\left\{\prod_{i=1}^n nw_i | \sum_{i=1}^n w_i(X_i - \theta) = 0, w_i \geq 0, \sum_{i=1}^n w_i = 1\right\}. \tag{2.1}$$

It may be verified that $0 < R(\theta) \leq 1$ iff $\theta \in \Theta_n$. Also, $R(\theta) = 0$ if $\theta \notin \Theta_n$. Hence, the domain of the OEL $l(\theta) = -2\log R(\theta)$ is Θ_n. For a $\theta \in \Theta_n$, the method of Lagrange multipliers may be used to show that

$$l(\theta) = 2\sum_{i=1}^n \log\{1 + \lambda^T(X_i - \theta)\}, \tag{2.2}$$

where the multiplier $\lambda = \lambda(\theta) \in \mathbb{R}^d$ satisfies

$$\sum_{i=1}^n \frac{X_i - \theta}{1 + \lambda^T(X_i - \theta)} = 0. \tag{2.3}$$

Under assumption (A_1), Owen (1990) showed that OEL $l(\theta)$ satisfies

$$l(\theta_0) \xrightarrow{\mathcal{D}} \chi_d^2 \qquad \text{as } n \to +\infty. \tag{2.4}$$

For an $\alpha \in (0,1)$, let c be the $(1-\alpha)$th quantile of the χ_d^2 distribution. Then, the $100(1-\alpha)\%$ OEL confidence region for θ_0 is given by

$$\mathcal{C}_{1-\alpha} = \{\theta : \theta \in \Theta_n \text{ and } l(\theta) \leq c\}. \tag{2.5}$$

Under assumptions (A_1) and (A_2), DiCiccio, Hall and Romano (1988, 1991) showed that the coverage error of $\mathcal{C}_{1-\alpha}$ is $O(n^{-1})$, that is,

$$P(\theta_0 \in \mathcal{C}_{1-\alpha}) = P(l(\theta_0) \leq c) = P(\chi_d^2 \leq c) + O(n^{-1}). \tag{2.6}$$

More importantly, they showed that the empirical likelihood is Bartlett correctable. To give a brief account of this surprising result, let

$$\mathcal{C}'_{1-\alpha} = \{\theta : l(\theta)(1 - bn^{-1}) \leq c\} \tag{2.7}$$

be the Bartlett corrected empirical likelihood ratio confidence region where b is the *Bartlett correction constant* and $(1 - bn^{-1})$ is the *Bartlett correction factor*, DiCiccio, Hall and Romano (1988, 1991) showed that $\mathcal{C}'_{1-\alpha}$ has a coverage error of only $O(n^{-2})$, that is,

$$P(\theta_0 \in \mathcal{C}'_{1-\alpha}) = P[l(\theta_0)(1 - bn^{-1}) \leq c] = P(\chi_d^2 \leq c) + O(n^{-2}). \tag{2.8}$$

In practice, the Bartlett correction constant b cannot be determined since it depends on the moments of X which are not available in the non-parametric setting of the empirical likelihood. However, replacing the Bartlett correction factor in (2.8) with $[1 - bn^{-1} + O_p(n^{-3/2})]$ does not affect the $O(n^{-2})$ term in its right-hand side, that is

$$P\{l(\theta_0)[1 - bn^{-1} + O_p(n^{-3/2})] \leq c\} = P(\chi_d^2 \leq c) + O(n^{-2}). \tag{2.9}$$

This allows us to replace b in (2.7) and (2.8) with a $\sqrt{n}$-consistent estimate $\hat{b}$ without invalidating (2.8). See DiCiccio, Hall and Romano (1991) and Hall and La Scala (1990) for detailed discussions on Bartlett correction.

2.2. *Extended empirical likelihood.* The OEL confidence region $\mathcal{C}_{1-\alpha}$ in (2.5) is confined to the OEL domain Θ_n. This is a main cause of the undercoverage problem associated with $\mathcal{C}_{1-\alpha}$ (Tsao, 2004). To alleviate the problem, Tsao (2013) proposed to expand Θ_n which will lead to larger EL confidence regions. Let $h_n : \mathbb{R}^d \to \mathbb{R}^d$ be a bijective mapping and define a new empirical log-likelihood ratio $l^*(\theta)$ through the OEL $l(\theta)$ as follows:

$$l^*(\theta) = l(h_n^{-1}(\theta)) \qquad \text{for } \theta \in \mathbb{R}^d. \tag{2.10}$$

Then, the domain for the new empirical log-likelihood ratio is $\Theta_n^* = h_n^{-1}(\Theta_n)$. Here, h_n plays the role of reassigning or extending the OEL values of points in Θ_n to points in Θ_n^*. Because of this, Tsao (2013) named $l^*(\theta)$ the extended empirical log-likelihood ratio or simply EEL. In particular, Tsao (2013) used the following $\tilde{\theta}$-centred similarity mapping $h_n^* : \mathbb{R}^d \to \mathbb{R}^d$

$$h_n^*(\theta) = \tilde{\theta} + \gamma_n(\theta - \tilde{\theta}), \tag{2.11}$$

where $\tilde{\theta}$ is the sample mean and $\gamma_n \in \mathbb{R}^1$ is a constant (which we will refer to as the *expansion factor*) satisfying $\gamma_n \geq 0$ and $\gamma_n \to 1$ as $n \to +\infty$. If we choose $\gamma_n > 1$, then $\Theta_n \subset \Theta_n^* \subset \mathbb{R}^d$, and Θ_n^* alleviates the mismatch problem of Θ_n. The EEL confidence region for θ_0 is given by

$$\mathcal{C}_{1-\alpha}^* = \{\theta : \theta \in \Theta_n^* \text{ and } l^*(\theta) \leq c\}. \tag{2.12}$$

The advantages of the EEL based on h_n^* in (2.11) are: (1) the EEL confidence regions are similarly transformed OEL confidence regions, as such they retain the natural centre and shape of the OEL confidence regions, (2) the EEL can be applied to empirical likelihood inference for a wide range of parameters, and (3) with a properly selected constant γ_n, EEL confidence regions can achieve the second order accuracy of $O(n^{-2})$.

Nevertheless, the EEL based on h_n^* is only a partial solution to the mismatch problem because the domain of this EEL Θ_n^* is also a proper subset of $\mathbb{R}^d$. A second order version of this EEL has been found to have good accuracy in one- and two-dimensional problems. But it also tends to undercover and no accurate first order version of this EEL is available. These motivated us to consider an EEL defined on the full parameter space.

3. Extended empirical likelihood on the full parameter space. Consider a bijective mapping from the OEL domain to the parameter space, $h_n : \Theta_n \to \Theta = \mathbb{R}^d$. Under such a mapping, the EEL $l^*(\theta)$ given by (2.10) is well defined throughout $\mathbb{R}^d$ and is thus free from the mismatch problem. In this section, we first construct such a mapping using h_n^* in (2.11). We call it the *composite similarity mapping* and denote it by $h_n^C : \Theta_n \to \mathbb{R}^d$. We then study the asymptotic properties of the EEL $l^*(\theta)$ based on h_n^C.

