THE EXTENDED EMPIRICAL LIKELIHOOD

by

Fan Wu
BA., University of Western Ontario, 2005
M.Sc., University of Victoria 2008

A Dissertation Submitted in Partial Fulfillment of the
Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Mathematics and Statistics

THE EXTENDED EMPIRICAL LIKELIHOOD

by

Fan Wu
BA., University of Western Ontario, 2005
M.Sc., University of Victoria 2008

Supervisory Committee

Dr. Min Tsao, Supervisor
(Department of Mathematics and Statistics)

Dr. Mary Lesperance, Departmental Member
(Department of Mathematics and Statistics)

Dr. Farouk Nathoo, Departmental Member
(Department of Mathematics and Statistics)

Dr. David Giles, Outside Member
(Department of Economics)

Supervisory Committee

Dr. Min Tsao, Supervisor
(Department of Mathematics and Statistics)

Dr. Mary Lesperance, Departmental Member
(Department of Mathematics and Statistics)

Dr. Farouk Nathoo, Departmental Member
(Department of Mathematics and Statistics)

Dr. David Giles, Outside Member
(Department of Economics)

ABSTRACT

The empirical likelihood method introduced by Owen (1988, 1990) is a powerful nonparametric method for statistical inference. It has been one of the most researched methods in statistics in the last twenty-five years and remains to be a very active area of research today. There is now a large body of literature on empirical likelihood method which covers its applications in many areas of statistics (Owen, 2001).

One important problem affecting the empirical likelihood method is its poor accuracy, especially for small sample and/or high-dimension applications. The poor accuracy can be alleviated by using high-order empirical likelihood methods such as the Bartlett corrected empirical likelihood but it cannot be completely resolved by high-order asymptotic methods alone. Since the work of Tsao (2004), the impact of the convex hull constraint in the formulation of the empirical likelihood on the finite-sample accuracy has been better understood, and methods have been developed to break this constraint in order to improve the accuracy. Three important methods along this direction are [1] the penalized empirical likelihood of Bartolucci (2007) and Lahiri and Mukhopadhyay (2012), [2] the adjusted empirical likelihood by Chen, Variyath and Abraham (2008), Emerson and Owen (2009), Liu and Chen (2010) and Chen and Huang (2012), and [3] the extended empirical likelihood of Tsao (2013) and Tsao and Wu (2013). The latter is particularly attractive in that it retains not only

the asymptotic properties of the original empirical likelihood, but also its important geometric characteristics. In this thesis, we generalize the extended empirical likelihood of Tsao and Wu (2013) to handle inferences in two large classes of one-sample and two-sample problems.

In Chapter 2, we generalize the extended empirical likelihood to handle inference for the large class of parameters defined by one-sample estimating equations, which includes the mean as a special case. In Chapters 3 and 4, we generalize the extended empirical likelihood to handle two-sample problems; in Chapter 3, we study the extended empirical likelihood for the difference between two p-dimensional means; in Chapter 4, we consider the extended empirical likelihood for the difference between two p-dimensional parameters defined by estimating equations. In all cases, we give both the first- and second-order extended empirical likelihood methods and compare these methods with existing methods. Technically, the two-sample mean problem in Chapter 3 is a special case of the general two-sample problem in Chapter 4. We single out the mean case to form Chapter 3 not only because it is a standalone published work, but also because it naturally leads up to the more difficult two-sample estimating equations problem in Chapter 4.

We note that Chapter 2 is the published paper Tsao and Wu (2014); Chapter 3 is the published paper Wu and Tsao (2014). To comply with the University of Victoria policy regarding the use of published work for thesis and in accordance with copyright agreements between authors and journal publishers, details of these published work are acknowledged at the beginning of these chapters. Chapter 4 is another joint paper Tsao and Wu (2015) which has been submitted for publication.

Contents

List of Tables

List of Figures

ACKNOWLEDGEMENTS

I would like to thank my supervisor Professor Tsao for his valuable advice. I have been extremely lucky to have Professor Tsao as my supervisor. He has guided me in the fields of non-parametric statistics and provided many ideas, and offered infinite help and patience during my Ph.D. study at the University of Victoria. Without his guidance and persistent help this dissertation would not have been possible.

DEDICATION

I must express my gratitude to Hong Li, my wife, for her unconditional love and continued support throughout the course of this dissertation.

Chapter 1

Introduction

The empirical likelihood method, first introduced by Owen (1988, 1990), is a powerful nonparametric method for statistical inference. Like the bootstrap and jackknife methods, it does not require strong distributional assumptions. It produces confidence regions which reflect the shape of the data without the need for a pivotal quantity, and it yields efficient non-parametric maximum likelihood estimates that make use of side information. In the last twenty-five years, the empirical likelihood method has found applications in virtually every area of statistical research (Owen, 2001). Today, it remains to be one of the most active areas of statistical research.

Since the early development of the empirical likelihood method, it has been widely observed that the empirical likelihood confidence regions tend to have poor coverage accuracy. In particular, there is an undercoverage problem in that the coverage probability of an empirical likelihood ratio confidence region tends to be lower than the nominal level; see, *e.g.,* Hall and La Scala (1990), Qin and Lawless (1994), Corcoran, Davison and Spady (1995), Owen (2001) and Liu and Chen (2010). The objective of this thesis is to improve the accuracy of the empirical likelihood inference for a large class of parameters defined by estimating equations where the poor accuracy problem is particularly serious and well-known. We tackle this problem in one-sample and two-sample sittings by generalizing the extended empirical likelihood of Tsao (2013) and Tsao and Wu (2013). Our main results are a first-order and a second-order extended empirical likelihood methods for such parameters. We show through simulation studies these new methods are substantially more accurate than the original empirical likelihood method.

The first important work addressing the accuracy issue of the empirical likelihood method is DiCiccio, Hall and Romano (1991) which showed that the empirical likelihood is Bartlett correctable. The Bartlett corrected empirical likelihood has the second-order accuracy, and the empirical likelihood is the only non-parametric likelihood that has been found to be Bartlett correctable. This surprising result added considerable theoretical appeal to the method of empirical likelihood. Although the Bartlett correction is an asymptotic technique, it leads to considerably more accurate empirical likelihood inference in finite-sample applications. Nevertheless, the undercoverage problem remains unresolved; the Bartlett corrected empirical likelihood also suffers from the undercoverage problem, albeit to a lesser degree. Further, the Bartlett correction is not always easy to compute.

Tsao (2004) approached the undercoverage issue from a finite-sample standpoint. He studied the finite-sample least-upper bound on the coverage probability of the empirical likelihood ratio confidence region which are the consequence of the convex hull constraint embedded in the formulation of the empirical likelihood. He derived the bounds for the large class of problems where the parameters of interest are defined by one-sample estimating equations. For small sample and/or high dimension situations, the bounds can be much lower than one. This suggests that the convex hull constraint is a main contributor to the undercoverage problem.

Since the work of Tsao (2004), various methods aimed at solving the undercoverage problem by breaking the convex hull constraint have been developed. Bartolucci

(2007) introduced a penalized empirical likelihood for the mean which removes the constraint from the formulation of the original empirical likelihood of Owen (1990, 2001) and replaces it with a penalizing term based on the Mahalanobis distance. Chen, Variyath and Abraham (2008) introduced an adjusted empirical likelihood which retains the formulation of original empirical likelihood but adds a pseudo-observation to the sample. The adjusted empirical likelihood is just the original empirical likelihood defined on the augmented sample, but due to the clever construction of the pseudo-observation the convex hull constraint will never be violated here. Emerson and Owen (2009) showed that the adjusted empirical likelihood statistic has a boundedness problem which may lead to trivial 100% confidence regions. They proposed an extension of the adjusted empirical likelihood 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 adjusted empirical likelihood confidence region achieves the second order accuracy of the Bartlett correction. Recently, Lahiri and Mukhopadhyay (2012) showed that under certain dependence structures, a modified penalized empirical likelihood for the mean works in the extremely difficult case of large dimension and small sample size.

Nevertheless, the penalized empirical likelihood is only available for the mean and it may be difficult to implement. The adjusted empirical likelihood is available for the large class of parameters defined by estimating equations, but the aforementioned boundedness problem requires more attention. More importantly, for both methods, the shape of their confidence regions no longer follow that of the original empirical likelihood region. Hence, they lose a celebrated advantage of the empirical likelihood method, that is, the shape of its confidence region reflects the shape of the data.

To deal with the undercoverage problem caused by the convex hull constraint while still keeping the shape of the empirical likelihood confidence regions, Tsao (2013) introduced a geometric approach to break the convex hull constraint by geometrically expanding the domain of the original empirical likelihood ratio. The empirical likelihood defined on this larger expanded domain is referred to as the extended empirical likelihood. With a large domain, the extended empirical likelihood produces larger confidence regions and hence more accurate coverage probabilities. Tsao and Wu (2013) made a significant step forward with this domain expansion idea where they derived an extended empirical likelihood ratio for the mean defined on the entire parameter space. The key technique that they developed to construct such an extended empirical likelihood is a composite similarity transformation which consists of a continuous sequence of simple similarity mappings of the original empirical likelihood ratio contours. This extended empirical likelihood for the mean is theoretically simple and appealing, and numerically substantially more accurate than the original empirical likelihood.

In this thesis, we generalize the extended empirical likelihood of Tsao (2013) and Tsao and Wu (2013) to improve the accuracy of the empirical likelihood inference in two directions. In Chapter 2, we generalize the extended empirical likelihood to handle inference for the large class of parameters defined by one-sample estimating equations, which includes the mean as a special case. In Chapters 3 and 4, we generalize the extended empirical likelihood to handle two-sample problems; in Chapter 3, we study the extended empirical likelihood for the difference between two p-dimensional means; in Chapter 4, we consider the extended empirical likelihood for the difference between two p-dimensional parameters defined by estimating equations. In all cases, we give both the first- and second-order extended empirical likelihood methods and compare these methods with existing methods.

It should be noted that Chapter 2 is the published paper Tsao and Wu (2014) and Chapter 3 is the published paper Wu and Tsao (2014); a detailed acknowledgement to this effect may be found at the beginning of these two chapters. Chapter 4 is also a joint paper Tsao and Wu (2015) which has been submitted for publication. At the request of the Thesis Supervisory Committee, we hereby acknowledge that Fan Wu is the principal author (defined as the co-author who is responsible for 60% or more of a joint paper's contents) for Wu and Tsao (2014), Min Tsao is the principal author for Tsao and Wu (2014), and both authors contributed roughly equally to Tsao and Wu (2015).

Chapter 2

Extended empirical likelihood for estimating equations

Acknowledgement: in accordance with copyright agreements between authors and journal publishers, we acknowledge that this chapter is the published paper under the same title by Tsao and Wu (2014), Biometrika, volume 101, issue 3, pages 703-710, with a 12-page Supplement Material available at Biometrika online.

2.1 Introduction

One important application of the empirical likelihood (Owen, 2001) is for inference on parameters defined by estimating equations that satisfy $E\{g(X, \theta_0)\} = 0$, where $g(x, \theta) \in \mathbb{R}^q$ is an estimating function for the parameter vector $\theta_0 \in \mathbb{R}^p$ of a random vector $X \in \mathbb{R}^d$ (Qin and Lawless, 1994). The estimating equations are said to be just-determined if $q = p$ and over-determined if $q > p$. The latter case arises when extra information about the parameter is available and results in an estimating function of dimension $q > p$. In principle, extra information should increase the precision of the inference. However, Qin and Lawless (1994) observed that empirical likelihood confidence regions for over-determined cases can have substantial undercoverage.

The poor accuracy of empirical likelihood confidence regions has also been noted by others, e.g., Hall and La Scala (1990), Owen (2001), Tsao (2004) and Chen, Variyath and Abraham (2008). In particular, as shown in a 1995 Nuffield College, Oxford, working paper by S. A. Corcoran, A. C. Davison and R. H. Spady, the second-order empirical likelihood method also has poor accuracy. This suggests that the principal cause of the poor accuracy is not the asymptotic orders of the methods. The main culprit turns out to be the mismatch between the domain of the empirical likelihood and the parameter space (Tsao, 2013; Tsao and Wu, 2013); whereas the parameter space is in general the whole of $\mathbb{R}^p$, the domain is usually a bounded subset of $\mathbb{R}^p$. This mismatch is a consequence of a convex hull constraint embedded in the formulation of the empirical likelihood; values of $\theta \in \mathbb{R}^p$ that violate this constraint are excluded from the domain, leading to the mismatch. Three variants of the original empirical likelihood of Owen (1988, 1990) tackle the convex hull constraint in different ways: the penalized empirical likelihood of Bartolucci (2007) and Lahiri and Mukhopadhyay (2012); the adjusted empirical likelihood by Chen, Variyath and Abraham (2008), Emerson and Owen (2009), Liu and Chen (2010) and Chen and Huang (2012); and the extended empirical likelihood of Tsao (2013) and Tsao and Wu (2013). The first replaces the convex hull constraint in the original empirical likelihood with a penalizing term based on the Mahalanobis distance. The second adds one or two pseudo-observations to the sample to ensure that the convex hull constraint is not violated. The third expands the domain of the original empirical likelihood geometrically to overcome the constraint. The adjusted empirical likelihood is available for parameters defined by estimating equations. The penalized and extended empirical likelihoods on $\mathbb{R}^p$ are available only for the mean. All three variants have the same asymptotic distribution as the original empirical likelihood, but the extended empirical likelihood is a more natural generalization because its contours have the same shape. The data-driven shape of the original empirical likelihood contours is a celebrated advantage, which is retained by the extended version.

In this paper, we generalize the results of Tsao and Wu (2013) for the mean to an extended empirical likelihood on $\mathbb{R}^p$ for the large collection of parameters defined by estimating equations. Under certain conditions, this new likelihood has the same asymptotic properties and identically shaped contours as the original one, and can attain the second-order accuracy of the Bartlett corrected likelihood ratio statistic of DiCiccio, Hall and Romano (1991) and Chen and Cui (2007). We highlight the first-order version of this extended empirical likelihood, which is not only easy-to-use but also much more accurate than the original version and available second-order methods. Because of its simplicity and accuracy, we recommend it to practitioners. A secondary objective of this paper is to provide details of techniques for deriving the extended empirical likelihood on $\mathbb{R}^p$ that may be applied to parameters beyond the standard estimating equations framework. Throughout this paper, we use $l(\theta)$ and $l^*(\theta)$ to denote the original and extended empirical log-likelihood ratios.

2.2 Extended empirical likelihood for estimating equations

2.2.1 Preliminaries

Let $X \in \mathbb{R}^d$ be a random vector with a parameter $\theta_0 \in \mathbb{R}^p$, let $g(X, \theta)$ be a q-dimensional estimating function for θ_0 and let $X_1, \ldots, X_n$ be independent copies of X, where the sample size $n > q$. We will need the following conditions on $g(X, \theta)$:

Condition 1. $E\{g(X, \theta_0)\} = 0$ and $\text{var}\{g(X, \theta_0)\} \in \mathbb{R}^{q \times q}$ is positive definite;

Condition 2. $\partial g(X,\theta)/\partial\theta$ and $\partial g^2(X,\theta)/\partial\theta\partial\theta^T$ are continuous in θ, and for θ in a neighbourhood of θ_0 they are each bounded in norm by an integrable function of X;

Condition 3. $\limsup_{\|t\|\to\infty} |E[\exp\{it^T g(X,\theta_0)\}]| < 1$ and $E\{\|g(X,\theta_0)\|^{15}\} < +\infty$.

These conditions ensure that the original empirical likelihood for estimating equations is Bartlett correctable (Chen and Cui, 2007). The empirical likelihood ratio for $\theta \in \mathbb{R}^p$ is

$$R(\theta) = \sup\left\{\prod_{i=1}^n nw_i : \sum_{i=1}^n w_i g(X_i,\theta) = 0, w_i \geq 0, \sum_{i=1}^n w_i = 1\right\},$$

where 0 denotes the origin in $\mathbb{R}^q$ (Owen, 2001). The original empirical log-likelihood ratio is $l(\theta) = -2\log R(\theta)$. An alternative to $l(\theta)$ is the statistic defined as $W_E(\theta) = l(\theta) - l(\tilde{\theta})$ in equation (3.9) of Qin and Lawless (1994), where $\tilde{\theta}$ is the maximum empirical likelihood estimator of θ_0. We will consider an extended empirical likelihood based on $W_E(\theta)$ in the Supplementary Material. Let $\bar{w} = (w_1,\ldots,w_n)$ denote a weight vector, with $w_i > 0$ and $\sum_{i=1}^n w_i = 1$. Define the common domain Θ_n of $R(\theta)$ and $l(\theta)$ as

$$\Theta_n = \{\theta :\ \theta \in \mathbb{R}^p \text{ and there exists a } \bar{w} \text{ such that } \textstyle\sum_{i=1}^n w_i g(X_i,\theta) = 0\}. \tag{2.1}$$

Then, Θ_n is the collection of all θ values satisfying $l(\theta) < +\infty$. Throughout this paper, we assume without loss of generality that Θ_n is a non-empty open set in $\mathbb{R}^p$. See the Appendix. For $\theta \in \Theta_n$, using the method of Lagrange multipliers we can show that

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

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

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

Under Condition 1, Qin and Lawless (1994) showed that $l(\theta_0)$ converges in distribution to a χ_q^2 random variable as n goes to infinity. Thus, the $100(1-\alpha)\%$ original empirical likelihood confidence region for θ_0 is

$$\mathcal{C}_{1-\alpha} = \{\theta : \theta \in \Theta_n, l(\theta) \leq c\}, \tag{2.3}$$

where c is the $(1-\alpha)$th quantile of the χ_q^2 distribution. The coverage error of $\mathcal{C}_{1-\alpha}$ is

$$\mathrm{pr}(\theta_0 \in \mathcal{C}_{1-\alpha}) = \mathrm{pr}\{l(\theta_0) \leq c\} = \mathrm{pr}(\chi_q^2 \leq c) + O(n^{-1}). \tag{2.4}$$

We now briefly review the Bartlett correction of DiCiccio, Hall and Romano (1991) and Chen and Cui (2007) for $l(\theta)$. Under Conditions 1, 2 and 3,

$$l(\theta_0) = nR^T R + O_p(n^{-3/2}), \tag{2.5}$$

where R is a q-dimensional vector which is a smooth function of general means. Through an Edgeworth expansion for the density function of $n^{1/2}R$, we can show that

$$\mathrm{pr}[nR^T R\{1 - bn^{-1} + O_p(n^{-3/2})\} \leq c] = \mathrm{pr}(\chi_q^2 \leq c) + O(n^{-2}), \tag{2.6}$$

where b depends on the moments of $g(X, \theta_0)$ and $(1 - bn^{-1})$ is the Bartlett correction factor. It follows from (2.5) and (2.6) that

$$\mathrm{pr}[l(\theta_0)\{1 - bn^{-1} + O_p(n^{-3/2})\} \leq c] = \mathrm{pr}(\chi_q^2 \leq c) + O(n^{-2}). \tag{2.7}$$

Let $l_B(\theta) = (1 - bn^{-1})l(\theta)$ be the Bartlett corrected empirical log-likelihood ratio, and denote by $\mathcal{C}'_{1-\alpha}$ the Bartlett corrected empirical likelihood confidence region for θ_0. Then,

$$\mathcal{C}'_{1-\alpha} = \{\theta : \theta \in \Theta_n, l_B(\theta) \leq c\}. \tag{2.8}$$

Equation (2.7) implies that the coverage error of $\mathcal{C}'_{1-\alpha}$ is $O(n^{-2})$, that is,

$$\text{pr}(\theta_0 \in \mathcal{C}'_{1-\alpha}) = \text{pr}\{l_B(\theta_0) \leq c\} = \text{pr}(\chi^2_q \leq c) + O(n^{-2}).$$

2.2.2 Composite similarity mapping

The mismatch between the original empirical likelihood domain Θ_n and the parameter space $\mathbb{R}^p$ is a main cause of the poor accuracy of the original empirical likelihood confidence region (Tsao, 2013). To solve the mismatch problem, we expand Θ_n to $\mathbb{R}^p$ through a composite similarity mapping $h_n^C : \Theta_n \to \mathbb{R}^p$ (Tsao and Wu, 2013). In order to define h_n^C, we assume that there exists a $\sqrt{n}$-consistent maximum empirical likelihood estimator $\tilde{\theta}$ for θ_0. See the Appendix for more discussion of this assumption. Using $l(\theta)$ and $\tilde{\theta}$, we define

$$h_n^C(\theta) = \tilde{\theta} + \gamma\{n, l(\theta)\}(\theta - \tilde{\theta}), \quad \theta \in \Theta_n, \tag{2.9}$$

where function $\gamma\{n, l(\theta)\}$ is the expansion factor given by

$$\gamma\{n, l(\theta)\} = 1 + \frac{l(\theta)}{2n}. \tag{2.10}$$

To see how h_n^C maps Θ_n onto $\mathbb{R}^p$, define the level-τ original empirical likelihood contour as

$$c(\tau) = \{\theta : \theta \in \Theta_n, l(\theta) = \tau\}, \tag{2.11}$$

where $\tau \geq \tilde{\tau} = l(\tilde{\theta}) \geq 0$. For the just-determined case, $R(\tilde{\theta}) = 1$ and $\tilde{\tau} = l(\tilde{\theta}) = 0$. The contours form a partition of the domain Θ_n; that is, $c(\tau_1) \cap c(\tau_2) = \varnothing$ for any $\tau_1 \neq \tau_2$ and

$$\Theta_n = \bigcup_{\tau \in [\tilde{\tau}, +\infty)} c(\tau). \tag{2.12}$$

In addition to Conditions 1, 2 and 3 above, we now introduce a new condition.

Condition 4. Each contour $c(\tau)$ is the boundary of a connected region and the contours are nested in that if $\tau_1 < \tau_2$, then $c(\tau_1)$ is contained in the interior of the region defined by $c(\tau_2)$.

Under Condition 4, (2.12) implies that $c(\tilde{\tau}) = \{\tilde{\theta}\}$ is the centre of Θ_n. It follows that the value of τ measures the outwardness of a $c(\tau)$ with respect to the centre; the larger the τ value, the more outward $c(\tau)$ is. Theorem 2.1 below gives three key properties of h_n^C.

Theorem 2.1. *Under Conditions 1 and 2, the mapping h_n^C defined by (2.9) and (2.10)*

(i) has a unique fixed point at $\tilde{\theta}$,

(ii) is a similarity transformation for each individual contour $c(\tau)$, and

(iii) is a surjection from Θ_n to $\mathbb{R}^p$.

Because of (ii), we call h_n^C the composite similarity mapping, as it may be viewed as a continuous sequence of similarity mappings from $\mathbb{R}^p$ to $\mathbb{R}^p$ that are indexed by $\tau \in [\tilde{\tau}, +\infty)$. The τ-th mapping has expansion factor $\gamma\{n, l(\theta)\} = \gamma(n, \tau)$ and is used exclusively to map the level-τ contour $c(\tau)$. Since $\gamma(n, \tau)$ is an increasing function of τ, contours farther away from the centre are expanded more so that images of the contours fill up $\mathbb{R}^p$. Regardless of the amount expanded, $c(\tau)$ and its image are identical in shape; Figure 1 illustrates this with the original empirical likelihood contours for parameters of a regression model and their expanded images.