3.1. *The composite similarity mapping.* The simple similarity mapping h_n^* in (2.11) maps OEL domain Θ_n onto a similar but bounded region in $\mathbb{R}^d$. If we think of Θ_n as a region consisting of distinct and nested contours of the OEL, then h_n^* expands all contours with the same constant expansion factor γ_n. In order to map Θ_n onto the full $\mathbb{R}^d$, we need to map contours on the outside more and more so that the images of the contours will fill up the entire $\mathbb{R}^d$. To achieve this, consider level-τ contour of the OEL $l(\theta)$,

$$c(\tau) = \{\theta : \theta \in \Theta_n \text{ and } l(\theta) = \tau\}, \tag{3.1}$$

where $\tau \geq 0$. The contours form a partition of the OEL domain,

$$\Theta_n = \bigcup_{\tau \in [0,+\infty)} c(\tau). \tag{3.2}$$

In light of (3.2), the centre of Θ_n is $c(0) = \{\tilde{\theta}\}$ and the outwardness of a $c(\tau)$ with respect to the centre is indexed by τ; the larger the τ value, the more outward $c(\tau)$ is. If we allow the expansion factor γ_n to be a continuous monotone increasing function of τ and allow γ_n to go to infinity when τ goes to infinity, then (such a variation of) h_n^* will map Θ_n onto $\mathbb{R}^d$. Hence we define the composite similarity mapping $h_n^C : \Theta_n \to \mathbb{R}^d$ as follows:

$$h_n^C(\theta) = \tilde{\theta} + \gamma(n, l(\theta))(\theta - \tilde{\theta}) \quad \text{for } \theta \in \Theta_n \ , \tag{3.3}$$

where $\gamma(n, l(\theta))$ is given by

$$\gamma(n, l(\theta)) = 1 + \frac{l(\theta)}{2n}. \tag{3.4}$$

Function $\gamma(n, l(\theta))$ is the new expansion factor which depends continuously on θ through the value of $l(\theta)$ or $\tau = l(\theta)$. For convenience, we will emphasis the dependence of $\gamma(n, l(\theta))$ on $l(\theta)$ instead of θ or τ. This new expansion factor has the two desired properties discussed above:

$$\text{for a fixed } n, \text{ if } l(\theta_1) < l(\theta_2), \text{ then } \gamma(n, l(\theta_2)) < \gamma(n, l(\theta_2)); \text{ and} \tag{3.5}$$

$$\text{for a fixed } n,\ \gamma(n, l(\theta)) \to +\infty \text{ as } l(\theta) \to +\infty. \tag{3.6}$$

The inclusion of the denominator $2n$ in (3.4) ensures that the expansion factor converges to 1, reflecting the fact that there is no need for domain expansion for large sample sizes. Also, the constant 2 in the denominator provides extra adjustment to the speed of expansion and may be replaced with other positive constants (see Figure 1). We choose to use 2 here as the corresponding $\gamma(n, l(\theta))$ in (3.4) is asymptotically equivalent to a likelihood based expansion factor $\gamma(n, L(\theta)) = L(\theta)^{-1/n}$ which we had first considered and was found to give accurate numerical results. The definition of $\gamma(n, l(\theta))$ in (3.4) uses $l(\theta)$ instead of $L(\theta)$ because of convenience for theoretical investigations. A more general form of $\gamma(n, l(\theta))$ will be considered later.

Theorem 3.1 below summarizes the key properties of the composite similarity mapping h_n^C. Its proof and that of subsequent theorems and lemmas may all be found in the Appendix.

THEOREM 3.1. *Under assumption (A_1), the composite similarity mapping $h_n^C : \Theta_n \to \mathbb{R}^d$ defined by (3.3) and (3.4) satisfies:*

(i) it has a unique fixed point at the mean $\tilde{\theta}$;
(ii) it is a similarity mapping for each individual $c(\tau)$; and
(iii) it is a bijective mapping from Θ_n to $\mathbb{R}^d$.

Because of (*ii*) above, we call h_n^C the composite similarity mapping as it may be viewed as a continuous sequence of simple similarity mappings from $\mathbb{R}^d$ to $\mathbb{R}^d$ indexed by $\tau = l(\theta) \in [0, +\infty)$. The "$\tau$-th" mapping from this sequence has expansion factor $\gamma(n, l(\theta)) = \gamma(n, \tau)$. It is just the simple similarity mapping h_n^* in (2.11) with $\gamma_n = \gamma(n, \tau)$, and is used exclusively to map the "τ-th" OEL contour $c(\tau)$. The latter has been implicitly built into h_n^C since for all $\theta \in c(\tau)$, $l(\theta) = \tau$ which implies the corresponding expansion factor of h_n^C is the constant $\gamma(n, \tau)$ that defines the "τ-th" mapping.

It should be noted that h_n^C is not a similarity mapping from $\mathbb{R}^d$ to $\mathbb{R}^d$ itself due to the dependence of the expansion factor $\gamma(n, l(\theta))$ on θ and its domain Θ_n which is only a bounded subset of $\mathbb{R}^d$.

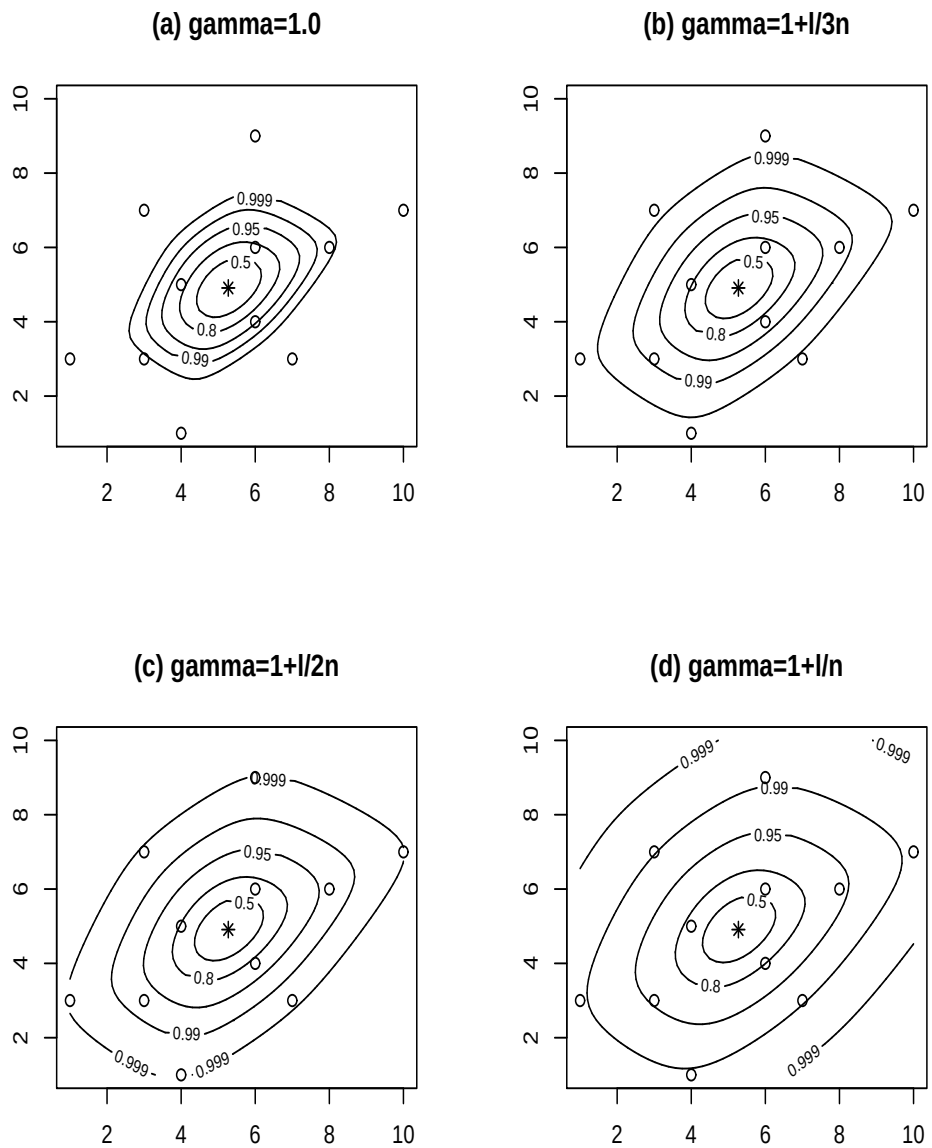

FIG 1. *(a) Contours of the OEL $l(\theta)$ (for which the expansion factor is 1.0); (b) contours of the EEL $l^*(\theta)$ with expansion factor $\gamma(11, l(\theta)) = 1+l(\theta)/3n$; (c) contours of the EEL $l^*(\theta)$ with $\gamma(11, l(\theta)) = 1 + l(\theta)/2n$; (d) contours of the EEL $l^*(\theta)$ with $\gamma(11, l(\theta)) = 1 + l(\theta)/n$. All four plots are based on the same random sample of 11 points shown in small circles. The star in the middle is the sample mean. The expansion factor increases as we go from plot (b) to plot (d), and correspondingly the EEL contours also become bigger in scale from plot (b) to (d). But the centre and shapes of contours are the same in all plots.*

3.2. *The extended empirical likelihood under the composite similarity mapping.* By Theorem 3.1, $h_n^C : \Theta_n \to \mathbb{R}^d$ is bijective. Hence it has an inverse which we denote by $h_n^{-C} : \mathbb{R}^d \to \Theta_n$. The EEL $l^*(\theta)$ under h_n^C is

$$l^*(\theta) = l(h_n^{-C}(\theta)) \quad \text{for } \theta \in \mathbb{R}^d,$$

which is defined throughout $\mathbb{R}^d$. The contours of $l^*(\theta)$ are larger in scale but have the same centre and identical shape as that of OEL $l(\theta)$. Figure 1 compares their contours with a sample of 11 two-dimensional observations. It shows that geometrically, mapping h_n^C is anchored at the sample mean $\tilde{\theta}$ as it is the fixed point of h_n^C that is not moved. From this anchoring point, the mapping pushes out/expands each OEL contour $c(\tau)$ proportionally in all directions at an expansion factor of $\gamma(n, \tau)$ to form an EEL contour. The boundary points of Θ_n are all pushed out to the infinity.

In the following, we will use θ' to denote the image of a $\theta \in \mathbb{R}^d$ under the inverse transformation h_n^{-C}, that is, $h_n^{-C}(\theta) = \theta' \in \Theta_n$. Of particular interest is the image of the unknown true mean θ_0,

$$h_n^{-C}(\theta_0) = \theta_0'. \tag{3.7}$$

Because the inverse transformation h_n^{-C} does not have an analytic expression, that for θ_0' is also not available. Nevertheless, Lemma 3.2 gives an asymptotic assessment on its distance to θ_0. The proof of Lemma 3.2 will need Lemma 3.1 below which shows that inside Θ_n, the OEL $l(\theta)$ is a "monotone increasing" function along each ray originating from the mean $\tilde{\theta}$.

LEMMA 3.1. *Under assumption (A_1), for a fixed point $\theta \in \Theta_n$ and any value $\alpha \in [0, 1]$ the OEL $l(\theta)$ satisfies*

$$l(\tilde{\theta} + \alpha(\theta - \tilde{\theta})) \leq l(\theta).$$

Denote by $[\tilde{\theta}, \theta_0]$ the line segment connecting $\tilde{\theta}$ and θ_0. Lemma 3.2 below shows that θ_0' is on $[\tilde{\theta}, \theta_0]$ and is asymptotically very close to θ_0.

LEMMA 3.2. *Under assumption (A_1), point θ_0' defined by equation (3.7) satisfies (i) $\theta_0' \in [\tilde{\theta}, \theta_0]$ and (ii) $\theta_0' - \theta_0 = O_p(n^{-3/2})$.*

Using θ_0', the EEL $l^*(\theta_0)$ can now be expressed as

$$l^*(\theta_0) = l(h_n^{-C}(\theta_0)) = l(\theta_0') = l(\theta_0 + (\theta_0' - \theta_0)). \tag{3.8}$$

The following theorem gives the asymptotic distribution of $l^*(\theta_0)$:

THEOREM 3.2. *Under assumption (A_1) and with the composite similarity mapping h_n^C defined by (3.3) and (3.4), the EEL $l^*(\theta)$ satisfies*

$$l^*(\theta_0) \xrightarrow{\mathcal{D}} \chi_d^2 \qquad \text{as } n \to +\infty. \tag{3.9}$$

The proof of Theorem 3.2 is based on the observation that $\|\theta_0' - \theta_0\|$ is asymptotically very small. This and (3.8) imply that $l^*(\theta_0) = l(\theta_0) + o_p(1)$. This proof demonstrates an advantage of the EEL: it has the simple relationship with the OLE shown in (3.8) through which we make use of known asymptotic properties of the OEL to study the EEL. Our derivation of a second order EEL below further explores this advantage.

3.3. *Second order extended empirical likelihood.* It may be verified that Theorems 3.1 and 3.2 also hold under any composite similarity mapping defined by (3.3) and the following general form of the expansion factor,

$$\gamma(n, l(\theta)) = 1 + \frac{\kappa[l(\theta)]^{\delta(n)}}{n^m}, \tag{3.10}$$

where κ and m are both positive constants and $\delta(n)$ is a bounded function of n satisfying $\varepsilon_n < \delta(n) \le a$ for some constants $a > \varepsilon_n > 0$. The availability of a whole family of $\gamma(n, l(\theta))$ functions for the construction of the EEL provides an opportunity to optimize our choice of this function to achieve the second order accuracy. Theorem 3.3 below gives the optimal choice.

THEOREM 3.3. *Assume (A_1) and (A_2) hold and denote by $l_s^*(\theta)$ the EEL under the composite similarity mapping (3.3) with expansion factor*

$$\gamma_s(n, l(\theta)) = 1 + \frac{b}{2n}[l(\theta)]^{\delta(n)}, \tag{3.11}$$

where $\delta(n) = O(n^{-1/2})$ and b is the Bartlett correction constant in (2.7). Then

$$l_s^*(\theta_0) = l(\theta_0)[1 - bn^{-1} + O_p(n^{-3/2})], \tag{3.12}$$

and

$$P(l_s^*(\theta_0) \le c) = P(\chi_d^2 \le c) + O(n^{-2}). \tag{3.13}$$

In our subsequent discussions, we will refer to an $l_s^*(\theta)$ defined by the expansion factor $\gamma_s(n, l(\theta))$ in (3.11) as a *second order EEL* on the full parameter space. The EEL $l^*(\theta)$ defined by $\gamma(n, l(\theta))$ in (3.4) will henceforth be referred to as the *first order EEL* on the full parameter space.

The utility of the $\delta(n)$ in $\gamma_s(n, l(\theta))$ is to provide an extra adjustment for the speed of the domain expansion which ensures that $l_s^*(\theta)$ will behave asymptotically like the BEL and hence will have the second order accuracy of the BEL. For convenience, we set $\delta(n) = n^{-1/2}$. The resulting second order EEL turns out to be competitive in accuracy to the BEL and the second order AEL. For small sample and/or high dimension situations, confidence regions based on this second order EEL can have undercoverage problems like those based on the OEL and BEL. Fine-tuning of $\delta(n)$ for such situations is needed and methods of fine-tuning are discussed in Wu (2013).

Finally, we noted after Theorem 3.2 that EEL $l^*(\theta_0)$ can be expressed in terms of the OLE $l(\theta_0)$ as $l^*(\theta_0) = l(\theta_0) + o_p(1)$. The $o_p(1)$ term can be improved and in fact we have $l^*(\theta_0) = l(\theta_0) + O_p(n^{-1})$. See the remark following the proof of Theorem 3.2 in the Appendix. An even stronger connection between $l^*(\theta_0)$ and $l(\theta_0)$ is given by Corollary 3.1 below.

COROLLARY 3.1. *Under assumptions (A_1) and (A_2), the first order EEL $l^*(\theta)$ satisfies,*

$$l^*(\theta_0) = l(\theta_0)[1 - l(\theta_0)n^{-1} + O_p(n^{-3/2})]. \tag{3.14}$$

The proof of Corollary 3.1 follows from that for Theorem 3.3. This result provides a partial explanation for the remarkable numerical accuracy of confidence regions based on the first order EEL $l^*(\theta)$.

4. Numerical examples and comparisons. We now present a simulation study comparing the EEL with the OEL, BEL and AEL. Throughout this section, we use $l_1^*(\theta)$ or EEL_1 to denote the first order EEL with expansion factor (3.4), and use $l_2^*(\theta)$ or EEL_2 to denote the second order EEL given by expansion factor (3.11) where $\delta(n) = n^{-1/2}$.