The proof of Theorem 2.1 is given in the Supplementary Material. A remark following the proof shows that if we are to add Condition 4 to Theorem 2.1, then (iii) can be strengthened to (iii') h_n^C is a bijection from Θ_n to $\mathbb{R}^p$. It is not clear how we may verify Condition 4 through $g(X,\theta)$, so we have kept it separate from the three conditions identified in the preliminaries. Nevertheless, we have not encountered any example where Condition 4 is violated.

2.2.3 Extended empirical likelihood on the full parameter space

Since $h_n^C : \Theta_n \to \mathbb{R}^d$ is surjective, for any $\theta \in \mathbb{R}^p$, $s(\theta) = \{\theta' : \theta' \in \Theta_n, h_n^C(\theta') = \theta\}$ is non-empty. When h_n^C is not injective, $s(\theta)$ may contain more than one point and h_n^C does not have an inverse. Hence, we define a generalized inverse $h_n^{-C} : \mathbb{R}^p \to \Theta_n$ as

$$h_n^{-C}(\theta) = \underset{\theta' \in s(\theta)}{\operatorname{argmin}}\{\|\theta' - \theta\|\}, \quad \theta \in \mathbb{R}^p. \tag{2.13}$$

The extended empirical log-likelihood ratio statistic $l^*(\theta)$ under h_n^{-C} is then

$$l^*(\theta) = l\{h_n^{-C}(\theta)\}, \quad \theta \in \mathbb{R}^p, \tag{2.14}$$

which is well-defined throughout $\mathbb{R}^p$. We now give the properties of the point θ_0' satisfying

$$h_n^{-C}(\theta_0) = \theta_0', \tag{2.15}$$

and the asymptotic distribution of $l^*(\theta_0) = l\{h_n^{-C}(\theta_0)\} = l(\theta_0')$. For convenience, we use $[\tilde{\theta}, \theta_0]$ to denote the line segment in $\mathbb{R}^p$ that connects $\tilde{\theta}$ and θ_0. We have

Lemma 2.1. *Under Conditions 1 and 2, the point θ_0' defined by equation (2.15) satisfies*

$$(i)\ \theta_0' \in [\tilde{\theta}, \theta_0], \quad (ii)\ \theta_0' - \theta_0 = O_p(n^{-3/2}).$$

Theorem 2.2. *Under Conditions 1 and 2, the extended empirical log-likelihood ratio statistic (2.14) satisfies $l^*(\theta_0) \longrightarrow \chi_q^2$ in distribution as $n \to +\infty$.*

Proofs of Lemma 2.1 and Theorem 2.2 are sketched in the Appendix. Detailed proofs are given in the Supplementary Material. A key element in the proof for Theorem 2.2 is the following simple relationship between $l(\theta)$ and $l^*(\theta)$:

$$l^*(\theta_0) = l\{h_n^{-C}(\theta_0)\} = l(\theta_0') = l\{\theta_0 + (\theta_0' - \theta_0)\}. \tag{2.16}$$

This and the fact that $\|\theta_0' - \theta_0\|$ is asymptotically very small imply that $l^*(\theta_0) = l(\theta_0) + o_p(1)$, which leads to Theorem 2.2. The relationship in (2.16) is also the key in the derivation of a second-order extended empirical likelihood in the next section.

2.2.4 Second-order extended empirical likelihood

The Bartlett corrected empirical likelihood of DiCiccio, Hall and Romano (1991) and Chen and Cui (2007) has second-order accuracy. Theorem 2.3 shows that for the just-determined case the extended empirical likelihood can also attain second-order accuracy.

Theorem 2.3. *Assume Conditions 1, 2 and 3 hold. For the just-determined case where $p = q$, let $l_2^*(\theta)$ be the extended empirical log-likelihood ratio under the composite similarity mapping (2.9) with expansion factor*

$$\gamma_2\{n, l(\theta)\} = 1 + \frac{b}{2n}\{l(\theta)\}^{\delta(n)}, \tag{2.17}$$

where $\delta(n) = O(n^{-1/2})$ and b is the Bartlett correction constant in (2.6) and (2.7).

Then

$$l_2^*(\theta_0) = l(\theta_0)\{1 - bn^{-1} + O_p(n^{-3/2})\}. \tag{2.18}$$

The proof of Theorem 2.3 is given in the Supplementary Material. By (2.18) and (2.7), confidence regions based on $l_2^*(\theta)$ have second-order accuracy. Hence, we call $l_2^*(\theta)$ the second-order extended empirical log-likelihood ratio. Correspondingly, we call $l^*(\theta)$ under an h_n^C defined by (2.9) and (2.10) the first-order extended empirical log-likelihood ratio. The utility of the $\delta(n)$ in $\gamma_2\{n, l(\theta)\}$ is to control the speed of domain expansion to ensure that $l_2^*(\theta)$ behaves asymptotically like $l_B(\theta)$. For convenience, we use $\delta(n) = n^{-1/2}$ in our numerical examples.

We noted after Theorem 2.2 that $l^*(\theta_0) = l(\theta_0) + o_p(1)$. An even stronger connection between $l^*(\theta_0)$ and $l(\theta_0)$ is given by Corollary 2.1 below. This result helps to explain the remarkable numerical accuracy of confidence regions based on $l^*(\theta)$ which we will discuss in Section 3.

Corollary 2.1. *Under Conditions 1, 2 and 3, the first-order extended empirical log-likelihood ratio $l^*(\theta)$ for the just-determined case satisfies*

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

2.3 Numerical examples

We compare the first-order extended empirical likelihood with the original and the Bartlett corrected empirical likelihoods through two regression examples. More examples are given in the Supplementary Material. Consider inference for β of a linear model $y = x^T\beta + \varepsilon$, where $\varepsilon \sim N(0, 1)$. We consider two models: Model 1, with $x = (1, x_1)^T$ and $\beta = (1, 2)^T$, and Model 2, with $x = (1, x_1, x_2)^T$ and $\beta = (1, 2, 3)^T$. Variables x_1 and x_2 are assumed to be uniform random variables on $[0, 30]$ and $[20, 50]$, respectively. The original empirical likelihood for β may be found on page 81 of Owen (2001). The extended empirical log-likelihood ratio $l^*(\beta)$ is defined by the composite similarity mapping (2.9) and (2.10) with $\tilde{\theta} = \hat{\beta}$, the least-squares estimate of β. The original and Bartlett corrected empirical likelihood confidence regions are given by (2.3) and (2.8), respectively. The extended empirical likelihood confidence region for β is $\mathcal{C}^*_{1-\alpha} = \{\beta : \beta \in \mathbb{R}^p, l^*(\beta) \leq c\}$ where c is the $(1-\alpha)$th quantile of the χ^2_q distribution. Table 2.1 compares the simulated coverage probabilities of these three confidence regions.

While none of the methods work well for small sample sizes, the extended empirical likelihood is more accurate than the original empirical likelihood for all combinations of sample size and confidence level. In particular, for $n \leq 30$ it is substantially more accurate than the original empirical likelihood. The extended empirical likelihood is also more accurate than the second-order Bartlett corrected empirical likelihood for $n \leq 30$. Remarkably, it remains more accurate than the Bartlett corrected empirical likelihood even for $n > 30$. This surprising observation may be partially explained by Corollary 2.1, where the extended empirical likelihood is seen as having a Bartlett correction type of expansion. See the Supplementary Material for more examples and further discussion.

The parameter vector of Model 2 has dimension $p = 3$ whereas that of Model 1 has $p = 2$. This allows us to assess the impact of an increase in dimension p. When p increases from 2 to 3, the coverage probability of the extended empirical likelihood is the least affected. For small to moderate sample sizes, that of the original empirical likelihood and Bartlett corrected empirical likelihood deteriorates a lot. This is due to the mismatch problem, whose negative impact on coverage accuracy becomes more serious when p increases. The extended empirical likelihood is not affected by the mismatch, so its accuracy holds up much better when p increases.

Table 2.1: Coverage probabilities (%) of confidence regions based on the original empirical likelihood (OEL), the first-order extended empirical likelihood (EEL) and the Bartlett corrected empirical likelihood (BEL)

		90% level			95% level			99% level		
	n	OEL	EEL	BEL	OEL	EEL	BEL	OEL	EEL	BEL
Model 1	10	66.9	80.0	76.3	73.4	88.5	80.9	81.5	98.4	87.5
	20	79.7	85.6	85.1	86.5	92.5	90.8	94.3	98.5	96.6
	30	84.3	87.8	87.2	90.1	93.9	92.6	96.5	98.6	97.5
	50	86.7	88.8	88.5	92.6	94.3	93.7	97.7	98.9	98.2
	100	88.8	89.8	89.6	94.0	94.8	94.5	98.4	99.0	98.6
Model 2	10	47.3	75.1	58.6	54.1	87.2	64.8	65.1	97.7	74.2
	20	69.9	81.2	77.6	77.3	89.7	84.2	88.0	97.8	92.3
	30	76.8	84.3	83.0	84.4	91.1	88.8	92.9	98.1	95.5
	50	83.5	87.2	86.8	89.8	93.1	92.0	96.3	98.5	97.6
	100	87.4	89.1	88.8	93.0	94.4	94.0	98.4	99.0	98.6

Each entry in the table is a simulated coverage probability for β based on 10,000 random samples of size n indicated in column 2 from the linear model indicated in column 1.

We conclude by briefly commenting on the computation of $l^*(\theta)$. Suppose h_n^C is also injective. Since $l^*(\theta) = l(\theta')$, we compute $l^*(\theta)$ by finding the θ' satisfying $h_n^C(\theta') = \theta$ first and then compute $l(\theta')$. We may find this θ' by computing the root for the multivariate function $d(\theta') = h_n^C(\theta') - \theta$, but it is more efficient to reformulate this function as a univariate function by using the fact that $\theta' \in [\tilde{\theta}, \theta]$. See the proof of Theorem 2.1 in the Supplementary Material. When h_n^C is not injective, we find one θ' satisfying $h_n^C(\theta') = \theta$ first, call it θ'_1. Then, look for another satisfying $h_n^C(\theta') = \theta$ in the interval $(\theta'_1, \theta]$, and iterate this process until no new solutions can be found. The last of these, θ'_l, is the solution closest to θ and hence $l^*(\theta) = l(\theta'_l)$.

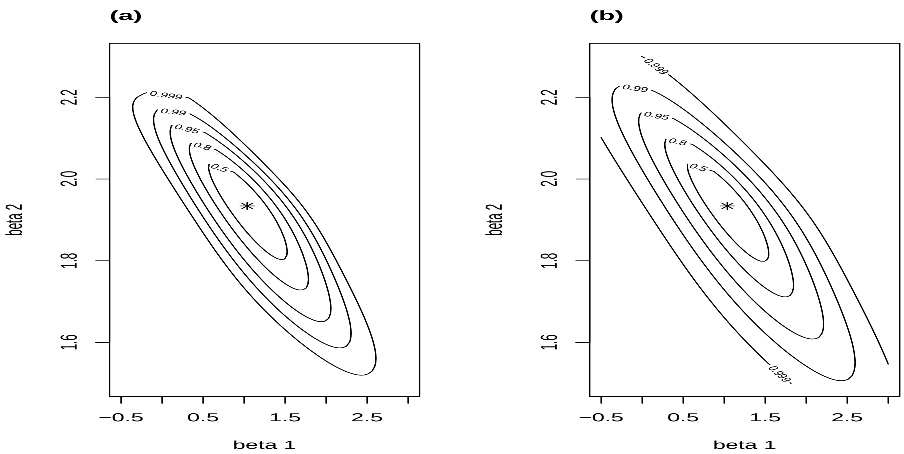

Figure 2.1: Contours of empirical likelihoods for β in Model 1. (a) original empirical likelihood; (b) extended empirical likelihood. Both plots are based on the same sample of 30 observations from Model 1. The star in the middle of each plot shows the least-squares estimate $\tilde{\beta} = (\tilde{\beta}_1, \tilde{\beta}_2) = (1.03, 1.93)^T$ based on this sample. Extended empirical likelihood contours are larger than but similar to original empirical likelihood contours with the same centre and identical shape.

2.4 Discussion

The impressive accuracy of the first-order extended empirical likelihood can also be seen through the examples in the Supplementary Material. We recommend it for practical applications due to its simplicity and superior accuracy. Although the focus of this paper is on extended empirical likelihood for parameters defined by estimating equations, the techniques employed in the proofs may be applied to handle parameters in other settings. In general, an extended empirical likelihood for a parameter θ_0 may be derived so long as a $\sqrt{n}$-consistent maximum empirical likelihood estimator $\tilde{\theta}$ is available. If the original empirical likelihood contours are nested, then the extended empirical likelihood retains not only all asymptotic properties of the original but also the geometric characteristics of its contours. Finally, we have only considered the case where the full parameter space Θ is $\mathbb{R}^p$. The case where Θ is a known subset of $\mathbb{R}^p$ may be handled by finding the extended empirical likelihood on $\mathbb{R}^p$ first and then redefining it as $+\infty$ for $\theta \notin \Theta$. See the Supplementary Material.

Appendix

We identify two assumptions used implicitly throughout this paper. We also sketch the proofs of Lemma 2.1 and Theorem 2.2. Detailed proofs of all results are in the Supplementary Material.

The two assumptions are (a) the original empirical likelihood domain Θ_n defined in (2.1) is an open set in $\mathbb{R}^p$ containing θ_0 and (b) there exists a $\sqrt{n}$-consistent maximum empirical likelihood estimator $\tilde{\theta}$ for θ_0. Assumption (a) ensures, among other things, that the domain Θ_n is non-degenerate. This is needed for domain expansion from Θ_n to $\mathbb{R}^p$. Assumption (b) is required as we need $\tilde{\theta}$ to construct the composite similarity mapping in (2.9) and (2.10). Under Conditions 1 and 2, we may assume without loss of generality that (a) holds. To see this, by Condition 1 and Lemma 11.1 in Owen (2001), with probability tending to 1 that the convex hull of the $g(X_i, \theta_0)$ contains 0 in its interior. Hence, we may assume for sufficiently large n that Θ_n contains θ_0 and it follows that Θ_n is non-empty. To see that Θ_n is open, suppose $\theta \in \Theta_n$. Then, the convex hull of the $g(X_i, \theta)$ contains 0 in its interior. That 0 is in the interior, not on the boundary, of this convex hull is a consequence of the restriction that the w_i in (2.1) are strictly positive. By Condition 2, $g(X_i, \theta)$ is continuous in θ, so a small change in θ will result in only a small change in the convex hull of the $g(X_i, \theta)$. Thus, there exists a small neighbourhood of θ such that for any θ' in that neighbourhood the convex hull of the $g(X_i, \theta')$ also contains 0. Hence, this neighbourhood is inside Θ_n, which implies that Θ_n is open. To see that we may assume (b) also holds under Conditions 1 and 2, we refer to Lemma 1 and Theorem 1 in Qin and Lawless (1994) which give, respectively, the existence and $\sqrt{n}$-consistency of the maximum empirical likelihood estimator.

Proof of Lemma 2.1. Differentiating both sides of equation (2.2) with respect to θ, we obtain $J(\theta_0) = \partial l(\theta)/\partial\theta|_{\theta=\theta_0} = O_p(n^{1/2})$. For θ in a small neighbourhood of θ_0, $\{\theta : \|\theta - \theta_0\| \leq \kappa n^{-1/2}\}$, where κ is a positive constant, Taylor expansion gives

$$l(\theta) = l\{\theta_0 + (\theta - \theta_0)\} = l(\theta_0) + J(\theta_0)(\theta - \theta_0) + O_p(1). \tag{2.19}$$

Since $J(\theta_0) = O_p(n^{1/2})$ and $l(\theta_0) = O_p(1)$, (2.19) implies that $l(\theta) = O_p(1)$. Also, $\gamma\{n, l(\theta)\} \geq 1$ and

$$\theta_0 - \tilde{\theta} = \gamma\{n, l(\theta_0')\}(\theta_0' - \tilde{\theta}), \tag{2.20}$$

so θ_0' is on the ray originating from $\tilde{\theta}$ through θ_0 and $\|\theta_0 - \tilde{\theta}\| \geq \|\theta_0' - \tilde{\theta}\|$. Hence, $\theta_0' \in [\tilde{\theta}, \theta_0]$. This and the $\sqrt{n}$-consistency of $\tilde{\theta}$ imply that $\theta_0' - \theta_0 = O_p(n^{-1/2})$. It follows that $l(\theta_0') = O_p(1)$ and

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

This and (2.20) then yield $\theta_0' - \theta_0 = O_p(n^{-3/2})$. $\square$

Proof of Theorem 2.2. By (ii) of Lemma 2.1, $\theta_0' - \theta_0 = O_p(n^{-3/2})$. Taylor expansion of $l^*(\theta_0)$ gives

$$l^*(\theta_0) = l(\theta_0') = l\{\theta_0 + (\theta_0' - \theta_0)\} = l(\theta_0) + J(\theta_0)(\theta_0' - \theta_0) + O_p(n^{-1}). \tag{2.21}$$

Since $J(\theta_0) = O_p(n^{1/2})$, (2.21) implies that $l^*(\theta_0) = l(\theta_0) + O_p(n^{-1})$. Hence, $l^*(\theta_0)$ has the same limiting χ_q^2 distribution as the original empirical log-likelihood ratio $l(\theta_0)$. $\square$

2.5 Supplement Material

Acknowledgement: we acknowledge that this section contains the supplement material to the published paper Tsao and Wu (2014), Biometrika, volume 101, issue 3, pages 703-710, and is available at Biometrika online.

This following section contains detailed proofs of lemmas and theorems in the paper and three more numerical examples. The first two examples provide a more comprehensive comparison between the extended empirical likelihood method and the existing empirical likelihood methods. The third example illustrates the construction of the extended empirical likelihood for the case where the parameter space Θ is not the full $\mathbb{R}^p$ but a known proper subset of $\mathbb{R}^p$.

Part I: Proofs of Lemmas and Theorems

Proofs in Tsao and Wu (2013) made use of the simple geometric structure of the original empirical likelihood contours for the mean and the simple expression of the original empirical log-likelihood ratio $l(\theta)$ for this special case. However, for parameters defined by estimating equations in general, the geometry of the original empirical likelihood contours is not well understood and difficult to characterize. The expression for $l(\theta)$ in (2.2) depends on the unspecified estimating function $g(X, \theta)$ and is thus also more complicated. In the following, we provide detailed proofs for lemmas and theorems in the paper which do not depend on any specific geometric structure and estimating function. The key components of the proofs are sufficiently general and as such they may also be useful for deriving the extended empirical likelihood on $\mathbb{R}^p$ for parameters beyond the standard estimating equations framework.

Proof of Theorem 2.1. Part (i) of Theorem 2.1 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 original empirical likelihood contour $c(\tau)$ defined by (11). 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 a constant. This is a similarity mapping from $\mathbb{R}^p$ to $\mathbb{R}^p$, and thus a similarity mapping for $c(\tau)$.

Under assumption (a) from the Appendix, the original empirical likelihood domain Θ_n is open. To show (iii), for any given $\theta' \in \mathbb{R}^p$ we need to find 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

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

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 $f(0) = \gamma\{n, l(\tilde{\theta})\} \times 0 = \gamma(n, \tilde{\tau}) \times 0 = 0$. Also,

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

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, $\theta'' \in \Theta_n$ because $\zeta'' \in [0, \zeta_b)$, and

$$\begin{aligned} h_n^C(\theta'') &= \tilde{\theta} + \gamma\{n, l(\theta'')\}(\theta'' - \tilde{\theta}) \\ &= \tilde{\theta} + \gamma\{n, l(\theta'')\}\zeta''\vec{\theta} \\ &= \tilde{\theta} + f(\zeta'')\vec{\theta} \\ &= \tilde{\theta} + \zeta'\vec{\theta} \\ &= \theta'. \end{aligned}$$

Hence, θ'' is the desired point in Θ_n satisfying $h_n^C(\theta'') = \theta'$. This completes the proof for Theorem 2.1. $\square$

Remark. If we add Condition 4 that the original empirical likelihood contours are nested to Theorem 2.1, then the composite similarity mapping h_n^C is also injective. To see this, first note that for a given $c(\tau)$, the mapping $h_n^C : c(\tau) \to c^*(\tau)$ is injective because by (ii) of Theorem 2.1, it is a similarity mapping of $c(\tau)$ and is thus bijective. By the partition of the original empirical likelihood domain Θ_n in (12), 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 because $c^*(\tau_1) \cap c^*(\tau_2) = \varnothing$. To see that $c^*(\tau_1) \cap c^*(\tau_2) = \varnothing$, since $\gamma\{n, l(\theta)\}$ is a strictly increasing function of $\tau = l(\theta)$, h_n^C expands outer contours more than inner ones. Under Condition 4, suppose $c(\tau_1)$ is the inner one relative to $c(\tau_2)$, then $c^*(\tau_1)$ is the inner one related to $c^*(\tau_2)$. As such, they cannot intersect.

Proof of Lemma 2.1. Differentiating $l(\theta)$ in (2.2) and evaluating the derivative at θ_0, we find

$$J(\theta_0) = \frac{\partial l(\theta)}{\partial \theta}|_{\theta=\theta_0} = 2\lambda^T(\theta_0) \sum_{i=1}^{n} \frac{g'(X_i, \theta_0)}{1 + \lambda^T(\theta_0) g(X_i, \theta_0)}, \tag{2.22}$$

where $g'(X_i, \theta_0) = \partial g(X_i, \theta)/\partial\theta|_{\theta=\theta_0}$. Under the conditions of the lemma, we can show that $\lambda(\theta_0) = O_p(n^{-1/2})$ and $J(\theta_0) = O_p(n^{1/2})$. Also, applying Taylor expansion to $l(\theta)$ in a small neighbourhood of θ_0, $\{\theta : \|\theta - \theta_0\| \le \kappa n^{-1/2}\}$, where κ is some positive constant, we obtain

$$l(\theta) = l\{\theta_0 + (\theta - \theta_0)\} = l(\theta_0) + J(\theta_0)(\theta - \theta_0) + O_p(1). \tag{2.23}$$

By Owen (2001), $l(\theta_0) = O_p(1)$. This and (2.23) imply that for a θ in the neighbourhood,

$$l(\theta) = O_p(1). \tag{2.24}$$

To show part (i), since $h_n^C(\theta_0') = \theta_0$, we have

$$\theta_0 - \tilde{\theta} = \gamma\{n, l(\theta_0')\}(\theta_0' - \tilde{\theta}). \tag{2.25}$$

Noting that $\gamma\{n, l(\theta)\} \ge 1$, (2.25) implies that θ_0' is on the ray originating from $\tilde{\theta}$ through θ_0 and

$$\|\theta_0 - \tilde{\theta}\| \ge \|\theta_0' - \tilde{\theta}\|.$$

Hence, $\theta_0' \in [\tilde{\theta}, \theta_0]$ and part (i) of the lemma is proven.