4.1. *Low-dimensional examples.* Tables 1 and 2 contain simulated coverage probabilities of confidence regions for the mean based on first order methods OEL, EEL_1 and second order methods BEL, AEL, EEL_2, BEL*, AEL* and EEL_2^*. Here, BEL, AEL and EEL_2 are based on the theoretical Bartlett correction constant b, and BEL*, AEL* and EEL_2^* are based on $\tilde{b}_n$ which is a bias corrected estimate for b given by Liu and Chen (2010).

Table 1 gives four one-dimensional (1-d) examples. Table 2 contains four bivariate (BV or 2-d) examples; the first three were taken from Liu and Chen (2010), the fourth is a "2-d chi-square", and here are the details:

(BV_1): $X_1|D \sim N(0, D^2)$ and $X_2|D \sim \text{Gamma}(D^{-1}, 1)$.
(BV_2): $X_1|D \sim \text{Poisson}(D)$ and $X_2|D \sim \text{Poisson}(D^{-1})$.

(BV_3): $X_1|D \sim \text{Gamma}(D, 1)$ and $X_2|D \sim \text{Gamma}(D^{-1}, 1)$.
(BV_4): X_1 and X_2 are independent copies of a χ^2_1 random variable.

The D in BV_1, BV_2 and BV_3 is a uniform random variable on $[1, 2]$ which is used to induce dependence between X_1 and X_2. We included $n = 10, 30, 50$ representing, respectively, small, medium and large sample sizes. Each entry in the tables is based on 10,000 random samples of size n, shown in column 2, from the distribution in column 1. Here are our observations:

Table 1 and Table 2 about here.

(1) **BEL, AEL and EEL$_2$:** For $n = 30$ and 50, these three theoretical second order methods are extremely close in terms of coverage accuracy. This is to be expected as their coverage errors are all $O(n^{-2})$ which is very small for medium or large sample sizes.
For $n = 10$, the AEL statistic suffers from a boundedness problem (Emerson and Owen, 2009) which may lead to trivial 100% confidence regions or inflated coverage probabilities. This explains the 1.000's in various places in the AEL column and renders the AEL unsuitable for such small sample sizes. Between BEL and EEL$_2$, the latter is more accurate, especially for the 2-d examples.
Overall, EEL$_2$ is the most accurate theoretical second order method.
(2) **BEL*, AEL* and EEL$^*_2$:** For $n = 30$ and 50, the AEL* and EEL$^*_2$ are slightly more accurate than the BEL*, especially in 2-d examples.
For $n = 10$, the AEL* has higher coverage probabilities but these are inflated by and unreliable due to the boundedness problem. Also, EEL$^*_2$ is more accurate than BEL*. For the "2-d chi-square" in example BV_4, EEL$^*_2$ is at least 12% more accurate than the BEL*.
Overall, EEL$^*_2$ is the most reliable and accurate among the three.
(3) **OEL and EEL$_1$:** These first order methods are simpler than the second order methods as they do not require computation of the theoretical or estimated Bartlett correction factor. The EEL$_1$ is consistently and substantially more accurate than the OEL. In particular, for 2-d examples with $n = 10$, the EEL$_1$ is more accurate by about 10%.
(4) **EEL$_1$ versus EEL$^*_2$:** These are the most accurate practical first and second order methods, respectively. Surprisingly, EEL$_1$ turns out to be slightly more accurate than EEL$^*_2$. Only the (impractical) theoretical second order EEL$_2$ is comparable to EEL$_1$ in accuracy. This intriguing observation may be partially explained by Corollary 3.1 where it was shown that $l^*_1(\theta_0) = l(\theta_0)[1 + l(\theta_0)n^{-1} + O(n^{-3/2})]$, which resembles

the Bartlett corrected OEL in (2.9) with the constant b replaced by $l(\theta_0)$. However, this does not account for its good accuracy for small sample sizes, which is due to the fact that EEL_1 makes good use of the dimension information through the composite similarity mapping. We will further elaborate on this in Section 4.2.

(5) **EEL_1:** Overall, it is the most accurate among the eight methods that we have compared. Importantly, it is not just accurate in relative terms. It is sufficiently accurate in absolute terms for practical applications in most 1-d examples, including cases of $n = 10$. It is also quite accurate for 2-d examples when $n = 30, 50$.

4.2. *High-dimensional examples.* Table 3 contains simulated coverage probabilities for the mean of three high-dimensional multivariate normal distributions ($d = 5, 10, 15$). Our main interest here is to probe the small sample behaviour of all methods in high-dimension situations. Because of this, we have included only combinations of n and d where n/d, which we will refer to as the *effective sample size*, is very small ($2 \leq n/d \leq 10$). The following are our observations based on Table 3:

Table 3 about here.

(1) For these high-dimension examples, EEL_1 is the most accurate, surpassing even the theoretical second order EEL_2. Whereas the OEL uses dimension d only once through the degrees of freedom in the chi-square calibration, EEL_1 uses d twice. The expansion factor for EEL_1 is $1 + l(\theta)/2n$ which implicitly depends on d; the $100(1-\alpha)\%$ EEL_1 confidence region is just the $100(1-\alpha)\%$ OEL confidence region expanded by a factor of $1 + \chi^2_{d,1-\alpha}/2n$. Hence, EEL_1 uses d through the chi-square calibration of the OEL region and the expansion factor.
For a fixed $\alpha \in (0, 1)$, the chi-square quantile $\chi^2_{d,1-\alpha}$ and consequently the EEL_1 expansion factor $1 + \chi^2_{d,1-\alpha}/2n$ are increasing functions of d. Hence at a fixed n, EEL_1 automatically provides higher degrees of expansion for higher dimensions where this is indeed needed.

(2) For multivariate normal means, Table 3 shows that EEL_1 is accurate when the effective sample size satisfies $n/d \geq 5$. However, when the underlying distribution is heavily skewed, the effective sample size needed to achieve similar accuracy needs to be 15 or larger. See Table 2 for some 2-d examples to this effect.

(3) The AEL and AEL* broke down in most cases with 100% coverage probabilities. This further illustrates the observation that AEL meth-

ods may not be suitable when the effective sample size is small. Among OEL, BEL, EEL_2 and EEL_2^*, the two EEL methods are consistently more accurate but they are not sufficiently accurate for practical applications except for the case of $(d, n) = (5, 50)$.

4.3. *Confidence region size comparison.* For the 1-d examples in Table 1, we computed the average interval lengths of the five practical methods OEL, EEL_1, BEL*, AEL* and EEL_2*. Table 4 gives the average length of 1000 intervals of each method and n combination for the $N(0, 1)$ case. For $n = 10$, the average for AEL* is not available due to occurrences of unbounded intervals; the number beside the N/A is the proportion of times where this occurred. Not surprisingly, intervals with higher coverage probabilities in

Table 4: Average lengths of EL confidence intervals for $N(0, 1)$ mean

n	Level	OEL	EEL_1	BEL*	AEL*	EEL_2^*
10	.90	0.965	1.096	1.044	N/A(0.026)	1.077
	.95	1.149	1.370	1.242	N/A(0.058)	1.298
	.99	1.499	1.996	1.615	N/A(0.172)	1.731
30	.90	0.589	0.616	0.606	0.606	0.608
	.95	0.706	0.752	0.726	0.727	0.731
	.99	0.940	1.044	0.967	0.969	0.976
50	.90	0.460	0.473	0.467	0.468	0.468
	.95	0.551	0.572	0.560	0.560	0.561
	.99	0.732	0.780	0.744	0.744	0.747

Table 1 have larger average lengths. That of EEL_1 is the largest but it is not excessive relative to averages of other methods. As such, length is not a big disadvantage for EEL_1 as other methods have lower coverage probabilities.