To show part (ii), since $\tilde{\theta}$ is $\sqrt{n}$-consistent and $\theta_0' \in [\tilde{\theta}, \theta_0]$, we have $\theta_0' - \theta_0 = O_p(n^{-1/2})$. It follows from (2.24) that $l(\theta_0') = O_p(1)$. This implies

$$\gamma\{n, l(\theta_0')\} = 1 + \frac{l(\theta_0')}{2n} = 1 + O_p(n^{-1}). \tag{2.26}$$

Adding and subtracting a θ_0 on the right-hand side of (2.25) gives

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

This implies that

$$[1 - \gamma\{n, l(\theta_0')\}](\theta_0 - \tilde{\theta}) = \gamma\{n, l(\theta_0')\}(\theta_0' - \theta_0). \tag{2.27}$$

It follows from (2.26), (2.27) and $\tilde{\theta} - \theta_0 = O_p(n^{-1/2})$ that

$$\theta_0' - \theta_0 = O_p(n^{-3/2}).$$

This proves part (ii) of the lemma. □

Remark. When the composite similarity mapping h_n^C is not injective, we may have more than one θ_0' values satisfying $h_n^C(\theta_0') = \theta_0$. The proof of Lemma 2.1 shows that all such θ_0' values are in the interval $[\tilde{\theta}, \theta_0]$ and within $O_p(n^{-3/2})$ distance from θ_0. Because of this, we may use any such θ_0' value to define the extended empirical likelihood $l^*(\theta_0) = l(\theta_0')$ and obtain the same asymptotic distribution for $l^*(\theta_0)$. But to ensure that $l^*(\theta_0)$ is well-defined, we have chosen through (13) the θ_0' value that is the closest to θ_0.

Proof of Theorem 2.2. By (ii) of Lemma 2.1, $\theta_0' - \theta_0 = O_p(n^{-3/2})$. Taylor expansion of $l^*(\theta_0)$ gives

$$l^*(\theta_0) = l(\theta_0') = l\{\theta_0 + (\theta_0' - \theta_0)\} = l(\theta_0) + J(\theta_0)(\theta_0' - \theta_0) + o_p(n^{-3/2}). \tag{2.28}$$

Since $J(\theta_0) = O_p(n^{1/2})$, (2.28) implies that $l^*(\theta_0) = l(\theta_0) + O_p(n^{-1})$. Thus, the extended empirical log-likelihood ratio $l^*(\theta_0)$ has the same limiting χ_q^2 distribution as the original empirical log-likelihood ratio $l(\theta_0)$. $\square$

To prove Theorem 2.3 and Corollary 2.1, we first give a more detailed review of Bartlett correction for the original empirical likelihood by DiCiccio, Hall and Romano (1991), Chen and Cui (2007) and Liu and Chen (2010). The latter two papers are concerned specifically with Bartlett correction for empirical likelihood for estimating equations, including the over-determined case, whereas the first paper is concerned with that for smooth functions of a mean. For simplicity of presentation, we assume that $\mathrm{var}\{g(X,\theta_0)\} = I_{q\times q}$. There is no loss of generality here since if $\mathrm{var}\{g(X,\theta_0)\} \neq I_{q\times q}$, we can replace $g(X,\theta)$ with $[\mathrm{var}\{g(X,\theta_0)\}]^{-1/2} g(X,\theta)$. For completeness, we begin by repeating the latter part of Section 2.1. Under Conditions 1, 2 and 3, we can show that $l(\theta_0)$ has the following expansion

$$l(\theta_0) = nR^T R + O_p(n^{-3/2}), \tag{2.29}$$

where R is a q-dimensional vector which is a smooth function of general means. Through an Edgeworth expansion for the density function of $n^{1/2}R$, we can show

$$\mathrm{pr}[nR^T R\{1 - bn^{-1} + O_p(n^{-3/2})\} \leq c] = \mathrm{pr}(\chi_q^2 \leq c) + O(n^{-2}), \tag{2.30}$$

where b is the Bartlett correction constant which depends the moments of $g(X,\theta_0)$. It follows from (2.29) and (2.30) that

$$\mathrm{pr}[l(\theta_0)\{1 - bn^{-1} + O_p(n^{-3/2})\} \leq c] = \mathrm{pr}(\chi_q^2 \leq c) + O(n^{-2}). \tag{2.31}$$

Let $l_B(\theta) = (1 - bn^{-1})l(\theta)$ be the Bartlett corrected empirical log-likelihood ratio, and denote by $\mathcal{C}'_{1-\alpha}$ the Bartlett corrected empirical likelihood confidence region for

θ_0. Then,

$$\mathcal{C}'_{1-\alpha} = \{\theta : \theta \in \Theta_n,\ l_B(\theta) \leq c\}.$$

Equation (2.31) implies that

$$\text{pr}(\theta_0 \in \mathcal{C}'_{1-\alpha}) = \text{pr}\{l_B(\theta_0) \leq c\} = \text{pr}(\chi^2_q \leq c) + O(n^{-2}). \tag{2.32}$$

Comparing (2.32) with (2.4), we see that the Bartlett corrected empirical likelihood confidence region has a smaller asymptotic error than the original empirical likelihood region. In practice, the exact/theoretical value of b is unknown as θ_0 and the moments of $g(X, \theta_0)$ are unknown. By (2.31), (2.32) still holds if b is replaced with a $\sqrt{n}$-consistent estimate $\hat{b}$.

Variable R in (2.29) can be written as

$$R = R_1 + R_2 + R_3.$$

This leads to another expression for $l(\theta_0)$,

$$l(\theta_0) = n(R_1 + R_2 + R_3)^T (R_1 + R_2 + R_3) + O_p(n^{-3/2}), \tag{2.33}$$

where each R_i is a function of

$$\alpha^{j_1 j_2 \cdots j_k} = E\left\{\prod_{i=1}^{k} g^{j_i}(X_i; \theta_0)\right\} \quad \text{and} \quad A^{j_1 j_2 \cdots j_k} = n^{-1} \sum_{i=1}^{n} \left\{\prod_{i=1}^{k} g^{j_i}(X_i; \theta_0)\right\} - \alpha^{j_1 j_2 \cdots j_k}.$$

Expressions for R_i in terms of $\alpha^{j_1 j_2 \cdots j_k}$ and $A^{j_1 j_2 \cdots j_k}$ may be found in Chen and Cui (2007) and Liu and Chen (2010). For our proofs, we need only the following observations based on these expressions:

$$(i) \quad R_i = O_p(n^{-j/2}) \quad \text{for } j = 1, 2, 3. \tag{2.34}$$

$$(ii) \quad R_1 = (A^1, A^2, \ldots, A^q)^T = \frac{1}{n}\sum_{i=1}^{n} g(X_i, \theta_0). \tag{2.35}$$

$$(iii) \quad \lambda(\theta_0) = R_1 + O_p(n^{-1}). \tag{2.36}$$

See Liu and Chen (2010) and Chen and Cui (2007) for detailed discussions on Bartlett correction for the original empirical likelihood for parameters defined by estimating equations. The proof of Theorem 2.3 needs the following lemma.

Lemma 2.2. *Assume Conditions 1, 2 and 3 hold. Under the composite similarity mapping (2.9) with expansion factor* $\gamma\{n, l(\theta)\} = \gamma_2\{n, l(\theta)\}$ *in (2.17), we have*

$$\theta_0' - \theta_0 = \frac{b}{2n}(\tilde{\theta} - \theta_0) + O_p(n^{-2}). \tag{2.37}$$

Proof of Lemma 2.2 . It may be verified that under the three conditions and with the composite similarity mapping h_n^C defined by (2.9) and (2.17), Theorem 2.1, Lemma 2.1 and Theorem 2.2 all hold. In particular, $\theta_0' - \theta_0 = O_p(n^{-3/2})$ and the extended empirical log-likelihood ratio $l_2^*(\theta_0)$ converges in distribution to a χ_q^2 random variable.

Since $\delta(n) = O(n^{-1/2})$ and $l(\theta_0') = l_2^*(\theta_0)$ which is asymptotically a χ_q^2 variable, we have

$$\{l(\theta_0')\}^{\delta(n)} = 1 + O_p(n^{-1/2}). \tag{2.38}$$

By $h_n^C(\theta_0') = \theta_0$, we have $\theta_0 - \tilde{\theta} = \gamma_2\{n, l(\theta_0')\}(\theta_0' - \tilde{\theta})$. Thus,

$$\theta_0' - \theta_0 = \frac{b\{l(\theta_0')\}^{\delta(n)}}{2n}(\tilde{\theta} - \theta_0') = \frac{b\{l(\theta_0')\}^{\delta(n)}}{2n}(\tilde{\theta} - \theta_0) + \frac{b\{l(\theta_0')\}^{\delta(n)}}{2n}(\theta_0 - \theta_0'). \tag{2.39}$$

It follows from (2.38), (2.39) and $\theta_0' - \theta_0 = O_p(n^{-3/2})$ that

$$\begin{aligned}\theta_0' - \theta_0 &= \frac{b\{l(\theta_0')\}^{\delta(n)}}{2n}(\tilde{\theta} - \theta_0) + O_p(n^{-5/2}) \\ &= \frac{b}{2n}(\tilde{\theta} - \theta_0) + O_p(n^{-2}),\end{aligned}$$

which proves the lemma. □

Proof of Theorem 2.3. By (2.37) from Lemma 2.2 and Taylor expansion (2.28), we have

$$\begin{aligned}l^*(\theta_0) &= l(\theta_0) + J(\theta_0)(\theta_0' - \theta_0) + o_p(n^{-3/2}) \\ &= l(\theta_0) + \frac{b}{2n}J(\theta_0)(\tilde{\theta} - \theta_0) + O_p(n^{-3/2}), \qquad (2.40)\end{aligned}$$

where $J(\theta_0)$ is given by (2.22). Under Condition 2, Taylor expansion of $g(X_i, \tilde{\theta})$ at θ_0 gives

$$g(X_i, \tilde{\theta}) = g(X_i, \theta_0) + g'(X_i, \theta_0)(\tilde{\theta} - \theta_0) + O_p(\|\theta_0 - \tilde{\theta}\|^2).$$

This and $\tilde{\theta} - \theta_0 = O_p(n^{-1/2})$ imply that for each $i \in \{1, 2, \ldots, n\}$,

$$g'(X_i, \theta_0)(\theta_0 - \tilde{\theta}) = g(X_i, \theta_0) - g(X_i, \tilde{\theta}) + O_p(n^{-1}).$$

Averaging the above equation over i gives

$$\frac{1}{n}\sum_{i=1}^{n} g'(X_i, \theta_0)(\theta_0 - \tilde{\theta}) = \frac{1}{n}\sum_{i=1}^{n} g(X_i, \theta_0) - \frac{1}{n}\sum_{i=1}^{n} g(X_i, \tilde{\theta}) + O_p(n^{-1}). \qquad (2.41)$$

Since the estimating equations are just-determined, $n^{-1}\sum_{i=1}^{n} g(X_i, \tilde{\theta}) = 0$. This and

(2.41) imply

$$\frac{1}{n}\sum_{i=1}^{n} g'(X_i,\theta_0)(\theta_0-\tilde{\theta}) = \frac{1}{n}\sum_{i=1}^{n} g(X_i,\theta_0) + O_p(n^{-1}). \tag{2.42}$$

Noting that $\lambda(\theta_0) = O_p(n^{-1/2})$ and $\theta_0 - \tilde{\theta} = O_p(n^{-1/2})$, we can show

$$\frac{1}{n}\sum_{i=1}^{n} \frac{g'(X_i,\theta_0)(\theta_0-\tilde{\theta})}{1+\lambda^T(\theta_0)g(X_i,\theta_0)} = \frac{1}{n}\sum_{i=1}^{n} g'(X_i,\theta_0)(\theta_0-\tilde{\theta}) + O_p(n^{-1}). \tag{2.43}$$

It follows from (2.42) and (2.43) that

$$\frac{1}{n}\sum_{i=1}^{n} \frac{g'(X_i,\theta_0)(\theta_0-\tilde{\theta})}{1+\lambda^T(\theta_0)g(X_i,\theta_0)} = \frac{1}{n}\sum_{i=1}^{n} g(X_i,\theta_0) + O_p(n^{-1}). \tag{2.44}$$

By (2.40), (2.22) and (2.44), we have

$$\begin{aligned}
l^*(\theta_0) &= l(\theta_0) + \frac{b}{2n} J(\theta_0)(\tilde{\theta}-\theta_0) + O_p(n^{-3/2}) \\
&= l(\theta_0) - \frac{b}{2n} 2\lambda^T(\theta_0) \sum_{i=1}^{n} \frac{g'(X_i,\theta_0)(\theta_0-\tilde{\theta})}{1+\lambda^T(\theta_0)g(X_i,\theta_0)} + O_p(n^{-3/2}) \\
&= l(\theta_0) - \frac{b}{n} n\lambda^T(\theta_0) n^{-1} \sum_{i=1}^{n} \frac{g'(X_i,\theta_0)(\theta_0-\tilde{\theta})}{1+\lambda^T(\theta_0)g(X_i,\theta_0)} + O_p(n^{-3/2}) \\
&= l(\theta_0) - \frac{b}{n} n\lambda^T(\theta_0) \left\{ n^{-1} \sum_{i=1}^{n} g(X_i,\theta_0) + O_p(n^{-1}) \right\} + O_p(n^{-3/2}) \\
&= l(\theta_0) - \frac{b}{n} n\lambda^T(\theta_0) \left\{ n^{-1} \sum_{i=1}^{n} g(X_i,\theta_0) \right\} + O_p(n^{-3/2}).
\end{aligned} \tag{2.45}$$

Finally, by (2.45), (2.34), (2.35), (2.36) and (2.33), we have

$$\begin{aligned} l^*(\theta_0) &= l(\theta_0) - \frac{b}{n} n R_1^T R_1 + O_p(n^{-3/2}) \\ &= l(\theta_0) - \frac{b}{n} n (R_1 + R_2 + R_3)^T (R_1 + R_2 + R_3) + O_p(n^{-3/2}) \\ &= l(\theta_0) - \frac{b}{n} l(\theta_0) + O_p(n^{-3/2}) \\ &= l(\theta_0) \left\{ 1 - \frac{b}{n} + O_p(n^{-3/2}) \right\}, \end{aligned}$$

which proves Theorem 2.3. □

Remark. The second-order result of Theorem 2.3 holds only for the just-determined case as the proof above used the condition that $n^{-1} \sum_{i=1}^n g(X_i, \tilde{\theta}) = 0$ to go from (2.41) to (2.42). For the over-determined case, a weaker condition $n^{-1} \sum_{i=1}^n g(X_i, \tilde{\theta}) = O_p(n^{-1})$ would also allow us to go from (2.41) to (2.42). However, we have yet to identify the type of estimating function $g(X, \theta)$ under which this weaker condition would hold for the over-determined case. When it does hold, the extended empirical log-likelihood ratio $l_2^*(\theta)$ defined in Theorem 2.3 has the second-order accuracy for the over-determined case as well. When it does not hold, $l_2^*(\theta)$ reduces to a first-order extended empirical log-likelihood ratio as Theorem 2.2 is still valid for $l_2^*(\theta)$.

Proof of Corollary 2.1. We first show that under the composite similarity mapping h_n^C defined by expansion factor (2.10), $\theta_0' = h_n^{-C}(\theta_0)$ satisfies

$$\theta_0' - \theta_0 = \frac{l(\theta_0)}{2n} (\tilde{\theta} - \theta_0) + O_p(n^{-5/2}). \tag{2.46}$$

In the proof of Theorem 2.2 above, we noted that

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

Since $l(\theta_0') = l^*(\theta_0)$, this implies

$$l(\theta_0') = l(\theta_0) + O_p(n^{-1}). \tag{2.47}$$

The expansion factor in (10) may be viewed as a special case of that in (17) where $\delta(n) = 1$ and $b = 1$. Setting $\delta(n) = 1$ and $b = 1$ in the proof Lemma 2.2 and replacing equation (2.38) with (2.47), we obtain (2.46) by following the rest of the steps in that proof. Finally, using (2.46) instead of equation (2.37) from Lemma 2.2 in (2.40) and following exactly the same steps in the proof of Theorem 2.3 after (2.40), we obtain Corollary 2.1. □

Part II: Additional Numerical Examples

We now present the following numerical examples to compare the extended empirical likelihood method with the existing empirical likelihood methods: [1] a simple linear model with three different error distributions, [2] an over-determined example from Qin and Lawless (1994) and Chen and Cui (2007) and [3] an example on simultaneous inference for the mean and variance of a univariate random variable. The third example involves two parameters for which the parameter space is the first quadrant instead of the entire $\mathbb{R}^2$.

For convenience, we first compare the first-order extended empirical likelihood with the original empirical likelihood and the Bartlett corrected empirical likelihood. The latter two methods serve as the benchmarks for evaluating the accuracy of the extended empirical likelihood. Then, we compare three second-order methods: the Bartlett corrected empirical likelihood (Chen and Cui, 2007), the second-order adjusted empirical likelihood (Liu and Chen, 2010) and the second-order extended empirical likelihood. The Bartlett correction constant used in all second-order methods is the biased corrected estimate $\tilde{b}$ given by Liu and Chen (2010).

Example 1: a simple linear model under three different error distributions

Table 2.2 contains simulated coverage probabilities of confidence regions based on the original empirical likelihood, the first-order extended empirical likelihood and the Bartlett corrected empirical likelihood for parameter vector $\boldsymbol{\beta}$ of the linear model

$$y = x^T\beta + \varepsilon,$$

where $x = (1, x_1)^T$ and $\beta = (1, 2)^T$. The error distributions considered are [i] $\varepsilon \sim N(0,1)$, [ii] $\varepsilon \sim EXP(1) - 1$ and [iii] $\varepsilon \sim \chi_1^2 - 1$. For the simulation, values of x_1 are randomly generated from a uniform distribution on $[0, 30]$. For symmetric error distribution [i], the extended empirical likelihood and Bartlett corrected empirical likelihood are more accurate than the original empirical likelihood and substantially so when the sample size is not large. The extended empirical likelihood is also competitive in accuracy to the Bartlett corrected empirical likelihood even when the sample size is large. This is surprising in that the Bartlett corrected empirical likelihood is a second-order method whereas the extended empirical likelihood in this table is only a first-order method. For skewed error distributions [ii] and [iii], the extended empirical likelihood and Bartlett corrected empirical likelihood are also substantially more accurate than the original empirical likelihood. The extended empirical likelihood is still more accurate than the Bartlett corrected empirical likelihood for small and moderate sample sizes but the Bartlett corrected empirical likelihood is slightly more accurate for large sample sizes.

Table 2.2: Example 1: Coverage probabilities (%) of confidence regions based on the original empirical likelihood (OEL), the first-order extended empirical likelihood (EEL) and the Bartlett corrected empirical likelihood (BEL)

		90% level			95% level			99% level		
Error Distribution	n	OEL	EEL	BEL	OEL	EEL	BEL	OEL	EEL	BEL
$N(0,1)$	10	66.0	77.6	75.5	72.8	84.9	80.7	81.5	93.4	87.2
	20	79.5	85.2	84.8	86.1	91.6	90.0	93.9	97.5	95.8
	30	84.1	87.4	87.0	90.0	93.5	92.3	96.3	98.5	97.4
	50	87.0	89.0	88.7	92.6	94.1	93.6	97.8	98.7	98.2
	100	89.1	90.2	90.0	94.4	95.2	94.9	98.6	98.9	98.8
$EXP(1)-1$	10	62.9	73.7	70.5	70.1	81.5	76.3	80.0	90.6	84.0
	20	75.0	80.7	81.1	81.8	87.7	86.4	90.5	95.2	93.1
	30	79.2	83.0	83.5	85.8	89.7	89.3	93.7	96.6	95.5
	50	83.8	86.1	87.0	90.0	92.0	92.0	96.2	97.9	97.2
	100	87.6	88.7	89.1	93.3	94.4	94.5	98.1	98.8	98.5
χ_1^2-1	10	59.9	70.0	65.6	66.6	77.3	70.9	76.1	86.7	77.7
	20	70.3	76.9	76.8	78.0	83.8	82.2	86.8	92.2	89.2
	30	76.3	80.2	81.3	83.2	87.0	86.5	91.2	94.4	92.4
	50	81.6	84.2	85.6	88.4	90.7	91.0	95.3	97.0	96.4
	100	86.5	87.5	88.4	92.3	93.4	93.6	97.7	98.2	98.1

Each entry in the table is a simulated coverage probability for $\boldsymbol{\beta}$ based on 10,000 random samples of size n indicated in column 2 from the linear model with error distribution indicated in column 1.

Table 2.3 compares the coverage accuracies of the three second-order methods. For small sample sizes, the second-order adjusted empirical likelihood coverage probability is seen to be the highest. For $n = 10$, it even exceeds the nominal levels. But this is due to the boundedness problem (Emerson and Owen, 2009) of the adjusted empirical likelihood statistic which artificially boosted the coverage probability of the adjusted empirical likelihood. The problem arises when the adjusted empirical likelihood statistic is bounded from the above by the Chi-square critical value for all θ values in the parameters space, resulting in trivial 100% confidence regions which coincide with the entire parameter space. When this occurs, the adjusted empirical likelihood confidence region trivially contains the true parameter value and this inflates the coverage probability of the adjusted empirical likelihood. Detecting and removing such cases when simulating the coverage probability is possible but time consuming, especially for multivariate problems. There are also variations of the adjusted empirical likelihood which do not have the boundedness problem. A more comprehensive comparison involving these will be reported elsewhere. Our experience suggests that if we remove the cases where the adjusted empirical likelihood statistic is bounded from our calculation, the coverage probability of the adjusted empirical likelihood is comparable to that of the Bartlett corrected empirical likelihood.

Putting aside the coverage probabilities of the adjusted empirical likelihood, Table 2.3 shows that the second-order extended empirical likelihood is consistently more accurate than the Bartlett corrected empirical likelihood for all sample sizes and error distributions. Interestingly, comparing Tables 2.2 and 2.3, we see that the first-order extended empirical likelihood is very competitive to the second-order extended empirical likelihood in all cases. While we do not have a full explanation for this, Corollary 2.1 shows the first-order extended empirical likelihood has an expansion similar to that of the Bartlett corrected empirical likelihood in (2.31) with the Bartlett

Table 2.3: Example 1: Coverage probabilities (%) of confidence regions based on the Bartlett corrected empirical likelihood (BEL), the second-order adjusted empirical likelihood (AEL) and the second-order extended empirical likelihood (EEL_2)

		90% level			95% level			99% level		
Error Distribution	n	BEL	AEL	EEL_2	BEL	AEL	EEL_2	BEL	AEL	EEL_2
$N(0,1)$	10	75.5	92.9	78.7	80.7	97.5	84.1	87.2	99.8	90.8
	20	84.8	87.8	85.7	90.0	93.5	91.1	95.8	98.7	96.4
	30	87.0	87.6	87.4	92.3	93.1	92.8	97.4	98.0	97.7
	50	88.7	89.0	89.0	93.6	93.8	93.9	98.2	98.3	98.3
	100	90.0	90.0	90.2	94.9	94.9	94.9	98.8	98.8	98.8
$EXP(1)-1$	10	70.5	85.3	73.4	76.3	92.9	79.5	84.0	98.8	87.3
	20	81.1	84.6	82.0	86.4	90.9	87.7	93.1	97.3	94.2
	30	83.5	85.3	84.1	89.3	90.9	89.8	95.5	96.9	95.9
	50	87.0	87.5	87.2	92.0	92.5	92.3	97.2	97.6	97.5
	100	89.1	89.2	89.3	94.5	94.5	94.6	98.5	98.5	98.6
χ^2_1-1	10	65.6	84.2	70.5	70.9	91.5	76.5	77.7	98.4	83.5
	20	76.8	83.5	79.0	82.2	89.3	84.3	89.2	96.4	91.3
	30	81.3	85.0	82.5	86.5	90.2	88.1	92.4	96.2	94.1
	50	85.6	86.6	86.0	91.0	92.0	91.4	96.4	97.3	96.9
	100	88.4	88.4	88.7	93.6	93.7	93.8	98.1	98.2	98.2

Each entry in the table is a simulated coverage probability for $\boldsymbol{\beta}$ based on 10,000 random samples of size n indicated in column 2 from the linear model with error distribution indicated in column 1.

correction constant b replaced by $l(\theta_0)$. This resemblance may be the reason that the first-order extended empirical likelihood behaves like the second-order Bartlett corrected empirical likelihood for large sample sizes. But the good accuracy of the first-order extended empirical likelihood for small sample sizes cannot be accounted for by any allusion to its asymptotic order; it is the benefit of being free from the mismatch problem between the domain and the parameter space which affects the original and Bartlett corrected empirical likelihoods.

Example 2: over-determined estimating equations

One over-determined estimating equations example used in both Qin and Lawless (1994) and Chen and Cui (2007) is the following: for a univariate random variable X, suppose $E(X) = \theta$ and $E(X^2) = 2\theta^2 + 1$. Then, θ is one-dimensional but there are two estimating equations. In this case, empirical likelihood inference may be conducted by using the original empirical likelihood defined by (2.2), or alternatively by using the $W_E(\theta)$ statistic given by (3.9) in Qin and Lawless (1994). The $W_E(\theta)$ statistic reduces to the original empirical likelihood statistic for just-determined cases but may be more efficient than the latter for over-determined cases. The Bartlett corrected empirical likelihood is also available for $W_E(\theta)$ (Chen and Cui, 2007) and an extended version of $W_E(\theta)$ can be easily defined by replacing the original empirical log-likelihood ratio $l(\theta)$ with $W_E(\theta)$ throughout Sections 2.2 to 2.4. We will use both the original empirical likelihood (2.2) and the W_E statistic. As in Qin and Lawless (1994) and Chen and Cui (2007), we assume that $X \sim N(\theta, \theta^2 + 1)$ and consider two cases (i) $\theta = 0$ and (ii) $\theta = 1$.

Table 2.4 compares the original empirical likelihood, the first-order extended empirical likelihood and the Bartlett corrected empirical likelihood. The extended empirical likelihood is again the most accurate among the three methods. It is worth mentioning here that when we do not use the extra information about the second moment by removing the second estimating equation $g(x, \theta) = x^2 - 2\theta^2 - 1$ from the empirical likelihood, we obtain higher coverage probabilities for all three methods. This suggests that the extra second moment information, when incorporated into the empirical likelihood through the second estimating equation, has a negative impact on the accuracy. This may seem to be counter-intuitive but is in fact another example of the serious negative impact of a higher dimension on the empirical likelihood; the benefit of the extra information represented by the second estimating equation is

Table 2.4: Example 2: Coverage probabilities (%) of confidence regions based on the original empirical likelihood (OEL), the first-order extended empirical likelihood (EEL) and the Bartlett corrected empirical likelihood (BEL)

		90% level			95% level			99% level		
Parameter value	n	OEL	EEL	BEL	OEL	EEL	BEL	OEL	EEL	BEL
$\theta = 0$	10	64.2	72.9	71.4	70.1	78.4	75.2	76.4	85.9	80.1
	30	82.2	85.9	85.6	88.5	91.9	90.9	95.1	97.0	96.1
	60	86.3	88.0	87.8	92.1	93.6	93.2	97.6	98.4	98.0
$\theta = 1$	10	64.1	72.7	70.7	69.9	78.4	74.8	76.5	85.7	80.1
	30	82.5	85.9	85.5	88.3	91.5	90.6	94.8	96.9	95.8
	60	85.9	87.6	87.6	91.8	93.4	93.0	97.4	98.3	97.9

The original empirical likelihood here is given by (2.2), and the extended empirical likelihood and Bartlett corrected empirical likelihood in this table are based on this original empirical likelihood. Each entry in the table is a simulated coverage probability for θ based on 10,000 random samples of size n from $N(\theta, \theta^2 + 1)$.

out-weighted by the negative impact of an increase in dimension, and the net effect of incorporating the extra information is a deterioration in coverage accuracy.

Table 2.5 compares the $W_E(\theta)$ statistic with the first-order extended empirical likelihood and the Bartlett corrected empirical likelihood based on the $W_E(\theta)$. The extended empirical likelihood is the most accurate. It is interesting that all three methods are more accurate when $\theta = 0$ and less so when $\theta = 1$. This is not the case in Table 2.4 where the original empirical likelihood (2.2) is compared to the extended and Bartlett corrected empirical likelihoods based on the original empirical likelihood. This suggests the performance of the $W_E(\theta)$ statistic and its associated extended and Bartlett corrected empirical likelihoods depend on the value of the unknown parameter. See also results in Qin and Lawless (1994) and Chen and Cui (2007).

Table 2.6 compares the three second-order methods: the Bartlett corrected empirical likelihood, the second-order adjusted empirical likelihood and the second-order extended empirical likelihood based on the original empirical likelihood (2.2). Table

Table 2.5: Example 2: Coverage probabilities (%) of confidence regions based on the W_E statistic of Qin and Lawless (1994), the first-order extended empirical likelihood (EEL) and the Bartlett corrected empirical likelihood (BEL)

		90% level			95% level			99% level		
Parameter value	n	W_E	EEL	BEL	W_E	EEL	BEL	W_E	EEL	BEL
$\theta = 0$	10	75.5	88.0	81.9	81.9	94.2	86.8	88.7	98.4	91.2
	30	86.5	89.3	88.9	92.6	95.2	94.5	98.3	99.4	98.9
	60	88.3	89.4	89.5	93.9	95.0	94.7	98.7	99.3	99.0
$\theta = 1$	10	67.7	76.0	73.2	73.4	82.2	77.7	80.4	89.5	83.3
	30	84.0	86.6	86.5	89.9	92.3	91.9	96.1	97.6	96.9
	60	87.4	88.5	88.6	92.7	93.7	93.5	98.0	98.6	98.3

The W_E statistics is defined by (3.9) in Qin and Lawless (1994). The EEL and BEL in this table are based on W_E. Each entry in the table is a simulated coverage probability for θ based on 10,000 random samples of size n from $N(\theta, \theta^2 + 1)$.

Table 2.6: Example 2: Coverage probabilities (%) of confidence regions based on the Bartlett corrected empirical likelihood (BEL), the second-order adjusted empirical likelihood (AEL) and the second-order extended empirical likelihood (EEL_2)

		90% level			95% level			99% level		
Parameter value	n	BEL	AEL	EEL_2	BEL	AEL	EEL_2	BEL	AEL	EEL_2
θ=0	10	71.4	82.9	74.6	75.2	88.2	78.6	80.1	95.6	83.9
	30	85.6	86.9	86.5	90.9	91.8	91.4	96.1	96.8	96.4
	60	87.8	88.1	88.2	93.2	93.4	93.4	98.0	98.0	98.0
θ=1	10	70.7	82.8	74.4	74.8	87.8	78.4	80.1	94.7	83.9
	30	85.5	86.7	86.4	90.6	91.5	91.2	95.8	96.5	96.1
	60	87.6	87.9	87.9	93.0	93.2	93.2	97.9	98.0	98.0

The BEL, AEL and EEL_2 here are based on the original empirical likelihood given by (2.2). Each entry in the table is a simulated coverage probability for θ based on 10,000 random samples of size n from $N(\theta, \theta^2 + 1)$.

Table 2.7: Example 2: Coverage probabilities (%) of confidence regions based on the Bartlett corrected empirical likelihood (BEL), the second-order adjusted empirical likelihood (AEL) and the second-order extended empirical likelihood (EEL_2)

		90% level			95% level			99% level		
Parameter value	n	BEL	AEL	EEL_2	BEL	AEL	EEL_2	BEL	AEL	EEL_2
θ=0	10	81.9	93.3	93.8	86.8	96.7	95.8	91.2	99.4	97.5
	30	88.9	90.2	91.2	94.5	95.6	96.0	98.9	99.4	99.3
	60	89.5	89.7	90.1	94.7	94.9	95.2	99.0	99.1	99.2
θ=1	10	73.2	86.5	82.0	77.7	91.1	85.2	83.3	96.1	89.2
	30	86.5	87.8	88.2	91.9	92.8	92.9	96.9	97.5	97.4
	60	88.6	88.8	89.3	93.5	93.7	94.0	98.3	98.5	98.5

The BEL, AEL and EEL_2 here are based on the W_E statistics defined by (3.9) in Qin and Lawless (1994). Each entry in the table is a simulated coverage probability for θ based on 10,000 random samples of size n from $N(\theta, \theta^2+1)$.

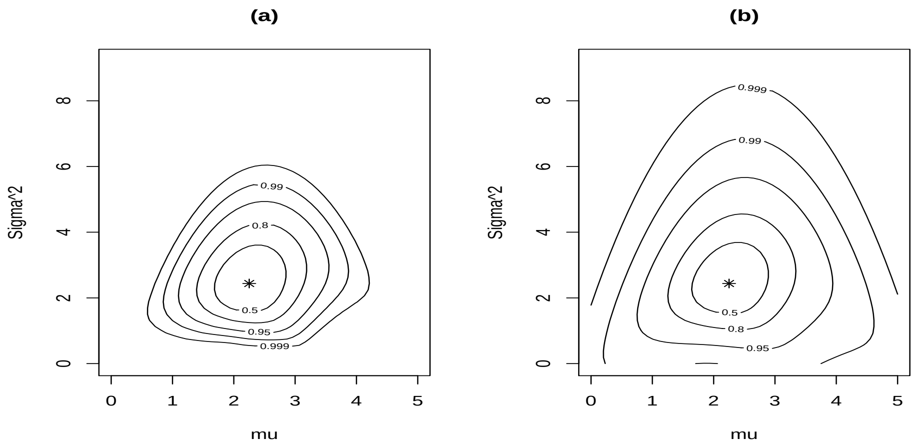

Figure 2.2: Contours of empirical likelihoods for (μ, σ^2). (a) original empirical likelihood; (b) extended empirical likelihood. Both plots are based the same sample of 10 observations from $N(2,3)$. The star in the middle of the plot is the maximum empirical likelihood estimate $(\tilde{\mu}, \tilde{\sigma}^2) = (2.25, 2.44)$. Extended empirical likelihood contours are larger than but similar to the original empirical likelihood contours with the same centre and identical shape, and by definition in Example 3 they are truncated at the boundaries of the first quadrant.

2.7 compares corresponding second-order methods based on the W_E statistic. In both tables, the adjusted empirical likelihood seems to have higher coverage probabilities for smaller sample sizes but these again are inflated by the boundedness problem of the adjusted empirical likelihood statistic. For the large sample size of $n = 60$ where the boundedness problem is not likely to occur, we see that the adjusted empirical likelihood is comparable to the Bartlett corrected empirical likelihood. Putting aside the inflated adjusted empirical likelihood coverage probabilities, we see again that the second-order extended empirical likelihood is more accurate than the Bartlett corrected empirical likelihood.

Finally, when computing the estimated Bartlett correction factor $\tilde{b}$ for the over-determined cases in Table 2.6 and Table 2.7, we have noted very high variations of the $\tilde{b}$ values for the same sample size and θ value. This is likely due to the high order sample moments used in the estimator which is highly variable. This unstable behaviour of $\tilde{b}$ for over-determined cases was also observed in Chen and Cui (2007) and Liu and Chen (2010), and led the latter to caution against the use of Bartlett correction for such cases.

Example 3: when the parameter space is not $\mathbb{R}^p$

The extended empirical likelihood $l^*(\theta)$ on the full $\mathbb{R}^p$ may violate known constraints on the parameter space Θ. Suppose we wish to make simultaneous inference about the mean μ and variance σ^2 of same univariate random variable with μ known to be non-negative. Then, the parameter space Θ is the first quadrant instead of the entire $\mathbb{R}^2$. As such, $l^*(\theta)$ is only meaningful for θ in the first quadrant. In this case, we redefine $l^*(\theta)$ for $\theta \notin \Theta$ as $l^*(\theta) = +\infty$ to ensure the extended empirical likelihood does not violate the known constraints. Consequently, the extended empirical likelihood contours stop at the boundaries of the parameter space Θ. Figure 2.2 shows an example of this based on a sample of size 10 from $N(\mu = 2, \sigma^2 = 3)$.

Chapter 3

Two-sample extended empirical likelihood for the mean

Acknowledgement: in accordance with copyright agreements between authors and journal publishers, we acknowledge that this chapter is the published paper under the same title by Wu and Tsao (2014), Statistics and Probability Letters, 2014, volume 84, issue C, pages 81-87, with a 11-page Supplement Material available at Statistics and Probability Letters online.

3.1 Introduction

The empirical likelihood introduced by Owen (1988, 1990) is a versatile non-parametric method of inference with many applications (Owen, 2001). One problem which the empirical likelihood method has been successfully applied to is the two-sample problem (Jing, 1995; Liu, Zou and Zhang 2008; Liu and Yu 2010; Wu and Yan, 2012) where the parameter of interest θ is the difference between the means of two populations. The well-known Behrens-Fisher problem is a special two-sample problem where the two populations are known to be normally distributed. Following DiCiccio, Hall and Romano (1991) who showed the surprising result that the (one-sample) empirical likelihood for a smooth function of the mean is Bartlett correctable, Jing (1995) and Liu, Zou and Zhang (2008) proved that the two-sample empirical likelihood for θ is also Bartlett correctable. The coverage error of a confidence region based on the original empirical likelihood is $O(n^{-1})$, but that based on the Bartlett corrected empirical likelihood is only $O(n^{-2})$.

For a one-sample empirical likelihood, there is a mismatch between its domain and the parameter space in that it is defined on only a part of the parameter space. This mismatch is a main cause of the undercoverage problem associated with empirical likelihood confidence regions (Tsao, 2013). The two-sample empirical likelihood for θ also has the mismatch problem as it is defined on a bounded region but the parameter space is $\mathbb{R}^d$. In this paper, we derive an extended version of the original two-sample empirical likelihood (OEL) by expanding its domain into $\mathbb{R}^d$ through the composite similarity mapping of Tsao and Wu (2013). The resulting two-sample extended empirical likelihood (EEL) for θ is defined on the entire $\mathbb{R}^d$ and hence free from the mismatch problem. Under mild conditions, this EEL has the same asymptotic properties as the OEL. It can also attain the second order accuracy of the two-sample Bartlett corrected empirical likelihood (BEL) of Jing (1995) and Liu, Zou and Zhang (2008). The first order version of this EEL is substantially more accurate than the OEL, especially for small sample sizes. It is also easy to compute and competitive in accuracy to the second order methods. We recommend it for two-sample empirical likelihood inference.

3.2 Two-sample empirical likelihood

Let $\{X_1, \ldots, X_m\}$ and $\{Y_1, \ldots, Y_n\}$ be independent copies of random vectors $X \in \mathbb{R}^d$ and $Y \in \mathbb{R}^d$, respectively. Denote by μ_x and Σ_x the mean and covariance matrix of X, and by μ_y and Σ_y the mean and covariance matrix of Y, respectively. The unknown parameter of interest is the difference in means $\theta_0 = \mu_y - \mu_x \in \mathbb{R}^d$ and the parameter space is the entire $\mathbb{R}^d$. We will need the following three conditions later in the paper:

C1. Σ_x and Σ_y are finite covariance matrix with full rank d;

C2. $\limsup_{\|t\|\to\infty} |E[\exp\{it^T X\}]| < 1$ and $\limsup_{\|t\|\to\infty} |E[\exp\{it^T Y\}]| < 1$;

C3. $E\|X\|^{15} < +\infty$ and $E\|Y\|^{15} < +\infty$.

Condition *C1* is needed to establish the first order result for the EEL and conditions *C2* and *C3* are needed for the second order result. Denote by $p = (p_1, ..., p_m)$ and $q = (q_1, ..., q_n)$ two probability vectors satisfying $p_i \geq 0$, $q_j \geq 0$, $\sum_{i=1}^m p_i = 1$ and $\sum_{i=1}^n q_j = 1$. Let $\mu_x(p) = \sum_{i=1}^m p_i X_i$ and $\mu_y(q) = \sum_{j=1}^n q_j Y_j$, and denote by $\theta(p,q)$ their difference, that is,

$$\theta(p,q) = \mu_y(q) - \mu_x(p).$$

The original two-sample empirical likelihood for a $\theta \in \mathbb{R}^d$, $L(\theta)$, is defined as

$$L(\theta) = \max_{(p,q):\theta(p,q)=\theta} \left(\prod_{i=1}^m p_i\right)\left(\prod_{j=1}^n q_j\right). \tag{3.1}$$

The corresponding two-sample empirical log-likelihood ratio for θ is thus

$$l(\theta) = -2 \max_{(p,q):\theta(p,q)=\theta} \left(\sum_{i=1}^m \log(mp_i) + \sum_{j=1}^n \log(nq_j)\right). \tag{3.2}$$

In order to develop our extended empirical likelihood, it is important to first investigate the domains of the original empirical likelihood ratio $L(\theta)$ and log-likelihood ratio $l(\theta)$. The domain of $L(\theta)$ is given by

$$D_\theta = \{\theta \in \mathbb{R}^d : \text{there exist } p \text{ and } q \text{ such that } \mu_x(p) = \textstyle\sum_{i=1}^m p_i X_i,$$

$$\mu_y(q) = \textstyle\sum_{j=1}^n q_j Y_j \text{ and } \theta = \theta(p, q) = \mu_y(q) - \mu_x(p)\}.$$

Since the "range" of $\mu_x(p)$ and $\mu_y(q)$ are the convex hulls of the X_i and Y_i, respectively, D_θ is a bounded, closed and connected region in $\mathbb{R}^d$ without voids. Detailed discussions about these and other geometric properties of D_θ may be found in the proof of Lemma 3.1. One of these properties is that θ is an interior point of D_θ if and only if it can be expressed as $\theta = \theta(p, q) = \mu_y(q) - \mu_x(p)$ for some p and q with straightly positive elements. Correspondingly, a boundary point of D_θ can only be expressed as $\theta(p, q) = \mu_y(q) - \mu_x(p)$ where one or more elements of p and q are zero. This implies that $L(\theta) = 0$ if θ is a boundary point of D_θ and $L(\theta) > 0$ if θ is an interior point of D_θ. We define the domain of the empirical log-likelihood ratio $l(\theta)$ as

$$\Theta_n = \{\theta : \theta \in D_\theta \text{ and } l(\theta) < +\infty\},$$

which excludes the boundary points of D_θ. To differentiate between the $l(\theta)$ in (3.2) and the extended version of $l(\theta)$ in the next section, we will refer to the $l(\theta)$ in (3.2) as the original two-sample empirical log-likelihood ratio or simply "OEL $l(\theta)$". The extended version will be referred to as the "EEL $l^*(\theta)$".

Let $N = m + n$, $f_m = N/m$ and $f_n = N/n$. Without loss of generality, assume that $m \geq n > d$. By the method of Lagrangian multipliers, we have

$$l(\theta_0) = 2\left[\sum_{i=1}^m \log\{1 - f_m\lambda^T(X_i - \mu_x)\} + \sum_{j=1}^n \log\{1 + f_n\lambda^T(Y_j - \mu_y)\}\right] \tag{3.3}$$

where the multiplier $\lambda = \lambda(\theta_0)$ satisfies

$$\sum_{i=1}^m \frac{X_i - \mu_x}{1 - f_m\lambda^T(X_i - \mu_x)} = 0 \quad \text{and} \quad \sum_{j=1}^n \frac{Y_j - \mu_y}{1 + f_n\lambda^T(Y_j - \mu_y)} = 0, \tag{3.4}$$

and

$$\sum_{j=1}^{n} \frac{Y_j}{1 + f_n \lambda^T (Y_j - \mu_y)} - \sum_{i=1}^{m} \frac{X_i}{1 - f_m \lambda^T (X_i - \mu_x)} = \theta_0. \tag{3.5}$$

Under the assumption $(C1)$, Jing (1995) and Liu, Zou and Zhang (2008) showed that

$$l(\theta_0) \xrightarrow{D} \chi_d^2 \quad \text{as} \;\; n \to +\infty. \tag{3.6}$$

Hence, the $100(1-\alpha)\%$ OEL confidence interval for θ_0 is

$$\mathcal{C}_{1-\alpha} = \{\theta : \theta \in \mathbb{R}^d \text{ and } l(\theta) \leq c_\alpha\} \tag{3.7}$$

where c_α is $(1-\alpha)$th quantile of the χ_d^2 distribution. The coverage error of $\mathcal{C}_{1-\alpha}$ is $O(n^{-1})$, that is

$$P(\theta_0 \in \mathcal{C}_{1-\alpha}) = P(l(\theta_0) \leq c_\alpha) = 1 - \alpha + O(n^{-1}). \tag{3.8}$$

Under assumptions $(C1)$, $(C2)$ and $(C3)$, Jing (1995) and Liu, Zou and Zhang (2008) also showed that the OEL $l(\theta)$ is Bartlett correctable, that is

$$P(\theta_0 \in \mathcal{C}'_{1-\alpha}) = P(l(\theta_0) \leq c_\alpha(1 + \eta N^{-1})) = \alpha + O(n^{-2}) \tag{3.9}$$

where $\mathcal{C}'_{1-\alpha} = \{\theta : l(\theta) \leq c_\alpha(1 + \eta N^{-1})\}$ is the Bartlett corrected empirical likelihood (BEL) confidence interval and η is the Bartlett correction constant in Theorem 2 of Liu, Zou and Zhang (2008). For the one-dimensional case, a formula for this constant was first given in Theorem 2 in Jing (1995) but the formula is incomplete (Liu, Zou and Zhang 2008). See also Wu and Yan (2012) and Qin (1994) for discussions about two-sample empirical likelihood methods.