For $d > 1$, sizes of the confidence regions are difficult to determine. But the relative size of an EEL region to the corresponding OEL region can be measured by the expansion factor. Table 5 contains values of the expansion factor for 95% EEL_1 regions at some combinations of n and d. The expansion

Table 5: Values of expansion factor for 95% EEL_1 confidence regions

n	$d = 1$	$d = 2$	$d = 3$	$d = 5$	$d = 10$	$d = 15$
10	1.192	1.299	1.390	1.553	N/A	N/A
15	1.128	1.199	1.260	1.369	1.610	N/A
20	1.096	1.149	1.195	1.276	1.457	1.624
30	1.064	1.099	1.130	1.184	1.305	1.416
50	1.038	1.059	1.078	1.110	1.183	1.249

factor increases when d goes up but decreases when n goes up, responding

to the need for more expansion in higher dimension situations and the need for less expansion when the sample size is large.

Finally, we briefly comment on the computation concerning the EEL $l_1^*(\theta)$. To compute $l_1^*(\theta)$ at a given $\theta \in \mathbb{R}^d$ which is just OEL $l(\theta')$ where θ' satisfies equation $h_n^C(\theta') = \theta$, we need to find the multivariate root for function $f(\theta') = h_n^C(\theta') - \theta$. This is seen as a nonlinear multivariate problem but it is easily reduced to a simpler *univariate* problem due to the fact that $\theta' \in [\tilde{\theta}, \theta]$ (see Lemma 3.2 and its proof). When using $l_1^*(\theta)$ for hypothesis testing or when simulating the coverage probabilities of the EEL confidence regions, we may use the fact that $l_1^*(\theta) \leq l(\theta)$. Hence we can compute $l(\theta)$ first and if it is smaller than the critical value, then there is no need to compute $l_1^*(\theta)$ because it must also be smaller than the critical value. Incorporating these observations, our R code for computing the EEL runs quite fast.

5. Concluding remarks. The geometric motivation of the domain expansion method is simple: since the OEL confidence region tends to be too small, an expansion of the OEL confidence region should help to ease its undercoverage problem. What needed to be determined then are the manner in which the expansion should take place and the amount of expansion that would be appropriate. The composite similarity mapping of the present paper is an effective way to undertake the expansion as it solves the mismatch problem and retains all important geometric characteristics of the OEL contours. With the impressive numerical accuracy of the EEL_1, the particular amount of expansion represented by its expansion factor (3.4) would be appropriate for general applications of the EEL method.

The EEL is readily constructed for parameters defined by general estimating equations. For such parameters, we use the maximum empirical likelihood estimator (MELE) $\tilde{\theta}$ to define the composite similarity mapping in (3.3). Under certain conditions on the estimating function which also guarantee the $\sqrt{n}$-consistency of the MELE, Lemma 3.2 and all three theorems of this paper remain valid. A detailed treatment of the EEL in this setting may be found in a technical report by Tsao and Fan (2013). See also Tsao (2013) for an EEL for estimating equations under the simple similarity transformation (2.11). For parameters outside of the standard estimating equations framework, the EEL on full parameter space may also be defined through a composite similarity mapping centred on the MELE, but its asymptotic properties need to be investigated for each case separately.

To conclude, the simple first order EEL_1 is a practical and reliable method that is remarkably accurate when the effective sample size is not too small. It is also easy to use. Hence we recommend it for real applications of the

empirical likelihood method. An intriguing question that remains largely unanswered is why this first order method has such good accuracy relative to the OEL and the second order methods. Corollary 3.1 and the first remark in Section 4.2 suggested, respectively, possible asymptotic and finite sample reasons, but a more convincing theoretical explanation is needed.

6. Appendix. We give proofs for lemmas and theorems below.

Proof of Theorem 3.1. Assumption (A_1) and $n > d$ imply that, with probability 1, the convex hull of the X_i is non-degenerate. This implies the OEL $l(\theta)$ has an open domain $\Theta_n \subset \mathbb{R}^d$, a condition that is required for the implementation of OEL domain expansion via composite similarity mapping. Subsequent proofs all require this condition which, hereafter, is assumed whenever (A_1) is, and for brevity will not explicitly restated.

Part (i) is a simple consequence of the observation that $\gamma(n, l(\theta)) \geq 1$. To show (ii), let n and τ be fixed, and consider the level-τ OEL contour $c(\tau)$ defined by (3.1). For $\theta \in c(\tau)$, $l(\theta) = \tau$. Thus the composite similarity mapping h_n^C simplifies to $h_n^C(\theta) = \tilde{\theta} + \gamma_n(\theta - \tilde{\theta})$ for $\theta \in c(\tau)$ where $\gamma_n = \gamma(n, \tau)$ is constant. This is the simple similarity mapping in (2.11).

To prove (iii), we need to show that $h_n^C : \Theta_n \to \mathbb{R}^d$ is both surjective and injective. We first show it is surjective, that is, for any given $\theta' \in \mathbb{R}^d$, there exists a $\theta'' \in \Theta_n$ such that $h_n^C(\theta'') = \theta'$. Consider the ray originating from $\tilde{\theta}$ and through θ'. Introduce a univariate parametrization of this ray,

$$\theta = \theta(\zeta) = \tilde{\theta} + \zeta\vec{\theta},$$

where $\vec{\theta}$ is the unit vector $(\theta' - \tilde{\theta})/\|\theta' - \tilde{\theta}\|$ in the direction of the ray and $\zeta \in [0, \infty)$ is the distance between θ (a point on the ray) and $\tilde{\theta}$. Define

$$\zeta_b = \inf\{\zeta : \zeta \in [0, +\infty) \text{ and } \theta(\zeta) \notin \Theta_n\}.$$

Then, $\theta(\zeta) \in \Theta_n$ for all $\zeta \in [0, \zeta_b)$. But $\theta(\zeta_b) \notin \Theta_n$ because Θ_n is open. It follows that $\zeta_b > 0$ as it represents the distance between $\tilde{\theta}$, an interior point of the open Θ_n, and $\theta(\zeta_b)$ which is a boundary point of Θ_n.

Now, consider the following univariate function defined on $[0, \zeta_b)$,

$$f(\zeta) = \gamma(n, l(\theta(\zeta)))\zeta.$$

We have $f(0) = \gamma(n, l(\tilde{\theta})) \times 0 = \gamma(n, 0) \times 0 = 0$. Also, by (3.6),

$$\lim_{\zeta \to \zeta_b} f(\zeta) = \lim_{\zeta \to \zeta_b} \gamma(n, l(\theta(\zeta)))\zeta = \zeta_b \lim_{\zeta \to \zeta_b} \gamma(n, l(\theta(\zeta))) = +\infty.$$

Hence by the continuity of $f(\zeta)$, for $\zeta' = \|\theta' - \tilde{\theta}\| \in [0, +\infty)$, there exists a $\zeta'' \in [0, \zeta_b)$ such that $f(\zeta'') = \zeta'$. Let $\theta'' = \theta(\zeta'')$. Then, $\theta'' \in \Theta_n$ since $\zeta'' \in [0, \zeta_b)$. Also, $h_n^C(\theta'') = \tilde{\theta} + \gamma(n, l(\theta''))(\theta'' - \tilde{\theta}) = \theta'$. Hence θ'' is the desired point in Θ_n that satisfies $h_n^C(\theta'') = \theta'$ and h_n^C is surjective.