3.3 Two-sample extended empirical likelihood

Like the one-sample empirical likelihood for the mean, the two-sample OEL $l(\theta)$ also suffers from the mismatch problem between its domain and the parameter space since the parameter space is $\mathbb{R}^d$ but $\Theta_n \subset \mathbb{R}^d$. This is a main cause of the undercoverage problem of empirical likelihood confidence regions (Tsao, 2013; Tsao and Wu, 2013). To overcome the mismatch, we now extend the OEL $l(\theta)$ by expanding its domain to the entire $\mathbb{R}^d$.

For simplicity, in addition to $m \geq n$ we further assume that $m/n = O(1)$ so that $O(n^{-1})$, $O(m^{-1})$ and $O(N^{-1})$, for example, are all interchangeable. A point estimator for θ_0 is $\hat{\theta} = \bar{Y} - \bar{X}$ where $\bar{X} = m^{-1}\sum X_i$ and $\bar{Y} = n^{-1}\sum Y_j$ are the sample means. It is easy to verify that $\hat{\theta}$ is the maximum empirical likelihood estimator (MELE) for θ_0. Following Tsao and Wu (2013), we define the composite similarity mapping $h_N^C : \Theta_n \to \mathbb{R}^d$ centred on $\hat{\theta}$ as

$$h_N^C(\theta) = \hat{\theta} + \gamma(N, l(\theta))(\theta - \hat{\theta}) \tag{3.10}$$

where function $\gamma(n, l(\theta))$ is the expansion factor given by

$$\gamma(N, l(\theta)) = 1 + \frac{l(\theta)}{2N}. \tag{3.11}$$

To investigate the properties of the composite similarity mapping h_N^C, we need Lemma 3.1 below which gives two properties of the two-sample OEL $l(\theta)$. For convenience, we denote by $[\hat{\theta}, \theta]$ the line segment that connects $\hat{\theta}$ and θ and by θ_b a boundary point of Θ_n. We have

Lemma 3.1. *The two-sample OEL $l(\theta)$ satisfies: (i) if $\theta \in \Theta_n$ and $\theta' \in [\hat{\theta}, \theta]$, then $l(\theta') \leq l(\theta)$ and (ii) for $\theta \in \Theta_n$, $\lim_{\theta \to \theta_b} l(\theta) = +\infty$.*

Lemma 3.1 shows the two-sample OEL $l(\theta)$ for the difference of two means behaves exactly like its one-sample counterpart for the mean in terms monotonicity and boundary behaviour: it is "monotone increasing" along each ray originating from the MELE and it goes to infinity as θ approaches a boundary point from within Θ_n. Nevertheless, the two-sample and one-sample cases are not entirely the same; the contours of the two-sample OEL may not be convex but that of the one-sample OEL always are. Theorem 3.1 below gives three key properties of composite similarity mapping $h_N^C : \Theta_n \to \mathbb{R}^d$.

Theorem 3.1. *Under the assumption (C1), $h_N^C : \Theta_n \to \mathbb{R}^d$ defined by (3.10) and (3.11) satisfies (i) it has a unique fixed point at $\hat{\theta}$, (ii) it is a similarity mapping for each individual contour of the OEL $l(\theta)$ and (iii) it is a bijective mapping from Θ_n to $\mathbb{R}^d$.*

Since $h_N^C : \Theta_n \to \mathbb{R}^d$ is bijective, it has an inverse function which we denote by $h_N^{-C}(\theta) : \mathbb{R}^d \to \Theta_n$. For any $\theta \in \mathbb{R}^d$, let $\theta' = h_N^{-C}(\theta) \in \Theta_n$. The two-sample extended empirical log-likelihood ratio EEL $l^*(\theta)$ is given by

$$l^*(\theta) = l(h_N^{-C}(\theta)) = l(\theta'), \tag{3.12}$$

which is defined for θ values throughout $\mathbb{R}^d$. Hence the EEL $l^*(\theta)$ is free from the mismatch problem of the OEL $l(\theta)$. Denote by θ_0' the image of θ_0 under the inverse transformation $h_N^{-C}(\theta) : \mathbb{R}^d \to \Theta_n$, that is

$$h_N^{-C}(\theta_0) = \theta_0'. \tag{3.13}$$

Then, the EEL $l^*(\theta)$ evaluated at θ_0 is given by

$$l^*(\theta_0) = l(h_N^{-C}(\theta_0)) = l(\theta_0') = l(\theta_0 + \theta_0' - \theta_0). \tag{3.14}$$

If $|\theta_0' - \theta_0|$ is asymptotically very small, then $l^*(\theta_0)$ will have the same asymptotic distribution as $l(\theta_0)$. Lemma 3.2 below shows that $\theta_0' \in [\hat{\theta}, \theta_0]$ and that $|\theta_0' - \theta_0|$ is indeed asymptotically very small.

Lemma 3.2. *Under assumption (C1), point θ_0' defined by equation (3.13) satisfies (i) $\theta_0' \in [\hat{\theta}, \theta_0]$ and (ii) $\theta_0' - \theta_0 = O_p(n^{-3/2})$.*

Theorem 3.2 below shows that EEL $l^*(\theta_0)$ has the same asymptotic chi-square distribution as the OEL $l(\theta_0)$.

Theorem 3.2. *Under assumption (C1), the EEL $l^*(\theta_0)$ defined by (3.14) satisfies*

$$l^*(\theta_0) \xrightarrow{D} \chi_d^2 \quad as \;\; n \to +\infty. \tag{3.15}$$

By Theorem 3.2, the $100(1-\alpha)\%$ EEL confidence region for θ_0 is

$$\mathcal{C}_{1-\alpha}^* = \{\theta : \theta \in \mathbb{R}^d \text{ and } l^*(\theta) \leq c_\alpha\}, \tag{3.16}$$

which has a coverage error of $O(n^{-1})$. The expansion factor in (3.11) is a convenient choice which also gives good numerical results. There are many other choices available under which Theorems 3.1 and 3.2 also hold. This provides an opportunity to optimize the choice of expansion factor to obtain the second order accuracy. Theorem 3.3 below gives such an optimal choice.

Theorem 3.3. *Under assumptions (C1), (C2) and (C3), and let $l_2^*(\theta)$ be the EEL defined by the composite similarity mapping (3.10) with the following expansion factor*

$$\gamma_2(N, l(\theta)) = 1 + \frac{\eta}{2N} [l(\theta)]^{\delta(n)} \tag{3.17}$$

where $\delta(n) = O(n^{-1/2})$ and η is the Bartlett correction factor for the two-sample empirical likelihood in (3.9). Then, we have

$$l_2^*(\theta_0) = l(\theta_0) \left[1 - \eta/N + O_p(n^{-3/2})\right], \tag{3.18}$$

and

$$P(l_2^*(\theta_0) \leq c) = P(\chi_d^2 \leq c) + O(n^{-2}). \tag{3.19}$$

Replacing EEL $l^*(\theta)$ in (3.16) with $l_2^*(\theta)$ gives an EEL confidence interval which, by (3.19), has a coverage error of $O(n^{-2})$. Because of this, we call $l_2^*(\theta)$ the *second order* EEL or EEL_2. Correspondingly, we call the EEL $l^*(\theta)$ defined by expansion factor (3.11) the *first order* EEL or EEL_1.

3.4 Numerical examples

We now compare the coverage accuracy of 95% confidence regions based on the OEL, BEL and EEL through numerical examples. A referee brought to our attention the adjusted two-sample empirical likelihood (AEL) by Liu and Yu (2010). The AEL is defined on the $\mathbb{R}^d$ and it can also attain the second order accuracy of the BEL. In our numerical comparison, we also include the second order AEL. Comparisons based on 90% and 99% confidence intervals give similar conclusions and are thus not included. They can also be found in Wu and Tsao (2013). In the following, $N(0, 1)$ and $BVN(0, I)$ represent the standard normal and standard bivariate normal distribution, and $X \sim (\chi_1^2, \chi_1^2)^T$, for example, represents a bivariate random vector X whose two elements are independent χ_1^2 random variables.

Example 1: $X \sim N(0,1)$ and $Y \sim N(0,1)$.

Example 2: $X \sim (\chi_1^2, \chi_1^2)^T$ and $Y \sim BVN(0,I)$.

Example 3: $X \sim (\chi_3^2, \chi_3^2)^T$ and $Y \sim (Exp(1), Exp(1))^T$.

To see the effect of the composite similarity mapping, Figure 1 compares contours for the OEL $l(\theta)$ and the corresponding contours for the EEL$_1$ $l(\theta)$ based on the same pair of X and Y samples from Example 2. We see that the contours in the two plots are identical in shape and the contours in both plots are centred on the MELE $\hat{\theta}$ as indicated in Theorem 3.1. Further, at any fixed level, the contour of the EEL $l^*(\theta)$ is larger in scale.

Simulated coverage probabilities for the three examples are given in Tables 1, 2 and 3, respectively. Each simulated probability in the tables is based on 10,000 pairs of random samples whose sizes are indicated by the row and column headings, respectively. The BEL, AEL and EEL$_2$ were computed by using the estimated Bartlett correction factor from page 1705 in Liu and Yu (2010). We summarize the tables with the following observations: (1) EEL$_1$ is consistently more accurate than the OEL. Surprisingly, in most cases it is also more accurate than the second order BEL and AEL. (2) EEL$_2$ is more accurate than OEL, BEL and AEL for small and moderate sample sizes. It is comparable to BEL and AEL when one or both sample sizes are large. (3) EEL$_1$ is slightly more accurate than EEL$_2$ overall.

Example 4: A real-data example. Interlining fabrics are used to support outer fabrics in order to create and maintain the shape and drape of different clothes. Fan, Leeuwner, and Hunter (1997) gave a method for selecting compatible fusible interlinings for different outer fabrics based on several variables. One of these variables is

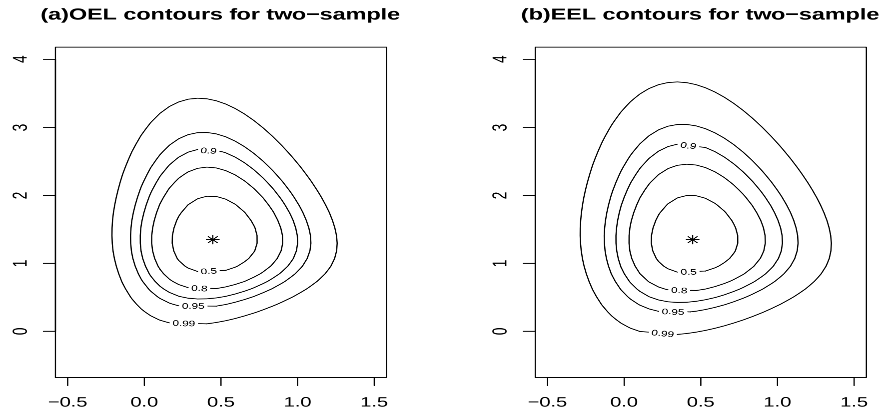

Figure 3.1: (a) Two-sample OEL contours; (b) Two-sample EEL contours. Both plots are based the same pair of X and Y samples from Example 2 with sample size $n = 20$ and $m = 20$. The star in the middle of the plot is the MELE $\hat{\theta}$. EEL_1 contours are larger than but similar to OEL contours with the same centre and identical shape.

Table 3.1: : Coverage probabilities of 95% OEL, EEL_1, BEL, AEL & EEL_2 confidence intervals for Example 1: $X \sim N(0,1)$ and $Y \sim N(0,1)$

		n=10	n=20	n=30	n=40
m=10	OEL	92.5	92.7	92.5	92.4
	EEL_1	94.5	94.3	93.6	93.3
	BEL	93.6	93.8	93.5	93.4
	AEL	94.0	94.4	94.0	93.9
	EEL_2	94.3	94.4	94.0	93.7
m=20	OEL	92.7	93.7	94.1	94.0
	EEL_1	94.3	94.9	95.0	94.7
	BEL	93.8	94.5	94.7	94.6
	AEL	94.2	94.6	94.8	94.6
	EEL_2	94.4	94.7	94.9	94.8
m=30	OEL	92.2	94.2	94.0	94.8
	EEL_1	93.4	95.2	94.7	95.3
	BEL	93.3	94.8	94.4	95.2
	AEL	93.8	94.8	94.5	95.2
	EEL_2	93.8	95.0	94.5	95.2
m=40	OEL	92.0	94.0	94.2	94.5
	EEL_1	92.9	94.8	94.9	95.0
	BEL	93.0	94.6	94.7	94.9
	AEL	93.6	94.7	94.8	94.9
	EEL_2	93.6	94.8	94.8	95.0

Table 3.2: : Coverage probabilities of 95% OEL, EEL_1, BEL, AEL & EEL_2 confidence regions for Example 2: $X \sim (\chi_1^2, \chi_1^2)$ and $Y \sim BVN(0, I)$

		n=10	n=20	n=30	n=40
m=10	OEL	84.2	89.0	89.7	88.7
	EEL_1	90.2	92.7	92.6	91.0
	BEL	86.9	90.9	91.8	90.7
	AEL	89.2	92.2	92.9	92.1
	EEL_2	89.1	92.5	93.0	92.2
m=20	OEL	82.8	89.6	91.6	92.8
	EEL_1	87.7	92.5	93.8	94.2
	BEL	85.7	91.2	93.1	93.8
	AEL	88.8	91.7	93.2	93.9
	EEL_2	87.7	92.1	93.6	94.2
m=30	OEL	80.3	89.2	91.9	92.5
	EEL_1	83.8	91.4	93.7	93.9
	BEL	83.0	90.6	93.1	93.4
	AEL	86.6	91.1	93.3	93.5
	EEL_2	84.8	91.3	93.5	93.7
m=40	OEL	78.8	88.5	90.9	92.3
	EEL_1	81.7	90.4	92.8	93.6
	BEL	81.3	89.9	92.2	93.1
	AEL	85.5	90.7	92.5	93.2
	EEL_2	83.2	90.7	92.7	93.4

Table 3.3: : Coverage probabilities of 95% OEL, EEL_1, BEL, AEL & EEL_2 confidence regions for Example 3: $X \sim (\chi^2_3, \chi^2_3)$ $Y \sim (Exp(1), Exp(1))$

		n=10	n=20	n=30	n=40
m=10	OEL	81.5	89.4	91.3	91.1
	EEL_1	89.2	93.0	93.9	93.6
	BEL	84.6	91.2	92.8	92.7
	AEL	87.5	91.8	93.1	93.5
	EEL_2	88.3	92.7	93.9	93.9
m=20	OEL	81.1	89.7	92.3	92.7
	EEL_1	86.0	92.6	94.3	94.5
	BEL	83.7	91.4	93.6	93.7
	AEL	86.8	91.8	93.7	93.9
	EEL_2	86.5	92.2	94.1	94.3
m=30	OEL	78.1	89.1	91.9	92.9
	EEL_1	84.2	91.4	93.5	94.3
	BEL	80.4	90.5	93.0	93.7
	AEL	83.9	91.1	93.1	93.8
	EEL_2	85.4	91.4	93.3	94.1
m=40	OEL	79.9	89.4	91.6	93.1
	EEL_1	82.9	91.5	93.4	94.3
	BEL	82.2	90.8	92.8	93.9
	AEL	85.1	91.3	93.0	94.0
	EEL_2	84.9	91.7	93.2	94.2

the extensibility of interlining fabrics. A dataset on page 139 of this paper contains percentage extensibility for 24 high and 23 medium quality fabrics. We used empirical likelihood methods to construct confidence intervals for the difference between the mean extensibility of these two grades of fabrics. The 95% OEL, EEL_1, BEL, AEL and EEL_2 confidence intervals are, respectively, $[-0.203, 0.274]$, $[-0.213, 0.338]$, $[-0.211, 0.281]$, $[-0.211, 0.281]$ and $[-0.212, 0.287]$. These results are consistent with the findings from the simulation results above; for example, EEL_1 interval is the widest which is consistent with the finding that EEL_1 has in general a higher coverage probability than other methods.

To conclude, EEL_1 is easy-to-compute and is the most accurate overall. Hence, we recommend EEL_1 for two-sample problems.

3.5 Supplement Material

Acknowledgement: we acknowledge that this section contains the supplement material to the published paper Tsao and Wu (2014) , Statistics and Probability Letters, 2014, vol. 84, issue C, pages 81-87 and is available at Statistics and Probability Letters online.

The following section includes the detailed proofs for the two lemmas and three theorems in the paper. It also includes simulated coverage probabilities for 90% and 99% confidence intervals for Examples 1 and 2 in Section 4.

3.5.1 Proofs of lemmas and theorems

Proof of Lemma 3.1. To prove (i), it is more convenient to work with the empirical likelihood ratio $L(\theta)$ instead of the log-likelihood ratio $l(\theta)$. Let $L_x(\mu_1)$ and $L_y(\mu_2)$ be the one-sample empirical likelihood ratios for the mean based the X and

Y samples, respectively. Then, for any $\theta \in \Theta_n$, the two-sample empirical likelihood ratio $L(\theta)$ in (3.1) can be expressed as

$$L(\theta) = \max_{(\mu_1,\mu_2)} \{L_x(\mu_1)L_y(\mu_2) : \mu_2 - \mu_1 = \theta\}. \tag{3.20}$$

In terms of $L(\theta)$, (i) is equivalent to $L(\theta') \geq L(\theta)$ for $\theta' \in [\hat{\theta}, \theta]$. In order to show this, it suffices to show that for any pair (μ_1, μ_2) such that $\mu_2 - \mu_1 = \theta$ there exists a pair (μ_1', μ_2') such that $\mu_2' - \mu_1' = \theta'$ and

$$L_x(\mu_1')L_y(\mu_2') \geq L_x(\mu_1)L_y(\mu_2). \tag{3.21}$$

To show (3.21), since $\theta \in \Theta_n$, there exist $\mu_1 = \mu_x(p)$ and $\mu_2 = \mu_y(q)$ such that $\theta = \theta(p,q) = \mu_y(q) - \mu_x(p)$. Without loss of generality, suppose $\theta \neq \hat{\theta}$ and consider only $\theta' \in (\hat{\theta}, \theta)$. By $\theta' \in (\hat{\theta}, \theta)$, there exists a $\beta \in (0,1)$ such that $\theta' = \beta\theta + (1-\beta)\hat{\theta} = \beta(\mu_y(q) - \mu_x(p)) + (1-\beta)(\bar{Y} - \bar{X})$, that is,

$$\begin{aligned}
\theta' &= \beta\left(\sum_{j=1}^{n} q_j Y_j - \sum_{i=1}^{m} p_i X_i\right) + (1-\beta)\left(n^{-1}\sum_{j=1}^{n} Y_j - m^{-1}\sum_{i=1}^{m} X_i\right) \\
&= \sum_{j=1}^{n}[\beta q_j + (1-\beta)n^{-1}]Y_j - \sum_{i=1}^{m}[\beta p_i + (1-\beta)m^{-1}]X_i \\
&= \sum_{j=1}^{n} q_j' Y_j - \sum_{i=1}^{m} p_i' X_i,
\end{aligned}$$

where $p_i' = \beta p_i + (1-\beta)m^{-1}$ for $i = 1, 2, \ldots, m$ and $q_j' = \beta q_j + (1-\beta)n^{-1}$ for $j = 1, 2, \ldots, n$. Since $0 \leq p_i \leq 1$, $0 < \beta < 1$ and $\sum p_i = 1$, we have $0 < p_i' < 1$ and $\sum p_i' = 1$. Similarly, $0 < q_j' < 1$ and $\sum q_j' = 1$. Thus the point $\theta' = \theta(p', q')$ where $p' = (p_1', p_2', \ldots, p_m')$ and $q' = (q_1', q_2', \ldots, q_n')$ satisfies $\theta' \in \Theta_n$. Letting $\mu_1' = \mu_x(p')$ and $\mu_2' = \mu_y(q')$, we have $\mu_2' - \mu_1' = \theta'$, $\mu_1' \in (\bar{X}, \mu_1)$ and $\mu_2' \in (\bar{Y}, \mu_2)$. It follows from Lemma 3.1 in Tsao and Wu (2013) that $L_x(\mu_1') \geq L_x(\mu_1)$ and $L_y(\mu_2') \geq L_y(\mu_2)$.

These imply (3.21) and prove (i).

The above proof also shows that (a) for any $\theta \in \Theta_n$, $[\hat{\theta}, \theta] \subset \Theta_n$ and (b) along the ray originating from $\hat{\theta}$ and through θ, each point θ' in the interior point of Θ_n can be expressed as $\theta' = \theta(p', q')$ for some p' and q' satisfying $0 < p'_i < 1$, $\sum p'_i = 1$, $0 < q'_j < 1$ and $\sum q'_j = 1$. Hence (b) implies $L(\theta') > 0$ for all such interior points. That Θ_n has a boundary point on the ray is clear from the boundedness of Θ_n which is due to the boundedness of both $\mu_x(p)$ and $\mu_y(q)$. Denoting by θ_b the boundary point on this ray, (a) implies this boundary point is unique. Further, for any pair (p_b, q_b) such that $\theta_b = \theta(p_b, q_b)$, one or more elements of p_b and q_b must be zero. To see this, if all of their elements p_{bi} $(i = 1, 2, \ldots, m)$ and q_{bj} $(j = 1, 2, \ldots, n)$ are straightly positive, we can find a β sufficiently close but not equal to 1 such that $0 < p^*_i = \beta^{-1} p_{bi} - \beta^{-1}(1-\beta)m^{-1} < 1$ and $0 < q^*_j = \beta^{-1} q_{bj} - \beta^{-1}(1-\beta)n^{-1} < 1$ uniformly for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. It is easy to verify that $\sum p^*_i = \sum q^*_j = 1$. Hence, $\theta^* = \theta(p^*, q^*) \in \Theta_n$. Noting that $p_{bi} = \beta p^*_i + (1-\beta)m^{-1}$ for $i = 1, 2, \ldots, m$ and $q_{bj} = \beta q^*_j + (1-\beta)n^{-1}$ for $j = 1, 2, \ldots, n$, we have $\theta_b = \theta(p_b, q_b) \in (\hat{\theta}, \theta^*)$ which shows θ_b is a *not* the boundary point on the ray from $\hat{\theta}$ and through θ_b, contradicting the assumption that θ_b is the boundary point. Hence, a boundary point of Θ_n can only be expressed as $\theta_b = \theta(p_b, q_b)$ where one or more elements of p_b and q_b are zero. This implies $L_x(\mu(p_b)) = 0$ and/or $L_y(\mu(q_b)) = 0$. It follows that $L(\theta_b) = 0$, $l(\theta_b) = +\infty$ and $\theta_b \in D_\theta$ but $\theta_b \notin \Theta_n$.

To summarize, the domain Θ_n of the OEL $l(\theta)$ is connected without voids since every $\theta \in \Theta_n$ is connected to $\hat{\theta}$ through a line segment $[\hat{\theta}, \theta] \subset \Theta_n$. Also, Θ_n is open since the boundary points of this connected region do not belong to the region. It is bounded due to the boundedness of $\mu_x(p)$ and $\mu_y(q)$. Along any ray originating from the MELE $\hat{\theta}$, there is a unique boundary point θ_b at which $L(\theta_b) = 0$ and $l(\theta_b) = +\infty$. The domain of $L(\theta)$, D_θ, is the closure of Θ_n. By the continuity of $L(\theta)$ over D_θ, we have (ii). □

Proof of Theorem 3.1. Proofs of parts (i) and (ii) of Theorem 3.1 follow easily from that for parts (i) and (ii) of Theorem 3.1 in Tsao and Wu (2013). Noting that part (i) of Lemma 3.1 implies the contours of the two-sample OEL $l(\theta)$ are nested around the MELE $\hat{\theta}$, the proof of part (iii) of Theorem 3.1 also follows from that for part (iii) of Theorem 3.1 in Tsao and Fan (2013). □

Proof of Lemma 3.2. Differentiating EEL $l(\theta)$ and evaluating the derivative at θ_0, we find

$$J(\theta_0) = \left[\frac{\partial l(\theta)}{\partial \theta}\right]_{\theta=\theta_0} = -2N\lambda^T(\theta_0). \tag{3.22}$$

Under the conditions of the lemma, we can show that $\lambda(\theta_0) = O_p(n^{-1/2})$ and $J(\theta_0) = O_p(n^{1/2})$. Also, applying Taylor expansion to $l(\theta)$ in a small neighbourhood of θ_0, $\mathcal{N}(\theta_0) = \{\theta : |\theta - \theta_0| \leq \kappa n^{-1/2}\}$, where κ is some positive constant, we obtain

$$l(\theta) = l(\theta_0 + (\theta - \theta_0)) = l(\theta_0) + J(\theta_0)(\theta - \theta_0) + O_p(1). \tag{3.23}$$

By Theorem 1 in Liu, Zou and Zhang (2008), $l(\theta_0) = O_p(1)$. This and (3.23) imply that

$$l(\theta) = O_p(1) \tag{3.24}$$

uniformly for $\theta \in \mathcal{N}(\theta_0)$. Since $h_N^C(\theta_0') = \theta_0$, by (3.10) we have

$$\theta_0 - \hat{\theta} = \gamma(N, l(\theta_0'))(\theta_0' - \hat{\theta}). \tag{3.25}$$

This and the fact that $\gamma(N, l(\theta)) \geq 1$ imply $\theta_0' \in [\hat{\theta}, \theta_0]$ which proves part (i) of Lemma 3.2. To show part (ii), by $\hat{\theta} - \theta_0 = O_p(n^{-1/2})$ and part (i), we have $\theta_0' - \theta_0 = O_p(n^{-1/2})$.

This and (3.24) imply

$$\gamma(N, l(\theta_0')) = 1 + \frac{l(\theta_0')}{2N} = 1 + O_p(n^{-1}). \tag{3.26}$$

Adding and subtracting a θ_0 on the right-hand side of (3.25) and simplifying, we find that

$$[1 - \gamma(N, l(\theta_0'))]\,(\theta_0 - \hat{\theta}) = \gamma(N, l(\theta_0'))(\theta_0' - \theta_0). \tag{3.27}$$

By (3.26), (3.27) and $\hat{\theta} - \theta_0 = O_p(n^{-1/2})$, we have

$$\theta_0' - \theta_0 = O_p(n^{-3/2}) \tag{3.28}$$

which proves part (ii) of Lemma 3.2. □

Proof of Theorem 3.2. By part (ii) of Lemma 3.2, $\theta_0' - \theta_0 = O_p(n^{-3/2})$. Applying Taylor expansion to $l^*(\theta_0) = l(\theta_0 + (\theta_0' - \theta_0))$, we obtain

$$\begin{aligned} l^*(\theta_0) &= l(\theta_0) + J(\theta_0)(\theta' - \theta_0) + o_p(n^{-3/2}) \\ &= l(\theta_0) + O_p(n^{-1}). \end{aligned} \tag{3.29}$$

This and $l(\theta_0) \to \chi^2_d$ in distribution imply Theorem 3.2. □

Proof of Theorem 3.3. Under assumptions $(C1)$, $(C2)$ and $(C3)$, based on the equation (2.4) on page 550 in Liu, Zou and Zhang (2008), we expand $l(\theta_0)$ as

$$l(\theta_0) = N(R_1 + R_2 + R_3)^T(R_1 + R_2 + R_3) + N\Delta + O_p(n^{-3/2}) \tag{3.30}$$

where R_i and Δ are functions of $G^{t_1t_2\ldots t_l}$ and $G_1^{t_1t_2\ldots t_l}$ with

$$
\begin{aligned}
R_1^r &= G^r, \quad R_2^r = -\frac{1}{2}G^{rs}G^s + \frac{1}{3}g^{rst}G^sG^t, && (3.31)\\
R_3^r &= \frac{3}{8}G^{rs}G^{st}G^t - \frac{5}{12}g^{rst}G^{tu}G^sG^u - \frac{5}{12}g^{stu}G^{rs}G^tG^u \\
&\quad + \frac{4}{9}g^{rst}g^{tuv}G^sG^uG^v + \frac{1}{3}G^{rst}G^sG^t - \frac{1}{4}g^{rstu}G^sG^tG^u, && (3.32)\\
\Delta &= (G^{rs} - G_1^{rs})G^rG^{s+} + \frac{2}{3}(G_1^{ruv} - G^{ruv})G^rG^uG^v, && (3.33)
\end{aligned}
$$

where for a vector P, P^r means its rth component. Expressions for $g^{t_1t_2\ldots t_l}$, $G^{t_1t_2\ldots t_l}$ and $G_1^{t_1t_2\ldots t_l}$ may be found in Liu, Zou and Zhang (2008). Let $V = f_mCov(X) + f_nCov(Y)$. Based on the expressions for λ_1^* and λ_2^* on page 553 in Liu, Zou and Zhang (2008), we have

$$\lambda(\theta_0) = V^{-1}D_1 + O_p(n^{-1}), \tag{3.34}$$

where $D_1 = n^{-1}\sum_{j=1}^{n}(y_j - \mu_y) - m^{-1}\sum_{i=1}^{m}(x_i - \mu_x)$. Noting that $G^{t_1t_2\ldots t_l} = O_p(n^{-1/2})$ and $G_1^{t_1t_2\ldots t_l} = O_p(n^{-1/2})$, by (3.31) to (3.34), we have

$$(i) \quad R_j = O_p(n^{-j/2}) \quad \text{for } j = 1, 2, 3, \tag{3.35}$$

$$(ii) \quad D_1 = \hat{\theta} - \theta_0 = O_p(n^{-1/2}), \tag{3.36}$$

$$(iii) \quad R_1^TR_1 = D_1^TV^{-1}D_1, \tag{3.37}$$

$$(iv) \quad \Delta = O_p(n^{-3/2}). \tag{3.38}$$

It may be verified that Lemma 3.2, Theorem 3.1 and Theorem 3.2 all hold under $\gamma_2(N, l(\theta))$. Hence, the limiting distribution of $l_2^*(\theta_0)$ is also χ_d^2. This and the condition that $\delta(n) = O(n^{-1/2})$ imply

$$[l(\theta_0')]^{\delta(n)} = 1 + O_p(n^{-1/2}). \tag{3.39}$$

Since $h_n^C(\theta_0') = \theta_0$, by (3.10) and (3.17), we have

$$\begin{aligned}
\theta_0' - \theta_0 &= \frac{\eta[l(\theta_0')]^{\delta(n)}}{2N}(\hat{\theta} - \theta_0') \\
&= \frac{\eta[l(\theta_0')]^{\delta(n)}}{2N}(\hat{\theta} - \theta_0) + \frac{\eta[l(\theta_0')]^{\delta(n)}}{2N}(\theta_0 - \theta_0'). \qquad (3.40)
\end{aligned}$$

By the part (ii) of Lemma 3.2, (3.39) and (3.40), we find that

$$\begin{aligned}
\theta_0' - \theta_0 &= \frac{\eta[l(\theta_0')]^{\delta(n)}}{2N}(\hat{\theta} - \theta_0) + O_p(n^{-5/2}) \\
&= \frac{\eta}{2N}(\hat{\theta} - \theta_0) + O_p(n^{-2}). \qquad (3.41)
\end{aligned}$$

It follows from (3.22), (3.29) and (3.41) that

$$\begin{aligned}
l_2^*(\theta_0) &= l(\theta_0) + J(\theta_0)(\theta' - \theta_0) + o_p(n^{-3/2}) \\
&= l(\theta_0) + \frac{\eta}{2N}J(\theta_0)(\hat{\theta} - \theta_0) + o_p(n^{-3/2}) \\
&= l(\theta_0) - \frac{\eta}{2N}\left[2N\lambda^T(\theta_0)\right](\hat{\theta} - \theta_0) + o_p(n^{-3/2}) \\
&= l(\theta_0) - \frac{\eta}{N}\left[N\lambda^T(\theta_0)(\hat{\theta} - \theta_0)\right] + o_p(n^{-3/2}). \qquad (3.42)
\end{aligned}$$

Finally, by (3.30), and (3.35) to (3.38), we obtain

$$\begin{aligned}
l_2^*(\theta_0) &= l(\theta_0) - \frac{\eta}{N}\left[N(V^{-1}D_1 + O_p(n^{-1}))^T D_1\right] + o_p(n^{-3/2}) \\
&= l(\theta_0) - \frac{\eta}{N}\left[ND_1^T V^{-1} D_1\right] + O_p(n^{-3/2}) \\
&= l(\theta_0) - \frac{\eta}{N}\left[NR_1^T R_1\right] + O_p(n^{-3/2}) \\
&= l(\theta_0) - \frac{\eta}{N}\left[N(R_1 + R_2 + R_3)^T(R_1 + R_2 + R_3) + N\Delta\right] + O_p(n^{-3/2}) \\
&= l(\theta_0) - \frac{\eta}{N}l(\theta_0) + O_p(n^{-3/2}) \\
&= l(\theta_0)\{1 - \frac{\eta}{N} + O_p(n^{-3/2})\}, \qquad (3.43)
\end{aligned}$$

which proves (3.18) in Theorem 3.3. Equation (3.19) in the theorem follows from equation (3.18) and Theorem 2 in Liu, Zou and Zhang (2008). $\square$

3.5.2 90% and 99% confidence intervals for Examples 1 and 2

The following tables compare simulated coverage probabilities of 90% and 99% confidence intervals based on the OEL, EEL_1, BEL and EEL_2. Observations from Section 4 based on 95% confidence intervals remain valid: (1) EEL_1 is consistently more accurate than the OEL. It is also more accurate than the second order BEL and EEL_2 for small and moderate sample sizes ($n, m \leq 20$) and competitive in accuracy when sample sizes are larger. (2) EEL_2 is more accurate than OEL and BEL for small and moderate sample size. It is comparable to BEL when one or both sample sizes are large.

Table 3.4: : Coverage probabilities of 90% OEL, EEL_1, BEL, AEL & EEL_2 confidence intervals: $X \sim N(0,1)$ and $Y \sim N(0,1)$

		n=10	n=20	n=30	n=40
m=10	OEL	86.8	87.6	87.1	86.8
	EEL_1	89.1	89.0	88.2	87.7
	BEL	88.6	88.9	88.5	88.2
	AEL	89.1	89.6	89.0	88.9
	EEL_2	89.2	89.5	89.1	88.7
m=20	OEL	86.8	88.2	89.0	88.8
	EEL_1	88.5	89.4	89.8	89.7
	BEL	88.5	89.3	89.8	89.8
	AEL	89.1	89.3	89.8	89.8
	EEL_2	89.1	89.5	90.0	90.0
m=30	OEL	86.5	89.2	88.8	89.6
	EEL_1	88.0	90.2	89.5	90.3
	BEL	88.3	90.2	89.4	90.2
	AEL	88.9	90.3	89.5	90.2
	EEL_2	88.8	90.4	89.6	90.3
m=40	OEL	86.6	88.5	89.2	89.2
	EEL_1	87.7	89.3	89.8	89.7
	BEL	88.2	89.4	89.8	89.7
	AEL	88.7	89.5	89.9	89.7
	EEL_2	88.6	89.5	89.9	89.8

Table 3.5: : Coverage probabilities of 99% OEL, EEL_1, BEL, AEL & EEL_2 confidence intervals: $X \sim N(0,1)$ and $Y \sim N(0,1)$

		n=10	n=20	n=30	n=40
m=10	OEL	97.5	97.8	97.5	97.7
	EEL_1	98.9	98.8	98.4	98.3
	BEL	98.1	98.2	98.0	98.1
	AEL	98.3	98.5	98.3	98.5
	EEL_2	98.4	98.5	98.3	98.4
m=20	OEL	97.7	98.6	98.6	98.5
	EEL_1	98.7	99.2	99.1	99.0
	BEL	98.1	98.8	98.8	98.7
	AEL	98.4	98.9	98.8	98.8
	EEL_2	98.5	99.0	98.9	98.8
m=30	OEL	97.4	98.6	98.7	98.8
	EEL_1	98.3	99.1	99.1	99.1
	BEL	97.9	98.8	98.9	98.9
	AEL	98.2	98.8	98.9	98.9
	EEL_2	98.2	98.9	99.0	98.9
m=40	OEL	97.5	98.5	98.6	98.8
	EEL_1	98.1	98.8	99.0	99.2
	BEL	97.9	98.7	98.8	98.9
	AEL	98.3	98.7	98.8	98.9
	EEL_2	98.2	98.7	98.8	99.0

Table 3.6: : Coverage probabilities of 90% OEL, EEL_1, BEL, AEL & EEL_2 confidence intervals: $X \sim (\chi_1^2, \chi_1^2)$ and $Y \sim BVN(0, I)$

		n=10	n=20	n=30	n=40
m=10	OEL	77.3	82.4	83.7	82.6
	EEL_1	83.4	86.5	86.4	85.0
	BEL	80.6	85.2	85.9	85.3
	AEL	83.1	86.4	87.1	86.4
	EEL_2	83.2	87.0	87.5	86.5
m=20	OEL	75.9	83.6	85.9	86.5
	EEL_1	80.0	86.3	88.1	88.2
	BEL	78.8	85.5	87.6	88.0
	AEL	81.7	86.2	87.8	88.2
	EEL_2	80.9	86.5	88.3	88.6
m=30	OEL	73.2	83.1	85.9	86.5
	EEL_1	76.4	85.2	87.7	88.2
	BEL	76.2	85.0	87.4	87.9
	AEL	79.7	85.6	87.7	88.0
	EEL_2	78.2	85.7	88.0	88.5
m=40	OEL	71.6	82.1	84.8	86.6
	EEL_1	74.4	84.1	86.5	88.0
	BEL	74.9	84.1	86.4	87.9
	AEL	78.9	85.0	86.5	87.9
	EEL_2	76.9	85.0	86.8	88.1

Table 3.7: : Coverage probabilities of 99% OEL, EEL_1, BEL, AEL & EEL_2 confidence intervals: $X \sim (\chi_1^2, \chi_1^2)$ and $Y \sim BVN(0, I)$

		n=10	n=20	n=30	n=40
m=10	OEL	92.2	95.8	96.0	96.0
	EEL_1	96.9	98.1	97.8	97.6
	BEL	93.7	96.8	96.9	97.0
	AEL	95.4	97.2	97.6	97.6
	EEL_2	95.4	97.6	97.6	97.7
m=20	OEL	92.2	96.4	97.4	98.0
	EEL_1	95.4	98.3	98.7	98.9
	BEL	93.4	97.2	98.0	98.4
	AEL	95.7	97.4	98.1	98.5
	EEL_2	94.7	97.6	98.3	98.6
m=30	OEL	89.9	96.2	97.6	98.0
	EEL_1	93.1	97.8	98.6	98.7
	BEL	91.5	97.0	98.1	98.3
	AEL	95.0	97.4	98.2	98.4
	EEL_2	92.8	97.4	98.3	98.5
m=40	OEL	88.0	95.5	97.4	97.8
	EEL_1	90.9	97.0	98.2	98.6
	BEL	89.8	96.2	97.9	98.2
	AEL	93.9	96.7	98.0	98.3
	EEL_2	91.3	96.7	98.0	98.3

Table 3.8: : Coverage probabilities of 90% OEL, EEL_1, BEL, AEL & EEL_2 confidence intervals: $X \sim (\chi_3^2, \chi_3^2)$ and $Y \sim (Exp(1), Exp(1))$

		n=10	n=20	n=30	n=40
m=10	OEL	75.2	83.3	85.4	85.0
	EEL_1	82.0	87.3	88.2	87.1
	BEL	78.6	86.0	87.4	86.8
	AEL	80.7	86.7	88.2	87.9
	EEL_2	81.6	87.4	88.8	88.0
m=20	OEL	74.2	84.3	86.7	87.0
	EEL_1	79.2	86.8	88.7	88.8
	BEL	77.7	86.2	88.2	88.5
	AEL	80.3	86.6	88.4	88.6
	EEL_2	80.5	87.0	88.8	88.9
m=30	OEL	72.3	83.2	85.8	87.2
	EEL_1	77.7	85.5	87.9	88.8
	BEL	75.1	85.2	87.6	88.5
	AEL	78.3	85.8	87.7	88.6
	EEL_2	79.5	86.1	88.0	89.0
m=40	OEL	73.8	83.6	86.2	87.6
	EEL_1	76.5	85.5	87.7	88.9
	BEL	76.8	85.5	87.6	88.8
	AEL	79.2	85.9	87.8	88.8
	EEL_2	79.1	86.5	88.0	89.1

Table 3.9: : Coverage probabilities of 99% OEL, EEL_1, BEL, AEL & EEL_2 confidence intervals: $X \sim (\chi_3^2, \chi_3^2)$ and $Y \sim (Exp(1), Exp(1))$

		n=10	n=20	n=30	n=40
m=10	OEL	89.5	95.7	97.0	97.1
	EEL_1	96.2	98.2	98.8	98.6
	BEL	91.0	96.5	97.8	97.8
	AEL	94.4	97.1	98.1	98.2
	EEL_2	94.6	97.8	98.5	98.5
m=20	OEL	89.3	96.0	97.7	98.1
	EEL_1	93.9	97.9	98.8	98.8
	BEL	90.9	96.7	98.2	98.4
	AEL	94.5	97.0	98.2	98.5
	EEL_2	93.1	97.3	98.5	98.6
m=30	OEL	85.0	95.5	97.4	98.2
	EEL_1	91.9	97.3	98.4	98.8
	BEL	86.4	96.3	97.9	98.5
	AEL	91.4	96.8	97.9	98.5
	EEL_2	92.0	97.0	98.0	98.5
m=40	OEL	87.8	95.7	97.5	97.9
	EEL_1	91.0	97.1	98.4	98.5
	BEL	89.4	96.3	98.0	98.2
	AEL	93.3	96.6	98.1	98.3
	EEL_2	91.7	97.0	98.2	98.3

Chapter 4

Two-sample empirical likelihood for estimating equations

4.1 Introduction

A two-sample problem is concerned with making inference for the difference between the corresponding parameters of two populations/models with two independent samples. The difference between two population means is a special case that has been extensively studied; when the sample sizes are not large and underlying distributions are normal, methods for the Behren-Fisher problem or a two-sample t method can be used; when the sample sizes are large, non-parametric z based procedures can be used. Recently, the empirical likelihood method (Owen, 2001) has been successfully applied to this special case. See Jing (1995), Liu, Zou and Zhang (2008), Liu and Yu (2010), Wu and Yan (2012) and Wu and Tsao (2013). These empirical likelihood methods complement existing methods as they do not require strong conditions and are more accurate than normal approximation based methods when the underlying distributions are skewed. In particular, the extended two-sample empirical likelihood for the difference between two p-dimensional means (Wu and Tsao, 2013) is defined on the whole of $\mathbb{R}^p$ and is more accurate than other empirical likelihood methods.

In this paper, we study empirical likelihood methods for the general two-sample problem concerning the difference between two p-dimensional parameters defined by general estimating equations. The main contribution of this paper is a new extended empirical likelihood for such a difference, which generalizes results of Tsao (2013) and Tsao and Wu (2013) to this two-sample problem. The empirical likelihood method was introduced by Owen (1988, 1990). It has since been applied to many problems in statistics; see Owen (2001) and references therein. In particular, Qin and Lawless (1994) showed that the empirical likelihood is effective for inference on parameters defined by estimating equations. DiCiccio, Hall and Romano (1991) and Chen and Cui (2007) proved that the empirical likelihood for estimating equations is Bartlett correctable; the Bartlett corrected empirical likelihood enjoys the second-order accuracy. Although there have been relatively few publications that apply empirical likelihood to the general two-sample problem, it is well-suited for this problem as the formulation of the one-sample empirical likelihood for estimating equations can be readily extended to handle the two-sample case; see, *e.g.,* Jing (1995), Qin and Zhao (2000), Liu, Zou and Zhang (2008), Liu and Yu (2010), Wu and Yan (2012) and Zi, Zou and Liu (2012). In particular, Qin and Zhao (2000) studied the standard two-sample empirical likelihood for the univariate version ($p = 1$) of the problem, and Zi, Zou and Liu (2012) considered the special case where the parameters are the coefficient-vectors of two linear models.

In Section 2, we study the standard two-sample empirical likelihood for estimating equations in the general multi-dimensional sitting where $p \geq 1$. Like its one-sample counterpart, this two-sample empirical likelihood also has an asymptotic chi-square distribution and is Bartlett correctable. Adopting the terminology in Tsao and Wu (2013), we refer to this standard two-sample empirical likelihood as the two-sample original empirical likelihood (OEL) for estimating equations. The OEL suffers from a mismatch problem (Tsao and Wu, 2013) in that it is only defined on a part of the parameter space. This problem affects the coverage accuracy of the OEL based confidence regions. To overcome this, in Section 3 we introduce a two-sample extended empirical likelihood (EEL) that is defined on the whole parameter space. The EEL is obtained by expanding the domain of the OEL to the full parameter space through a composite similarity mapping. We show that the EEL has the same asymptotic chi-square distribution as the OEL and that it can also achieve the second-order accuracy of the Bartlett correction. In Section 4, we discuss two applications of the two-sample OEL and EEL. The first application is concerned with the inference for the difference between two Gini indices, and the second application is concerned with that between coefficient vectors of two regression models. We also make use of these applications to compare the numerical accuracy of the OEL and EEL confidence regions and to illustrate the superior accuracy of the EEL.

Proofs of theoretical results on two-sample OEL and EEL are all relegated to the Appendix. Note that some of these results can be proved by slightly modifying the proofs of similar results for other empirical likelihoods in the literature. For brevity, we will not include detailed proofs for such results in the Appendix but will give relevant references containing similar proofs.