To show that h_n^C is also injective, first note that for a given OEL contour $c(\tau)$, the mapping $h_n^C : c(\tau) \to c^*(\tau)$ is injective because for $\theta \in c(\tau)$, h_n^C is equivalent to the similarity mapping in (2.11) which is bijective from $\mathbb{R}^d$ to $\mathbb{R}^d$. By the partition of the OEL domain Θ_n in (3.2), two different points θ_1, θ_2 from Θ_n are either [a] on the same contour $c(\tau)$ where $\tau = l(\theta_1) = l(\theta_2)$ or [b] on two separate contours $c(\tau_1)$ and $c(\tau_2)$, respectively, where $\tau_1 = l(\theta_1) \neq l(\theta_2) = \tau_2$. Under [a], $h_n^C(\theta_1) \neq h_n^C(\theta_2)$ because $h_n^C : c(\tau) \to c^*(\tau)$ is injective. Under [b], $h_n^C(\theta_1) \neq h_n^C(\theta_2)$ also holds as (3.5) implies $c^*(\tau_1) \cap c^*(\tau_2) = \varnothing$. □

Proof of Lemma 3.1. For a fixed $\theta \in \Theta_n$, $l(\theta)$ is a fixed quantity in $[0, +\infty)$. Define an OEL confidence region for the mean using $l(\theta)$ as follows,

$$\mathcal{C}_\theta = \{\theta' : \ \theta' \in \mathbb{R}^d \text{ and } l(\theta') \leq l(\theta) \ \}. \tag{6.1}$$

Then, $\mathcal{C}_\theta$ is a convex set in $\mathbb{R}^d$. See Owen (1990) and Hall and La Scala (1990). Since $l(\tilde{\theta}) = 0$ and $l(\theta) \geq 0$, $\tilde{\theta}$ is in $\mathcal{C}_\theta$. Further, by construction, θ itself is also in $\mathcal{C}_\theta$. It follows from the convexity of $\mathcal{C}_\theta$ that for any $\alpha \in [0, 1]$,

$$\theta^* = (1 - \alpha)\tilde{\theta} + \alpha\theta = \tilde{\theta} + \alpha(\theta - \tilde{\theta})$$

must also be in $\mathcal{C}_\theta$. By (6.1), $l(\theta^*) \leq l(\theta)$. Thus, $l(\tilde{\theta} + \alpha(\theta - \tilde{\theta})) \leq l(\theta)$. □

Proof of Lemma 3.2. Since $\theta_0 = h_n^C(\theta_0') = \tilde{\theta} + \gamma(n, l(\theta_0'))(\theta_0' - \tilde{\theta})$,

$$\theta_0 - \tilde{\theta} = \gamma(n, l(\theta_0'))(\theta_0' - \tilde{\theta}). \tag{6.2}$$

Noting that $\gamma(n, l(\theta)) \geq 1$, (6.2) implies θ_0' is on the ray originating from $\tilde{\theta}$ and through θ_0 and $\|\theta_0 - \tilde{\theta}\| \geq \|\theta_0' - \tilde{\theta}\|$. Hence, $\theta_0' \in [\tilde{\theta}, \theta_0]$.

Without loss of generality, we assume that $\theta_0 \in \Theta_n$. See Owen (1990). By the convexity of Θ_n, $[\tilde{\theta}, \theta_0] \subset \Theta_n$. It follows from Lemma 3.1 that

$$0 = l(\tilde{\theta}) \leq l(\theta_0') \leq l(\theta_0).$$

This and the fact that $l(\theta_0) = O_p(1)$ imply $l(\theta_0') = O_p(1)$. Hence,

$$\gamma(n, l(\theta_0')) = 1 + \frac{l(\theta_0')}{2n} = 1 + O_p(n^{-1}). \tag{6.3}$$

Replacing $(\theta_0' - \tilde{\theta})$ in (6.2) with $(\theta_0' - \theta_0 + \theta_0 - \tilde{\theta})$, we obtain

$$[1 - \gamma(n, l(\theta_0'))](\theta_0 - \tilde{\theta}) = \gamma(n, l(\theta_0'))(\theta_0' - \theta_0). \tag{6.4}$$

By $\tilde{\theta} - \theta_0 = O_p(n^{-1/2})$, (6.3) and (6.4), we have $\theta_0' - \theta_0 = O_p(n^{-3/2})$. □

It may be verified using the same steps in the above proof that if the expansion factor $\gamma(n, l(\theta))$ in (3.4) is replaced with a more general $\gamma(n, l(\theta)) = 1 + O_p(n^{-m})$ where $m > 0$, then Lemma 3.2 (i) still holds and (ii) becomes $\theta_0' - \theta_0 = O_p(n^{-m-1/2})$. In particular, under expansion factor $\gamma_s(n, l(\theta))$ in (3.11), we also have $\theta_0' - \theta_0 = O_p(n^{-3/2})$.

Proof of Theorem 3.2. Differentiating both sides of (2.2), we obtain $\partial l(\theta)/\partial\theta = -2n\lambda^T$. Applying Taylor expansion to $l^*(\theta_0)$ in (3.8), we have

$$l^*(\theta_0) = l(\theta_0 + (\theta_0' - \theta_0)) = l(\theta_0) - 2n\lambda(\theta_0)^T(\theta_0' - \theta_0) + o_p(\|(\theta_0' - \theta_0)\|). \tag{6.5}$$

By Owen (1990), $\lambda(\theta_0) = O_p(n^{-1/2})$. Also, by (ii) in Lemma 3.2, $\theta_0' - \theta_0 = o_p(n^{-1/2})$. These and (6.5) imply that $l^*(\theta_0) = l(\theta_0) + o_p(1)$, which together with (2.4), imply Theorem 3.2. □

The above proof requires only $\theta_0' - \theta_0 = o_p(n^{-1/2})$. This and the remark after the proof of Lemma 3.2 imply that Theorem 3.2 holds for any EEL defined by expansion factor (3.10). Also, $\theta_0' - \theta_0 = O_p(n^{-3/2})$. Plugging this into (6.5), we obtain a stronger result that $l^*(\theta_0) = l(\theta_0) + O_p(n^{-1})$.

Proof of Theorem 3.3. First note that $\gamma_s(n, l(\theta))$ in (3.11) satisfies conditions (3.5) and (3.6). Thus it may be verified that Theorem 3.1, Lemma 3.2 and Theorem 3.2 hold under the composite similarity mapping given by $\gamma_s(n, l(\theta))$. In particular, the EEL corresponding to this composite similarity mapping, $l_s^*(\theta_0)$, converges in distribution to a χ_d^2 random variable.

Since $\delta(n) = O(n^{-1/2})$ and $l(\theta_0') = l_s^*(\theta_0) = O_p(1)$, we have

$$[l(\theta_0')]^{\delta(n)} = 1 + O_p(n^{-1/2}). \tag{6.6}$$

With expansion factor $\gamma_s(n, l(\theta))$ in (3.11), equation (6.2) becomes

$$\theta_0 - \tilde{\theta} = \gamma_s(n, l(\theta_0'))(\theta_0' - \tilde{\theta}).$$

This implies

$$\theta_0' - \theta_0 = \frac{b[l(\theta_0')]^{\delta(n)}}{2n}(\tilde{\theta} - \theta_0') = \frac{b[l(\theta_0')]^{\delta(n)}}{2n}(\tilde{\theta} - \theta_0) + \frac{b[l(\theta_0')]^{\delta(n)}}{2n}(\theta_0 - \theta_0'). \tag{6.7}$$

It follows from (6.6), (6.7) and $\theta_0' - \theta_0 = O_p(n^{-3/2})$ that

$$\theta_0' - \theta_0 = \frac{b[l(\theta_0')]^{\delta(n)}}{2n}(\tilde{\theta} - \theta_0) + O_p(n^{-5/2}) = \frac{b}{2n}(\tilde{\theta} - \theta_0) + O_p(n^{-2}). \tag{6.8}$$

By assumptions (A_1) and (A_2), the OEL $l(\theta_0)$ has expansion

$$l(\theta_0) = n(\tilde{\theta} - \theta_0)^T \Sigma_0^{-1} (\tilde{\theta} - \theta_0) + O_p(n^{-1/2}), \tag{6.9}$$

and the Lagrange multipliers λ at θ_0 can be written as

$$\lambda = \lambda(\theta_0) = \Sigma_0^{-1}(\tilde{\theta} - \theta_0) + O_p(n^{-1}). \tag{6.10}$$

See Hall and La Scala (1990) and DiCiccio, Hall and Romano (1991). By (6.8), (6.10) and the Taylor expansion (6.5), we have

$$\begin{aligned} l_s^*(\theta_0) &= l(\theta_0 + (\theta_0' - \theta_0)) = l(\theta_0) - 2n\lambda(\theta_0)^T(\theta_0' - \theta_0) + o_p(\|(\theta_0' - \theta_0)\|) \\ &= l(\theta_0) - 2n\lambda(\theta_0)^T \frac{b}{2n}(\tilde{\theta} - \theta_0) + O_p(n^{-3/2}). \end{aligned} \tag{6.11}$$

It follows from (6.9), (6.10), (6.11) and $\tilde{\theta} - \theta_0 = O_p(n^{-1/2})$ that

$$\begin{aligned} l_s^*(\theta_0) &= l(\theta_0) - \frac{b}{n}\left\{ n[(\tilde{\theta} - \theta_0)^T \Sigma_0^{-1} + O_p(n^{-1})](\tilde{\theta} - \theta_0) \right\} + O_p(n^{-3/2}) \\ &= l(\theta_0)[1 - bn^{-1} + O_p(n^{-3/2})]. \end{aligned} \tag{6.12}$$

This proves (3.12) and shows that $l_s^*(\theta_0)$ is equivalent to the BEL in the left-hand side of (2.9). Finally, (3.13) follows from (3.12) and (2.9). □

Proof of Corollary 3.1. It is convenient to view the expansion factor of the first order EEL as a special case of that for the second order EEL (3.11) where $b = 1$ and $\delta(n) = 1$. The condition of $\delta(n) = O(n^{-1/2})$ imposed on the $\delta(n)$ in (3.11) is not needed here. Noting that $l^*(\theta_0) = l(\theta_0) + O_p(n^{-1})$ and $l(\theta_0') = l^*(\theta_0)$, equation (6.6) in the proof Theorem 3.3 is now

$$l(\theta_0') = l(\theta_0) + O_p(n^{-1}).$$

Thus equation (6.8) becomes

$$\theta_0' - \theta_0 = \frac{l(\theta_0)}{2n}(\tilde{\theta} - \theta_0) + O_p(n^{-5/2}).$$

Using the above equation and following the steps given by (6.9) to (6.12), we obtain Corollary 3.1. □

Acknowledgements. We thank two anonymous referees and an Associate Editor for their valuable comments which have led to improvements in this paper. We also thank the Editor, Professor Peter Hall, and the Associate Editor for their exceptionally quick handling of this manuscript.

References.

[1] Bartolucci, F. (2007). A penalized version of the empirical likelihood ratio for the population mean. *Statist. Probab. Lett.*, **77**, 104–110. MR2339024

[2] Chen, J. and Huang, Y. (2012). Finite-sample properties of the adjusted empirical likelihood. *J. Nonparametric Statistics* **25** 147–159. MR2439967

[3] Chen, J., Variyath, A. M. and Abraham, B. (2008). Adjusted empirical likelihood and its properties. *J. Comput. Graph. Statist.* **3** 426–443. MR2439967

[4] Chen, S. X. and Cui, H. J. (2007). On the second-order properties of empirical likelihood with moment restrictions. *J. Econometrics*, 141, 492–516. MR2413478

[5] Corcoran, S. A., Davison, A. C. and Spady, R. H. (1995). Reliable inference from empirical likelihood. *Ecnomics Working Paper 10.* Nuffield College, University of Oxford.

[6] DiCiccio, T. J., Hall, P. and Romano, J. P. (1988). Bartlett adjustment for empirical likelihood. Technical Report 298, Dept. Statistics, Stanford Univ. **19** 1053–1061. MR1105861

[7] DiCiccio, T. J., Hall, P. and Romano, J. P. (1991). Empirical likelihood is Bartlett Correctable. *Ann. Statist.* **19** 1053–1061. MR1105861

[8] Emerson, S. C. and Owen, A. B. (2009). Calibration of the empirical likelihood method for a vector mean. *Electron. J. Statist.* **3** 1161–1192. MR2566185

[9] Hall, P. and La Scala, B. (1990). Methodology and algorithm of empirical likelihood. *Internat. Statist. Rev.* **58** 109–127.

[10] Lahiri, S. N. and Mukhopadhyay, S. (2012). A penalized empirical likelihood method in high dimensions. *Ann. Statist.*, **40** 2511-2540.

[11] Liu, Y. and Chen, J. (2010). Adjusted empirical likelihood with high-order precision. *Ann. Statist.* **38** 1341–1362. MR2662345

[12] Owen, A. B. (1988). Empirical likelihood ratio confidence regions for single functional. *Biometrika.* **75**, 237-249. MR0946049

[13] Owen, A. B. (1990). Empirical likelihood confidence regions. *Ann. Statist.* **18** 90–120. MR1041387

[14] Owen, A. B. (2001). *Empirical Likelihood*, Chapman & Hall/CRC, London. MR1700749

[15] Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22** 300–325. MR1272085

[16] Tsao, M. (2004). Bounds on coverage probabilities of the empirical likelihood ratio confidence regions. *Ann. Statist.* **32** 1215–1221. MR2065203

[17] Tsao, M. (2013). Extending the empirical likelihood by domain expansion. To appear in the *Canadian Journal of Statistics*.

[18] Tsao, M. and Wu, F. (2013). Extended empirical likelihood for parameters defined by estimating equations. Technical report (in preparation), Dept. of Math. & Stat., Univ. of Victoria, Canada, V8W 3R4.

[19] Wu, F. (2013). *The Extended Empirical Likelihood*. Ph.D. Dissertation, in progress; Dept. of Math. & Stat., Univ. of Victoria, Canada, V8W 3R4.

Department of Mathematics and Statistics
University of Victoria
Victoria, British Columbia
Canada V8W 3R4
E-mail: mtsao@uvic.ca
fwu@uvic.ca

Table 1: Simulated coverage probabilities for one-dimensional examples

	n	Level	OEL	$\mathbf{EEL}_1$	BEL	AEL	EEL_2	BEL*	AEL*	EEL_2^*
$N(0,1)$	10	.90	.8506	**.8914**	.8753	.8788	.8813	.8767	.8867	.8824
		.95	.9039	**.9452**	.9246	.9294	.9317	.9242	.9352	.9324
		.99	.9580	**.9867**	.9677	.9753	.9738	.9656	.9771	.9734
	30	.90	.8920	**.9071**	.9007	.9008	.9022	.9017	.9019	.9030
		.95	.9398	**.9548**	.9461	.9461	.9476	.9466	.9468	.9474
		.99	.9866	**.9925**	.9882	.9883	.9885	.9883	.9884	.9884
	50	.90	.8941	**.9024**	.8995	.8996	.9003	.8992	.8993	.9000
		.95	.9447	**.9541**	.9481	.9479	.9486	.9483	.9484	.9490
		.99	.9880	**.9920**	.9892	.9892	.9892	.9894	.9894	.9895
t_5	10	.90	.8277	**.8765**	.9226	.9979	.9209	.8520	.8873	.8782
		.95	.8882	**.9394**	.9599	1.000	.9651	.9036	.9367	.9307
		.99	.9556	**.9851**	.9820	1.000	.9887	.9499	.9798	.9751
	30	.90	.8690	**.8852**	.8999	.9028	.9017	.8852	.8885	.8882
		.95	.9265	**.9436**	.9476	.9509	.9502	.9385	.9428	.9420
		.99	.9797	**.9888**	.9875	.9901	.9886	.9831	.9863	.9861
	50	.90	.8862	**.8967**	.9040	.9048	.9052	.8977	.8983	.8987
		.95	.9410	**.9491**	.9515	.9518	.9518	.9465	.9471	.9474
		.99	.9861	**.9918**	.9907	.9913	.9913	.9881	.9882	.9886
χ_1^2	10	.90	.7764	**.8174**	.8726	1.000	.8634	.6792	.8456	.8291
		.95	.8314	**.8781**	.9068	1.000	.9030	.7239	.8918	.8779
		.99	.8973	**.9378**	.9417	1.000	.9461	.7677	.9391	.9253
	30	.90	.8594	**.8759**	.8887	.8901	.8890	.8658	.8847	.8829
		.95	.9115	**.9249**	.9319	.9343	.9330	.9105	.9278	.9272
		.99	.9659	**.9764**	.9759	.9786	.9769	.9565	.9735	.9733
	50	.90	.8722	**.8833**	.8936	.8941	.8943	.8887	.8912	.8909
		.95	.9318	**.9411**	.9441	.9458	.9459	.9388	.9419	.9415
		.99	.9779	**.9847**	.9837	.9845	.9845	.9804	.9831	.9830
$.3N(0,1)+.7N(2,1)$	10	.90	.8470	**.8908**	.8551	.8556	.8569	.8761	.8826	.8821
		.95	.9036	**.9433**	.9094	.9097	.9127	.9215	.9299	.9285
		.99	.9564	**.9867**	.9592	.9601	.9631	.9657	.9760	.9741
	30	.90	.8930	**.9054**	.8956	.8956	.8960	.9016	.9013	.9017
		.95	.9438	**.9582**	.9455	.9455	.9460	.9501	.9501	.9507
		.99	.9873	**.9943**	.9883	.9883	.9884	.9901	.9901	.9909
	50	.90	.8965	**.9048**	.8989	.8988	.8990	.9014	.9014	.9016
		.95	.9465	**.9556**	.9475	.9476	.9477	.9494	.9496	.9499
		.99	.9876	**.9911**	.9879	.9877	.9879	.9883	.9883	.9886