4.2 Two-sample original empirical likelihood (OEL) for estimating equations

We first describe the general two-sample problem for estimating equations as follows. Let $X \in \mathbb{R}^d$ and $Y \in \mathbb{R}^d$ be two random vectors with unknown parameters $\theta_{x_0} \in \mathbb{R}^p$ and $\theta_{y_0} \in \mathbb{R}^p$, respectively. Let $g(X, \theta_x)$ and $g(Y, \theta_y)$ be two q-dimensional estimating functions for θ_{x_0} and θ_{y_0} satisfying $E\{g(X, \theta_{x_0})\} = 0$ and $E\{g(Y, \theta_{y_0})\} = 0$, respectively. The unknown parameter of interest is the difference $\pi_0 = \theta_{y_0} - \theta_{x_0} \in \mathbb{R}^p$ and the parameter space is the entire $\mathbb{R}^p$. A more general version of this problem allows the estimating function for θ_{x_0} to be different from that for θ_{y_0}. For simplicity, we consider only the common case where the two estimating functions are the same. We assume that $\{X_1, \ldots, X_m\}$ are independent copies of X, $\{Y_1, \ldots, Y_n\}$ are independent copies of Y, and X_i and Y_j are independent.

We now generalize the one-sample OEL for estimating equations (Qin and Lawless, 1994) to obtain a two-sample OEL for π_0 and study its asymptotic properties. We will need the following four conditions on $g(X, \theta_x)$ and $g(Y, \theta_y)$.

Condition 1. $E\{g(X, \theta_{x_0})\} = 0$ and $E\{g(Y, \theta_{y_0})\} = 0$, and $\mathrm{var}\{g(X, \theta_{x_0})\} \in \mathbb{R}^{q\times q}$ and $\mathrm{var}\{g(Y, \theta_{y_0})\} \in \mathbb{R}^{q\times q}$ are both positive definite.

Condition 2. $\partial g(X, \theta_x)/\partial\theta_x$ and $\partial g^2(X, \theta_x)/\partial\theta_x\partial\theta_x^T$ are continuous in θ_x, and for θ_x in a neighbourhood of θ_{x_0} they are each bounded in norm by an integrable function of X.

Condition 3. $\partial g(Y, \theta_y)/\partial\theta_y$ and $\partial g^2(Y, \theta_y)/\partial\theta_y\partial\theta_y^T$ are continuous in θ_y, and for θ_y in a neighbourhood of θ_{y_0} they are each bounded in norm by an integrable function of Y.

Condition 4. $\limsup_{\|t\|\to\infty} |E[\exp\{it^T g(X, \theta_x)\}]| < 1$ and $E\|g(X, \theta_x)\|^{15} < +\infty$; $\limsup_{\|t\|\to\infty} |E[\exp\{it^T g(Y, \theta_y)\}]| < 1$ and $E\|g(Y, \theta_y)\|^{15} < +\infty$.

Denote by $\bar{p} = (p_1, ..., p_m)$ and $\bar{q} = (q_1, ..., q_n)$ two probability vectors satisfying $p_i \geq 0$, $q_j \geq 0$, $\sum_{i=1}^m p_i = 1$ and $\sum_{i=1}^n q_j = 1$. Let θ_y and θ_x be points in $\mathbb{R}^p$ and denote by $\theta_y(\bar{q})$ and $\theta_x(\bar{p})$ values that satisfy

$$\sum_{i=1}^{m} p_i g(X_i, \theta_x(\bar{p})) = 0, \qquad \sum_{j=1}^{n} p_j g(Y_j, \theta_y(\bar{q})) = 0.$$

Let $\pi = \theta_y - \theta_x \in \mathbb{R}^p$ and let $\pi(\bar{p}, \bar{q}) = \theta_y(\bar{q}) - \theta_x(\bar{p})$. Then, the two-sample OEL for a possible value of the difference π, $L(\pi)$, is defined as

$$L(\pi) = \sup_{(\bar{p},\bar{q}):\pi(\bar{p},\bar{q})=\pi} \left(\prod_{i=1}^{m} p_i \right) \left(\prod_{j=1}^{n} q_j \right), \tag{4.1}$$

which is the maximum of the product of the one-sample OEL for θ_y and the one-sample OEL for θ_x taken over all pairs (θ_x, θ_y) that satisfies $\pi = \theta_y - \theta_x$. The corresponding two-sample empirical log-likelihood ratio for π is thus

$$l(\pi) = -2 \sup_{(\bar{p},\bar{q}):\pi(\bar{p},\bar{q})=\pi} \left\{ \sum_{i=1}^{m} \log(mp_i) + \sum_{j=1}^{n} \log(nq_j) \right\}. \tag{4.2}$$

For convenience, we will also use OEL for the original empirical log-likelihood ratio. We will write "OEL $L(\pi)$" and "OEL $l(\pi)$" for the original empirical likelihood ratio (4.1) and log-likelihood ratio (4.2), respectively.

Define the domain of $L(\pi)$, D_n, as

$$D_n = \{\pi \in \mathbb{R}^p : \text{there exist } \theta_y(\bar{q}) \text{ and } \theta_x(\bar{p}) \text{ such that } \pi = \theta_y(\bar{q}) - \theta_x(\bar{p}).\},$$

and define the domain of $l(\pi)$, Π_n, as

$$\Pi_n = \{\pi : \pi \in D_n \text{ and } l(\pi) < +\infty\}.$$

Let $N = m + n$, $f_m = N/m$ and $f_n = N/n$. Without loss of generality, we assume that $m \geq n > q$. We also assume that $m/n = O(1)$ so that $O(n^{-1})$, $O(m^{-1})$ and $O(N^{-1})$, for example, are all interchangeable. By the method of Lagrangian multipliers, we have

$$l(\pi_0) = 2 \left[\sum_{j=1}^{n} \log\{1 + f_n(\lambda^*)^T g(Y_j, \theta_y^*)\} + \sum_{i=1}^{m} \log\{1 - f_m(\lambda^*)^T g(X_i, \theta_x^*)\} \right] \tag{4.3}$$

where $(\lambda^*, \theta_y^*, \theta_x^*)$ is the solution of the following non-linear system

$$\begin{aligned}
&\sum_{j=1}^{n} \frac{g(Y_j, \theta_y)}{1 + f_n \lambda^T g(Y_j, \theta_y)} = 0, \\
&\sum_{i=1}^{m} \frac{g(X_i, \theta_x)}{1 - f_m \lambda^T g(X_i, \theta_x)} = 0, \\
&\pi_0 = \theta_y - \theta_x.
\end{aligned} \tag{4.4}$$

Hence, we may write $l(\pi_0) = l(\lambda^*, \theta_y^*, \theta_x^*)$. The following theorem gives the asymptotic distribution of $l(\pi_0)$.

Theorem 4.1. *Under Conditions 1, 2, 3 and 4, the two-sample OEL $l(\pi_0)$ defined by (4.3) satisfies*

$$l(\pi_0) \xrightarrow{D} \chi_q^2 \quad \textit{as } n \to +\infty. \tag{4.5}$$

By Theorem 4.1, the $100(1-\alpha)\%$ two-sample OEL confidence region for π_0 is

$$\mathcal{C}_{1-\alpha} = \{\pi : \pi \in \mathbb{R}^p \text{ and } l(\pi) \le c_\alpha\} \tag{4.6}$$

where c_α is $(1-\alpha)$th quantile of the χ_q^2 distribution. The coverage error of $\mathcal{C}_{1-\alpha}$ is $O(n^{-1})$, that is

$$P(\pi_0 \in \mathcal{C}_{1-\alpha}) = P\{l(\pi_0) \le c_\alpha\} = 1 - \alpha + O(n^{-1}). \tag{4.7}$$

Theorem 4.1 is the standard first-order result for an OEL. The error rate of $O(n^{-1})$ in (4.7) follows from an argument in DiCiccio, Hall and Romano (1991) for that of the one-sample empirical likelihood. See also Hall and La Scala (1990).

DiCiccio, Hall and Romano (1991) and Chen and Cui (2007) showed that the one-sample OEL for estimating equations is Bartlett correctable; the Bartlett correction reduces the coverage error of the empirical likelihood confidence region to $O(n^{-2})$. Theorem 3.2 shows that the two-sample OEL for estimating equations (4.3) is also Bartlett correctable. The key result for proving Theorem 3.2 is Lemma 4.1 below. In order to present Lemma 4.1, we need to first introduce some new notations.

Denote by θ_y^k, θ_x^k and ζ^k approximations of θ_y^*, θ_x^* and λ^*, respectively. For brevity, the analytic expressions of θ_y^k, θ_x^k and ζ^k will be given later in the Appendix. For these three notations, we note that the k in say θ_y^k is a superscript (not to the power of k) which indicates the order of the approximation is $O(n^{-(k+1)/2})$, *i.e.*, $\theta_y^k = \theta_y + O(n^{-(k+1)/2})$. Let $V_1 = f_n \mathrm{var}\{g(Y, \theta_y)\}$, $V_2 = f_m \mathrm{var}\{g(X, \theta_x)\}$, $V = V_1 + V_2$ and $W = V_1 V^{-1} V_2$. Further, define

$$
\begin{aligned}
& z_{j0} = V^{-1/2} g(y_j, \theta_y^0), \quad z_{i0} = V^{-1/2} g(x_i, \theta_x^0) \quad z_{j1} = V^{-1/2} g(y_j, \theta_y^1), \qquad (4.8)\\
& z_{i1} = V^{-1/2} g(x_i, \theta_x^1), \quad s^{t_1 t_2 \ldots t_l} = f_n^{l-1} E(z_{j0}^{t_1} z_{j0}^{t_2} \ldots z_{j0}^{t_l}) + (-1)^l f_m^{l-1} E(z_{i0}^{t_1} z_{i0}^{t_2} \ldots z_{i0}^{t_l}),\\
& S^{t_1 t_2 \ldots t_l} = \frac{f_n^{l-1}}{n} \sum_{j=1}^{n} (z_{j0}^{t_1} z_{j0}^{t_2} \ldots z_{j0}^{t_l}) + \frac{(-1)^l f_m^{l-1}}{m} \sum_{i=1}^{m} (z_{i0}^{t_1} z_{i0}^{t_2} \ldots z_{i0}^{t_l}) - s^{t_1 t_2 \ldots t_l},\\
& S_1^{t_1 t_2 \ldots t_l} = \frac{f_n^{l-1}}{n} \sum_{j=1}^{n} (z_{j1}^{t_1} z_{j1}^{t_2} \ldots z_{j1}^{t_l}) + \frac{(-1)^l f_m^{l-1}}{m} \sum_{i=1}^{m} (z_{i1}^{t_1} z_{i1}^{t_2} \ldots z_{i1}^{t_l}) - s^{t_1 t_2 \ldots t_l},
\end{aligned}
$$

and

$$
\begin{aligned}
\Delta_1 &= S^\tau S^\tau - S^{\tau\upsilon} S^\tau S^\upsilon + \frac{2}{3} s^{\tau\alpha\beta} S^\tau S^\alpha S^\beta + S^{\tau\upsilon} S^{\upsilon\omega} S^\tau S^\omega + \frac{2}{3} S^{\tau\alpha\beta} S^\tau S^\alpha S^\beta,\\
&\quad -2 s^{\tau\upsilon\omega} S^{\tau\alpha} S^\upsilon S^\alpha S^\upsilon + s^{\tau\upsilon\omega} s^{\tau\alpha\beta} S^\upsilon S^\omega S^\alpha S^\beta - \frac{1}{2} s^{\tau\upsilon\omega\alpha} S^\tau S^\upsilon S^\omega S^\alpha,\\
\Delta_2 &= (S^{\tau\upsilon} - S_1^{\tau\upsilon}) S^\tau S^\upsilon + \frac{2}{3} (S_1^{\tau\alpha\beta} - S^{\tau\alpha\beta}) S^\tau S^\alpha S^\beta.
\end{aligned}
$$

where we have used the common summation convention that if an index appears more than once in an expression, summation over the index is understood.

Lemma 4.1. *With above notations and under condition 1, 2, 3, and 4, we have*

$$\frac{l(\pi_0)}{N} = \Delta_1 + \Delta_2 + O_p(n^{-5/2}). \tag{4.9}$$

To see the connection between expansion (4.9) and that of other high-order expansions of empirical log-likelihood ratios, we note that the Δ_1 term in (4.9) is similar to the expansion of the one-sample empirical log-likelihood ratio at the true parameter value given by DiCiccio, Hall and Romano (1991) and Chen and Cui (2007). In the present case, the expansion at the true difference π_0 depends on the true parameter values θ_{x0} and θ_{y0}, both of which need to be estimated. The use of the estimated values of these parameters resulted in the extra term Δ_2 in expansion (4.9). See also a similar Δ_2 term in the expansion of the two-sample empirical log-likelihood ratio for the mean in Liu, Zou and Zhang (2008).

We now use Lemma 4.1 to derive the two-sample Bartlett corrected empirical likelihood confidence region for the difference between two parameters defined by estimating equations. Let η be the Bartlett correction factor where

$$\eta = -\frac{1}{3d}s^{\tau\upsilon\omega}s^{\tau\alpha\beta} + \frac{1}{2d}s^{\tau\tau\alpha\alpha} + \frac{f_m f_n}{d}tr(V^{-1/2}WV^{-1/2}). \tag{4.10}$$

The derivation of (4.10) is similar to that for the Bartlett correction factor for the difference between two means in Liu, Zou and Zhang (2008), which involves taking expectations of Δ_1 and Δ_2 and omitting terms of order $O(n^{-1})$. With η, the two-sample Bartlett corrected empirical log-likelihood ratio (BEL) is given by

$$l_B(\pi) = l(\pi)(1 - \eta N^{-1}).$$

It follows that the two-sample BEL confidence region $\mathcal{C}'_{1-\alpha}$ for π_0 is

$$\mathcal{C}'_{1-\alpha} = \{\pi : \pi \in \mathbb{R}^p \text{ and } l_B(\pi) \le c\}. \tag{4.11}$$

Theorem 3.2 below shows the coverage error of $\mathcal{C}'_{1-\alpha}$ is $O(n^{-2})$.

Theorem 4.2. *Under Conditions 1, 2, 3 and 4, for any $c > 0$ the Bartlett corrected two-sample empirical likelihood confidence region satisfies*

$$P(\pi_0 \in \mathcal{C}'_{1-\alpha}) = P[l(\pi_0)\{1 - \eta N^{-1}\} \le c] = P(\chi^2_d \le c) + O(n^{-2}). \tag{4.12}$$

A stronger result due to DiCiccio, Hall and Romano (1991) is that

$$P[l(\pi_0)\{1 - \eta N^{-1} + O_p(n^{-3/2})\} \le c] = P(\chi^2_d \le c) + O(n^{-2}). \tag{4.13}$$

The Bartlett correction factor η in (4.11) and (4.12) depends on the moments of $g(X;\theta_{x0})$ and $g(Y;\theta_{y0})$ which are not available in empirical likelihood applications. Fortunately, by (4.13) we can use a $\sqrt{n}$-consistent estimator $\hat{\eta}$ in place of the η in (4.12) without affecting the $O(n^{-2})$ error term in (4.12). In real applications of the Bartlett correction, the $\sqrt{n}$-consistent estimator $\hat{\eta}$ is usually used instead of the exact η; see for example, Chen and Cui (2007) and Liu and Chen (2010). For two-sample BEL for the difference between two means, Liu, Zou and Zhang (2008) gave a moment estimator $\hat{\eta}$ for η. Liu and Yu (2010) reported that $\hat{\eta}$ tends to underestimate η and proposed a less biased estimator $\tilde{\eta}$ for η. This less biased $\tilde{\eta}$ is also applicable to our two-sample BEL for π_0, and we will use this $\tilde{\eta}$ for our simulation studies.

4.3 Two-sample extended empirical likelihood (EEL) for estimating equations

4.3.1 Composite similarity mapping

Like the one-sample OEL for estimating equations, the two-sample OEL $l(\pi)$ also suffers from the mismatch problem between its domain Π_n and the parameter space since the parameter space is $\mathbb{R}^p$ but $\Pi_n \subset \mathbb{R}^p$. The mismatch problem is a main contributor to the undercoverage problem of the OEL confidence regions. To solve this problem, we now expand Π_n to match the parameter space $\mathbb{R}^p$ through a composite similarity mapping (Tsao and Wu, 2013). This leads to an EEL defined on $\mathbb{R}^p$ and hence is free from the mismatch problem.

Denote by $\tilde{\theta}_x$ and $\tilde{\theta}_y$ the $\sqrt{n}$-consistent maximum empirical likelihood estimators (MELEs) for θ_{x0} and θ_{y0}, respectively. Then, it is not difficult to show that the MELE of π_0 is $\tilde{\pi} = \tilde{\theta}_y - \tilde{\theta}_x$ which is $\sqrt{n}$-consistent for π_0. We define the composite similarity mapping $h_N^C : \Pi_n \to \mathbb{R}^p$ as

$$h_N^C(\pi) = \tilde{\pi} + \gamma\{N, l(\pi)\}(\pi - \tilde{\pi}) \quad \text{for } \pi \in \Pi_n, \tag{4.14}$$

where function $\gamma\{N, l(\pi)\}$ is the expansion factor given by the following expression which depends continuously on π

$$\gamma\{N, l(\pi)\} = 1 + \frac{l(\pi)}{2N}. \tag{4.15}$$

To see how h_N^C maps Π_n onto $\mathbb{R}^p$, define the level-τ OEL contour as

$$c(\tau) = \{\pi : \pi \in \Pi_n \text{ and } l(\pi) = \tau\}, \tag{4.16}$$

where $\tau \geq \tilde{\tau} = l(\tilde{\pi}) \geq 0$. For the just-determined case, the one-sample OEL's satisfy $l(\tilde{\theta}_x) = 1$ and $l(\tilde{\theta}_y) = 1$. Thus, $L(\tilde{\pi}) = 1$ and $\tilde{\tau} = l(\tilde{\pi}) = 0$. The contours form a partition of the domain Π_n; that is, $c(\tau_1) \cap c(\tau_2) = \varnothing$ for any $\tau_1 \neq \tau_2$ and

$$\Pi_n = \bigcup_{\tau \in [\tilde{\tau}, +\infty)} c(\tau). \tag{4.17}$$

In addition to conditions 1 to 4 above, we now introduce a new condition.

Condition 5. Each contour $c(\tau)$ is the boundary of a connected region in $\mathbb{R}^p$, and the contours are nested in that if $\tau_1 < \tau_2$, then $c(\tau_1)$ is contained in the interior of the region defined by $c(\tau_2)$.

Under Condition 5 and in view of (4.17), the MELE $c(\tilde{\tau}) = \{\tilde{\pi}\}$ may be regarded as the centre of domain Π_n. It follows that the value of τ measures the outwardness of a $c(\tau)$ with respect to the centre; the larger the τ value, the more outward $c(\tau)$ is. The following theorem gives three key properties of h_N^C.

Theorem 4.3. *Under conditions 1, 2 and 3, mapping h_N^C defined by (4.14) and (4.15) satisfies (i) it has a unique fixed point at $\tilde{\pi}$, (ii) it is a similarity transformation for each individual contour $c(\tau)$ and (iii) it is a surjection from Π_n to $\mathbb{R}^p$.*

As a result of (ii), h_N^C may be viewed as a continuous sequence of similarity mappings from $\mathbb{R}^p$ to $\mathbb{R}^p$ that are indexed by $\tau \in [\tilde{\tau}, +\infty)$. The τ-th mapping has expansion factor $\gamma\{N, l(\pi)\} = \gamma(N, \tau)$ and it maps only points on the level-τ contour $c(\tau)$. Regardless of the amount expanded, $c(\tau)$ and its image are identical in shape. By (4.15), the expansion factor $\gamma(N, \tau)$ is an increasing function of τ which approaches infinity when τ does. Hence, contours farther away from the centre are expanded more and images of the contours fill up the entire $\mathbb{R}^p$.

If we are to add Condition 5 to Theorem 4.3, then (iii) can be strengthened to (iii') h_N^C is a bijection from Π_n to $\mathbb{R}^p$. See, for example, the proof of Theorem 1 in Tsao and Wu (2013). It is not clear how we may verify condition 5 through $g(X, \theta_x)$ and $g(Y, \theta_y)$. This is why we have not added it to Theorem 4.3. Nevertheless, we have not encountered any example where Condition 5 is violated.

4.3.2 Extended empirical likelihood on the full parameter space

By Theorem 4.3, $h_N^C : \Pi_n \to \mathbb{R}^p$ is surjective. Thus, for any $\pi \in \mathbb{R}^p$, $s(\pi) = \{\pi' : \pi' \in \Pi_n \text{ and } h_N^C(\pi') = \pi\}$ is non-empty. When h_N^C is not injective, $s(\pi)$ may contain multiple points and h_N^C does not have an inverse. Hence, we define a generalized inverse $h_N^{-C} : \mathbb{R}^p \to \Pi_n$ as follows

$$h_N^{-C}(\pi) = \text{argmin}_{\pi' \in s(\pi)}\{\|\pi' - \pi\|\} \quad \text{for } \pi \in \mathbb{R}^p. \tag{4.18}$$

If $s(\pi)$ contains exactly one point π', then $h_N^{-C}(\pi) = \pi'$. If $s(\pi)$ has multiple points, then $h_N^{-C}(\pi)$ equals the point $\pi' \in s(\pi)$ that is the closest to π.

We now define the EEL $l^*(\pi)$ under h_N^{-C} as follows

$$l^*(\pi) = l\{h_N^{-C}(\pi)\} \quad \text{for } \pi \in \mathbb{R}^p. \tag{4.19}$$

It is clear that $l^*(\pi)$ is well-defined throughout $\mathbb{R}^p$ since $h_N^{-C}(\pi) \in \Pi_n$ for any $\pi \in \mathbb{R}^p$ and thus the right-hand side of (4.19) is always well-defined. Let π_0' be the image of π_0 under the inverse mapping $h_N^{-C}(\pi)$, that is,

$$h_N^{-C}(\pi_0) = \pi_0'. \tag{4.20}$$

Then, $l^*(\pi_0) = l\{h_N^{-C}(\pi_0)\} = l(\pi_0')$. Denote by $[\tilde{\pi}, \pi_0]$ the line segment in $\mathbb{R}^p$ that connects the two points $\tilde{\pi}$ and π_0. Lemma 4.2 below shows that π_0' is on $[\tilde{\pi}, \pi_0]$ and that it is asymptotically very close to π_0.

Lemma 4.2. *Under conditions 1, 2 and 3, the point π_0' defined by equation (4.20) satisfies*

$$\text{(i) } \pi_0' \in [\tilde{\pi}, \pi_0] \quad \text{and} \quad \text{(ii) } \pi_0' - \pi_0 = O_p(n^{-3/2}).$$

Theorem 4.4 below gives the asymptotic distribution of $l^*(\pi_0)$.

Theorem 4.4. *Under conditions 1, 2, 3 and 4, the two-sample EEL $l^*(\pi)$ defined by (4.19) satisfies*

$$l^*(\pi_0) \longrightarrow \chi^2_q$$

in distribution as $n \to +\infty$.