Table 2: Simulated coverage probabilities for two-dimensional examples

	n	Level	OEL	**EEL$_1$**	BEL	AEL	EEL$_2$	BEL*	AEL*	EEL*$_2$
$BV1$	10	.90	.7134	**.8118**	.7965	.9777	.8212	.7561	.8350	.7989
		.95	.7717	**.8758**	.8407	1.000	.8718	.8000	.8941	.8478
		.99	.8484	**.9422**	.8945	1.000	.9268	.8570	.9680	.9083
	30	.90	.8549	**.8888**	.8824	.8856	.8872	.8786	.8813	.8822
		.95	.9120	**.9426**	.9313	.9348	.9361	.9296	.9336	.9337
		.99	.9689	**.9868**	.9772	.9798	.9796	.9757	.9783	.9779
	50	.90	.8699	**.8917**	.8869	.8874	.8894	.8859	.8869	.8886
		.95	.9259	**.9428**	.9354	.9361	.9374	.9344	.9351	.9363
		.99	.9806	**.9908**	.9846	.9852	.9856	.9839	.9848	.9851
$BV2$	10	.90	.7513	**.8521**	.8035	.8573	.8282	.7942	.8451	.8229
		.95	.8061	**.9095**	.8627	.9397	.8861	.8499	.9103	.8833
		.99	.8879	**.9693**	.9202	1.000	.9430	.9116	.9721	.9405
	30	.90	.8714	**.9019**	.8864	.8872	.8897	.8864	.8881	.8906
		.95	.9256	**.9549**	.9406	.9413	.9428	.9403	.9409	.9432
		.99	.9789	**.9907**	.9823	.9826	.9838	.9820	.9829	.9836
	50	.90	.8826	**.9037**	.8935	.8937	.8954	.8938	.8939	.8952
		.95	.9348	**.9528**	.9423	.9426	.9438	.9423	.9425	.9435
		.99	.9839	**.9914**	.9862	.9864	.9871	.9861	.9864	.9864
$BV3$	10	.90	.7001	**.7979**	.7979	1.000	.8162	.7333	.8363	.7911
		.95	.7608	**.8581**	.8374	1.000	.8624	.7765	.8922	.8375
		.99	.8331	**.9263**	.8817	1.000	.9151	.8286	.9639	.8942
	30	.90	.8429	**.8775**	.8749	.8789	.8788	.8719	.8780	.8764
		.95	.9015	**.9363**	.9266	.9326	.9319	.9221	.9282	.9280
		.99	.9648	**.9817**	.9740	.9776	.9760	.9709	.9750	.9740
	50	.90	.8619	**.8836**	.8807	.8820	.8836	.8787	.8801	.8808
		.95	.9212	**.9403**	.9351	.9364	.9379	.9325	.9346	.9347
		.99	.9758	**.9848**	.9810	.9816	.9817	.9802	.9806	.9810
$BV4$	10	.90	.6408	**.7371**	.7882	1.000	.7940	.6240	.8382	.7596
		.95	.7030	**.8027**	.8212	1.000	.8377	.6637	.8896	.8051
		.99	.7788	**.8808**	.8580	1.000	.8914	.7129	.9576	.8602
	30	.90	.8229	**.8595**	.8709	.8820	.8760	.8598	.8738	.8681
		.95	.8857	**.9191**	.9215	.9329	.9255	.9079	.9212	.9170
		.99	.9520	**.9734**	.9689	.9819	.9717	.9591	.9696	.9667
	50	.90	.8494	**.8707**	.8758	.8783	.8783	.8716	.8755	.8740
		.95	.9060	**.9251**	.9256	.9287	.9282	.9221	.9259	.9251
		.99	.9675	**.9807**	.9781	.9797	.9793	.9753	.9778	.9765

Table 3: Simulated coverage probabilities for the mean of d-dimensional multivariate normal distributions

	n	Level	OEL	**EEL$_1$**	BEL	AEL	EEL$_2$	BEL*	AEL*	EEL$_2^*$
$d = 5$	10	.90	.3007	**.5897**	.3839	1.000	.5306	.3691	1.000	.5005
		.95	.3368	**.6794**	.4135	1.000	.5842	.4028	1.000	.5500
		.99	.3946	**.7984**	.4498	1.000	.6642	.4422	1.000	.6287
	30	.90	.7790	**.8862**	.8258	.8554	.8455	.8273	.8585	.8468
		.95	.8497	**.9436**	.8880	.9208	.9047	.8884	.9256	.9052
		.99	.9337	**.9881**	.9532	.9889	.9629	.9539	.9899	.9634
	50	.90	.8476	**.9036**	.8752	.8803	.8820	.8757	.8808	.8825
		.95	.9089	**.9522**	.9297	.9341	.9349	.9297	.9349	.9354
		.99	.9728	**.9913**	.9804	.9833	.9830	.9804	.9839	.9831
$d = 10$	20	.90	.1889	**.5367**	.2845	1.000	.4297	.2823	1.000	.4235
		.95	.2281	**.6260**	.3209	1.000	.4905	.3191	1.000	.4824
		.99	.2895	**.7708**	.3747	1.000	.5783	.3727	1.000	.5717
	30	.90	.4689	**.7752**	.5944	1.000	.6750	.5954	1.000	.6752
		.95	.5432	**.8594**	.6627	1.000	.7492	.6635	1.000	.7480
		.99	.6698	**.9442**	.7670	1.000	.8514	.7675	1.000	.8527
	50	.90	.7097	**.8806**	.7921	.9531	.8189	.7933	.9582	.8198
		.95	.7959	**.9393**	.8577	.9968	.8827	.8588	.9974	.8838
		.99	.9027	**.9864**	.9392	1.000	.9546	.9396	1.000	.9549
$d = 15$	30	.90	.1224	**.4850**	.2199	1.000	.3581	.2196	1.000	.3569
		.95	.1513	**.5761**	.2504	1.000	.4130	.2502	1.000	.4124
		.99	.2155	**.7490**	.3100	1.000	.5077	.3099	1.000	.5054
	50	.90	.4769	**.7983**	.6177	1.000	.6883	.6191	1.000	.6894
		.95	.5665	**.8776**	.6971	1.000	.7630	.6985	1.000	.7646
		.99	.7065	**.9600**	.8097	1.000	.8682	.8103	1.000	.8686
	100	.90	.7696	**.9031**	.8325	.9309	.8472	.8328	.9341	.8484
		.95	.8484	**.9514**	.8985	.9852	.9086	.8989	.9865	.9096
		.99	.9405	**.9900**	.9639	1.000	.9693	.9641	1.000	.9692