The proof of Theorem 4.4 makes use of the observation that

$$l^*(\pi_0) = l\{h_N^{-C}(\pi_0)\} = l(\pi_0') = l\{\pi_0 + (\pi_0' - \pi_0)\}. \tag{4.21}$$

Since by Lemma 4.2 $\|\pi_0' - \pi_0\|$ is asymptotically very small, (4.21) implies that $l^*(\pi_0) = l(\pi_0) + o_p(1)$. This and the fact that $l(\pi_0)$ has an asymptotic χ^2_q distribution lead to Theorem 4.4. The relationship in (4.21) is also the key in the derivation of a second-order two-sample EEL in the next section.

4.3.3 Second-order extended empirical likelihood

We have seen in Theorem 4.2 that the two-sample OEL admits a Bartlett correction which reduces the coverage error of the empirical likelihood confidence region to $O(n^{-2})$. The following theorem shows that for the just-determined case, the two-sample EEL can also attain the second-order accuracy.

Theorem 4.5. *Assume conditions 1, 2, 3 and 4 hold. For the just-determined case where $p = q$, let $l_2^*(\pi)$ be the EEL defined by the composite similarity mapping (4.14) with expansion factor $\gamma\{N, l(\pi)\} = \gamma_2\{N, l(\pi)\}$ given by*

$$\gamma_2\{N, l(\pi)\} = 1 + \frac{\eta}{2N}\{l(\pi)\}^{\delta(n)}, \tag{4.22}$$

where $\delta(n) = O(n^{-1/2})$ and η is the Bartlett correction constant in (4.10). Then,

$$l_2^*(\pi_0) = l(\pi_0)\{1 - \eta N^{-1} + O_p(n^{-3/2})\}, \tag{4.23}$$

and for any fixed $c > 0$,

$$P(l_2^*(\pi_0) \leq c) = P(\chi_d^2 \leq c) + O(n^{-2}). \tag{4.24}$$

Equation (4.24) follows from (4.23) and (4.13). It shows that confidence regions based on $l_2^*(\pi)$ have a coverage error of $O(n^{-2})$. Hence, we call $l_2^*(\pi)$ the second-order EEL or EEL$_2$. Correspondingly, we call $l^*(\pi)$ in (4.19), which is defined with the expansion factor $\gamma\{N, l(\pi)\}$ in (4.15), the first-order EEL or EEL$_1$. The $\delta(n)$ function in $\gamma_2\{N, l(\pi)\}$ is used to control the speed of domain expansion to achieve the second-order accuracy. For convenience, we will use $\delta(n) = n^{-1/2}$ when we compute EEL$_2$ in our numerical examples.

4.4 Applications and numerical comparison

The need for comparing two populations/models in terms of some numerical aspect of interest arises frequently in applied research. Whenever the numerical aspect of interest can be represented by a parameter defined by estimating equations, the two-sample OEL, BEL and EEL discussed here may be applied to make the comparison. In this section, we consider two such applications. The first is concerned with comparing two populations in terms of the inequality of income distribution. The second is concerned with comparing two linear regression models. Through these two examples, we also compare the numerical accuracy of the three two-sample empirical likelihood methods.

4.4.1 Application 1: Comparing two Gini indices

The Gini index was introduced by Corrado Gini, an Italian statistician of the early 20th century, as a measure of inequality of income or wealth distribution in a country. The value of the Gini index is bounded between 0 and 1, with 0 representing complete equality where all individuals have equal income and 1 representing complete inequality where one individual has all the income and others have none. Gini index has been widely used in social and economic studies of income distributions [e.g., Gini (1936), Chen (2009), Domeij Domeij and Flodén (2010), and Bee (2012)]. There are also a lot of work on the estimation and inference of the Gini index in both the statistical and econometric literature.

Qin, Rao and Wu (2010) and Peng (2011) applied the method of empirical likelihood to make inference about the Gini index. In particular, Peng (2011) derived an interesting estimating equation for the Gini index with which the existing theory of Qin and Lawless (1994) was readily applied to make empirical likelihood inference for the index. Peng (2011) also derived empirical likelihoods for the difference between two Gini indexs with paired data and two independent samples. We now apply the two-sample methods to make inference about the difference between two Gini indices using the estimating equation of Peng (2011). For this application, the two-sample OEL coincides with that given by Peng (2011).

Let $X_1, ..., X_n$ be i.i.d. observations from an income distribution $F(x)$ supported on $[0, +\infty)$. Define $T_i = \{X_i + X_{[n/2]+i}\}/2$ and $Z_i = \min\{X_i, X_{[n/2]+i}\}$ for $i = 1, ..., [n/2]$ where $[n/2]$ is the integer part of $n/2$. Then, Peng (2011) showed that the Gini index, θ_0, of distribution $F(x)$ satisfies,

$$E(T_i - Z_i - T_i\theta_0) = 0. \tag{4.25}$$

Let $F_A(x)$ be the income distribution of Country A with Gini index θ_{x0} and $F_B(y)$ be that of Country B with Gini index θ_{y0}. Suppose we have two random samples of sizes m and n, respectively, from $F_A(x)$ and $F_B(y)$. Then, we can compute confidence intervals for the difference $\pi_0 = \theta_{y0} - \theta_{x0}$ by using the two-sample OEL, BEL, EEL_1 and EEL_2. To illustrate their use and to compare the coverage accuracy of confidence intervals based these methods, we consider the following two examples:

Example 1: F_A is log-normal with $\log(X) \sim N(0,1)$ and F_B is χ^2_1.

Example 2: F_A is Pareto(5) and F_B is Exp(1).

The true value of the Gini index for Exp(1) is 0.5 and the true values of Gini index for $\log(X) \sim N(0,1)$, χ^2_1 and Pareto(5) are approximately 0.5205, 0.6366 and 0.1111, respectively, obtained through Monte Carlo simulations. Before presenting numerical results, note that the two-sample OEL confidence interval for π_0 is $\mathcal{C}_{1-\alpha}$ given by (4.6). The BEL confidence interval is $\mathcal{C}'_{1-\alpha}$ given in (4.11). The EEL_1 $l^*(\pi)$ and EEL_2 $l^*_2(\pi)$ are both defined through the OEL $l(\pi)$ and the inverse of the composite similarity mapping $h_N^{-C}(\pi)$ in (4.19); the expansion factor in $h_N^{-C}(\pi)$ corresponding to $l^*(\pi)$ is given by (4.15) and that corresponding to $l^*_2(\pi)$ is given by (4.22). The EEL_1 confidence interval is $\mathcal{C}^*_{1-\alpha} = \{\pi : \pi \in \mathbb{R}^p \text{ and } l^*(\pi) \leq c\}$ and the EEL_2 confidence interval is $\mathcal{C}'^*_{1-\alpha} = \{\pi : \pi \in \mathbb{R}^p \text{ and } l^*_2(\pi) \leq c\}$. The Bartlett correction factor η needs to be estimated when computing the BEL and EEL_2 confidence intervals, and in both cases we have used the less biased estimator $\tilde{\eta}$ given by Liu and Yu (2010).

Table 4.1 contains simulated coverage probabilities of the four confidence intervals for the difference between the two Gini indexes of F_A and F_B in Example 1. Table 4.2 contains that for Example 2. Each entry in these tables is based on 10,000 pairs of random samples whose sizes are given in the first two columns; it is the proportion of confidence intervals containing the true difference among the 10,000 confidence intervals computed using the 10,000 pairs of samples. We make the following observations based on Tables 4.1 and 4.2.

1. All four confidence intervals give coverage probabilities lower than the nominal level. The OEL interval, in particular, gives the lowest coverage probabilities that may be as much as 10% lower than the nominal level.

2. The BEL, EEL_1 and EEL_2 intervals are consistently more accurate than the OEL intervals. The two EEL intervals are more accurate than the BEL interval for all combinations of sample sizes and confidence level. Surprisingly, the first-order EEL_1 is overall the best, more accurate than the second-order BEL and EEL_2 intervals. Hence, we recommend EEL_1 for this application.

3. In column 1 of Table 4.2, we see that the OEL coverage probability for $(m, n) = (20, 40)$ is lower than that for $(20, 30)$; in this case the larger sample sizes did not give higher coverage probability. This surprising phenomenon occurs sometimes for other other two-sample methods as well. See also Table 2 in Liu and Yu (2010) for similar results. Noting that $m - n$ is smaller in $(20, 30)$, it seems that for two-sample inference a large difference in sample sizes can negatively affect the accuracy of the EL based confidence intervals.

4.4.2 Application 2: Comparing two linear regression models

Consider two simple linear regression models having the same predictor variable but possibly different slopes and intercepts. To compare the parameters of the two models with two independent random samples (one from each model), a commonly used method is to introduce a dummy/indicator variable and the comparison is then done through a multiple linear regression model with two covariates; the predictor variable and the dummy variable, and an interaction may also be included. This method, however, requires the assumption that error distributions of the two models are the same. Without making this assumption, we now use two-sample empirical likelihood methods to compare the model parameters. Specifically, we compare models

$$(a)\;\; y = x^T\beta_a + \varepsilon_a \quad \text{and} \quad (b)\;\; y = x^T\beta_b + \varepsilon_b,$$

where $\beta_a = (\beta_{a0}, \beta_{a1})^T$, $\beta_b = (\beta_{b0}, \beta_{b1})^T$, ε_a and ε_b are random errors with possibly different distributions, but $x = (1, x_1)^T$ is the same in both models. The parameter vector of interest is the difference $\pi = \beta_a - \beta_b$.

For our simulation study, x_1 is assumed to be a uniform random variable on $[0, 30]$. We consider the following two examples:

Example 3: Model (a) with $\varepsilon_a \sim N(0,1)$ and $\beta_a = (2,1)^T$ and Model (b) with $\varepsilon_b \sim N(0,1)$ and $\beta_b = (2,2)^T$.

Example 4: Model (a) with $\varepsilon_a \sim Exp(1) - 1$ and $\beta_a = (2,1)^T$ and Model (b) with $\varepsilon_b \sim N(0,1)$ and $\beta_b = (2,2)^T$.

The simulated coverage probabilities for π given by the four empirical likelihood methods are shown in Table 4.3. Although Examples 3 and 4 are multi-dimensional examples ($p = 2$), the three observations made above also apply to Table 4.3. In particular, overall EEL_1 has better accuracy than OEL, BEL, and EEL_2. We recommend

Table 4.1: Coverage probabilities (%) of confidence regions based on OEL, EEL_1, BEL and EEL_2 for Example 1

		90% level				95% level				99% level			
m	n	OEL	EEL_1	BEL	EEL_2	OEL	EEL_1	BEL	EEL_2	OEL	EEL_1	BEL	EEL_2
20	20	80.0	81.9	81.4	82.4	86.5	88.7	87.7	88.4	94.0	95.8	94.5	95.2
	30	81.1	83.3	82.4	83.9	87.6	89.8	88.8	89.8	95.2	96.8	95.7	96.3
	40	82.0	84.1	83.2	84.9	88.5	90.9	89.6	91.0	95.3	96.9	95.7	96.3
	60	82.1	84.2	83.1	85.7	88.1	90.7	89.2	91.3	95.9	97.2	96.4	97.0
30	20	79.7	81.7	80.9	82.2	86.6	88.4	87.5	88.5	94.0	95.8	94.6	95.2
	30	82.6	84.0	83.7	84.5	89.1	90.3	89.9	90.3	95.7	96.6	96.0	96.3
	40	83.2	84.5	84.2	85.1	89.5	90.9	90.3	91.0	96.0	97.1	96.4	96.9
	60	84.2	85.5	84.9	86.3	90.3	91.7	91.1	92.0	97.0	97.8	97.3	97.7
40	20	80.4	82.2	81.5	82.9	87.0	88.5	87.8	88.6	94.2	95.7	94.7	95.2
	30	82.9	84.2	83.8	84.8	89.6	90.9	90.4	90.9	96.2	97.2	96.6	96.9
	40	84.4	85.5	85.2	86.0	90.6	91.6	91.3	91.7	96.8	97.6	97.1	97.4
	60	85.6	86.6	86.5	87.2	91.4	92.5	92.1	92.8	97.2	97.8	97.5	97.7
60	20	79.7	81.5	80.8	82.7	86.7	88.6	87.5	88.8	94.6	95.8	95.0	95.6
	30	83.5	84.7	84.2	85.3	89.6	90.6	90.2	91.0	96.0	97.0	96.3	96.8
	40	85.1	86.0	85.9	86.7	91.3	92.1	91.9	92.3	97.2	97.8	97.4	97.6
	60	85.8	86.4	86.5	86.9	91.6	92.4	92.2	92.5	97.4	97.8	97.5	97.7

Each entry in the table is a simulated coverage probability for π based on 10,000 random samples of size m and n indicated in column 1 and 2 from the distribution log-normal (i.e. log $N(0,1)$) and χ_1^2, respectively.

EEL_1 due to its simplicity and accuracy.

4.5 Appendix

We now present proofs of theorems and lemmas in the order as they appeared in the paper. For brevity, for results that are minor variations of existing results in the literature, we give only references to the existing results instead of detailed proofs which may be found in the references.

Theorem 4.1 is the standard first-order result for an OEL. It is implied by Theorem 4.2 which gives the second-order result. Hence, its proof is omitted. We now prove Lemma 4.1 by following that for equation (3) in Liu, Zou and Zhang (2008).

Proof of Lemma 4.1

First note that, under conditions of Lemma 4.1, $\lambda^* = O_p(n^{-1/2})$; see the proof of Theorem 1 in Owen (1990). For clarity of presentation, we break the proof of Lemma 4.1 into the following three steps.

Table 4.2: Coverage probabilities (%) of confidence regions based on OEL, EEL_1, BEL and EEL_2 for Example 2

		90% level				95% level				99% level			
m	n	OEL	EEL_1	BEL	EEL_2	OEL	EEL_1	BEL	EEL_2	OEL	EEL_1	BEL	EEL_2
20	20	80.8	83.4	81.8	83.4	86.8	89.7	87.6	89.3	93.6	96.5	94.1	95.8
	30	82.0	84.5	83.1	84.8	88.1	90.1	88.7	89.9	94.3	96.2	94.7	95.7
	40	80.1	84.4	81.0	84.7	85.6	90.4	86.4	90.2	91.7	96.2	92.2	95.9
	60	82.1	83.8	83.0	84.7	88.0	89.8	88.8	90.1	95.0	96.4	95.3	96.1
30	20	84.1	86.0	85.1	86.2	90.1	92.0	90.7	91.7	95.8	97.7	96.2	97.2
	30	84.2	86.1	85.0	86.3	89.9	91.9	90.4	91.7	95.8	97.5	96.1	97.2
	40	84.2	86.2	85.0	86.4	89.9	92.2	90.6	92.1	95.2	97.5	95.5	97.1
	60	84.1	85.4	84.9	86.0	90.2	91.6	90.8	91.7	96.2	97.2	96.5	97.0
40	20	85.3	87.3	86.2	87.5	91.1	93.0	91.6	92.9	96.6	98.3	96.8	97.9
	30	85.1	87.2	85.8	87.5	90.8	92.9	91.2	92.8	96.4	98.3	96.7	98.0
	40	86.0	87.1	86.7	87.3	91.3	92.5	91.8	92.5	97.0	98.0	97.1	97.7
	60	85.4	86.4	86.1	86.7	91.2	92.1	91.7	92.2	96.8	97.6	97.1	97.4
60	20	86.2	88.0	86.8	88.7	91.8	93.9	92.3	93.8	97.6	99.0	97.7	98.7
	30	86.4	88.6	87.0	89.0	91.8	94.1	92.2	94.1	96.9	98.6	97.1	98.4
	40	87.2	88.4	87.8	88.7	92.8	93.9	93.3	93.9	98.0	98.6	98.1	98.4
	60	86.9	88.0	87.5	88.2	92.5	93.3	92.7	93.3	97.4	98.4	97.6	98.2

Each entry in the table is a simulated coverage probability for π based on 10,000 random samples of size m and n indicated in column 1 and 2 from the distribution $Pareto(5)$ and $Exp(1)$, respectively.

Table 4.3: Coverage probabilities (%) of confidence regions based on OEL, EEL_1, BEL and EEL_2 for Example 3 (Ex-3) and Example 4 (Ex-4)

		90% level				95% level				99% level			
	(m,n)	OEL	EEL_1	BEL	EEL_2	OEL	EEL_1	BEL	EEL_2	OEL	EEL_1	BEL	EEL_2
Ex-3	(20, 20)	80.9	85.8	83.0	85.2	86.9	91.6	88.6	90.5	93.2	96.9	94.1	95.8
	(20, 40)	81.2	85.2	83.0	84.7	86.8	90.7	88.1	89.8	93.2	96.6	93.8	95.5
	(30, 30)	84.2	87.5	85.8	87.3	90.2	93.2	91.5	92.9	95.4	97.7	95.7	97.1
	(40, 30)	83.1	87.0	84.2	86.5	88.4	92.6	89.5	91.7	93.9	97.2	94.3	96.4
	(40, 40)	87.2	88.6	88.5	88.6	92.9	94.2	93.8	94.0	98.1	98.9	98.6	98.7
	(50, 30)	80.7	86.3	81.9	85.5	86.1	91.5	87.0	90.7	91.3	96.0	91.6	95.1
	(50, 50)	87.7	88.6	88.5	88.7	93.6	94.5	94.2	94.3	98.5	99.0	98.8	98.8
Ex-4	(20, 20)	78.0	83.3	80.6	83.0	84.4	89.5	86.4	88.7	91.5	95.6	92.6	94.6
	(20, 40)	79.4	84.0	81.4	83.5	85.3	89.4	86.5	88.6	91.6	95.2	92.2	94.3
	(30, 30)	82.7	86.4	84.6	86.3	88.7	92.1	90.0	91.7	94.4	97.2	95.0	96.5
	(40, 30)	80.3	85.1	81.8	84.8	86.4	91.2	87.6	90.5	91.8	95.9	92.3	95.1
	(40, 40)	86.1	87.7	87.6	88.0	92.4	93.6	93.4	93.6	97/7	98.5	98.2	98.3
	(50, 30)	78.7	85.0	80.0	84.1	84.2	90.3	85.3	89.3	89.7	95.0	90.1	94.1
	(50, 50)	86.7	87.8	87.8	88.2	92.2	93.4	93.2	93.4	97.7	98.5	98.2	98.3

Each entry in the table is a simulated coverage probability for π based on 10,000 pairs of random samples with sizes (m, n) indicated in column 2 from the linear models indicated in column 1.

Step 1: Let $C_{11} = \frac{1}{n}\sum g(y_j, \theta_y^0)$ and $C_{12} = \frac{1}{m}\sum g(x_i, \theta_x^0)$. Taylor expansion of the first equation of (4.4) gives

$$\frac{1}{n}\sum g(y_j, \theta_y^*) - \left\{\frac{f_n}{n}\sum g(y_j, \theta_y^*)g^T(y_j, \theta_y^*) - V_1\right\}\lambda^* - V_1\lambda^* + O_p(n^{-1}) = 0. \quad (4.26)$$

It follows that

$$\lambda^* = V_1^{-1}\left\{\frac{1}{n}\sum g(y_j, \theta_y^0)\right\} + O_p(n^{-1}) = V_1^{-1}C_{11} + O_p(n^{-1}). \quad (4.27)$$

Similarly, expansion of the second equation of (4.4) gives,

$$\lambda^* = -V_2^{-1}\left\{\frac{1}{m}\sum g(x_i, \theta_x^0)\right\} + O_p(n^{-1}) = -V_2^{-1}C_{12} + O_p(n^{-1}). \quad (4.28)$$

Based on (4.27) and (4.28),

$$V_1\lambda^* = C_{11} + O_p(n^{-1}),$$

$$-V_2^{-1}\lambda^* = C_{12} + O_p(n^{-1}),$$

thus, we have

$$\zeta^0 = V^{-1}(C_{11} - C_{12}) = V^{-1}D_1.$$

Step 2. Further expanding the left-hand side of (4.26), we have

$$\begin{aligned}&\frac{1}{n}\sum g(y_j, \theta_y^*) - \left\{\frac{f_n}{n}\sum g(y_j, \theta_y^0)g^T(y_j, \theta_y^0) - V_1\right\}\zeta^0 \\ &-V_1\lambda^* + \frac{f_n^2}{n}\sum\left\{(\zeta^0)^T g(y_j, \theta_y^0)\right\}^2 g(y_j, \theta_y^0) + O_p(n^{-3/2}) = 0.\end{aligned} \quad (4.29)$$

Let

$$C_{21} = -\left\{\frac{f_n}{n}\sum g(y_j,\theta_y^0)g^T(y_j,\theta_y^0) - V_1\right\}\zeta^0, \tag{4.30}$$
$$C_{31} = \frac{f_n^2}{n}\sum\left\{(\zeta^0)^T g(y_j,\theta_y^0)\right\}^2 g(y_j,\theta_y^0),$$
$$C_{22} = \left\{\frac{f_m}{m}\sum g(x_i,\theta_x^0)g^T(x_i,\theta_x^0) - V_2\right\}\zeta^0,$$
$$C_{32} = -\frac{f_m^2}{m}\sum\left\{(\zeta^0)^T g(x_i,\theta_x^0)\right\}^2 g(x_i,\theta_x^0).$$

It follows that

$$\lambda^* = V_1^{-1}(C_{11}+C_{21}+C_{31}) + O_p(n^{-3/2}),$$
$$\lambda^* = -V_2^{-1}(C_{12}+C_{22}+C_{32}) + O_p(n^{-3/2}),$$

and we obtain

$$\zeta^1 = V^{-1}\left\{(C_{11}-C_{12}) + (C_{21}-C_{22}) + (C_{31}-C_{32})\right\} = V^{-1}(D_1+D_2+D_3).$$

Step 3. Further expanding (4.29) gives

$$\frac{1}{n}\sum g(y_j,\theta_y^*) - \left\{\frac{f_n}{n}\sum g(y_j,\theta_y^1)g^T(y_j,\theta_y^1) - V_1\right\}\zeta^1 - V_1\lambda^* +$$
$$\frac{f_n^2}{n}\sum\left\{(\zeta^1)^T g(y_j,\theta_y^1)\right\}^2 g(y_j,\theta_y^1) + \frac{f_n^3}{n}\sum\left\{(\zeta^0)^T g(y_j,\theta_y^0)\right\}^3 g(y_j,\theta_y^0) + O_p(n^{-2}) = 0.$$

Let

$$
\begin{aligned}
C_{21}^{*} &= -\left\{\frac{f_n}{n}\sum g(y_j,\theta_y^1)g^T(y_j,\theta_y^1) - V_1\right\}\zeta^1, \\
C_{31}^{*} &= \frac{f_n^2}{n}\sum\left\{(\zeta^1)^T g(y_j,\theta_y^1)\right\}^2 g(y_j,\theta_y^1), \\
C_{41} &= -\frac{f_n^3}{n}\sum\left\{(\zeta^0)^T g(y_j,\theta_y^0)\right\}^3 g(y_j,\theta_y^0), \\
C_{22}^{*} &= \left\{\frac{f_m}{m}\sum g(x_i,\theta_x^1)g^T(x_i,\theta_x^1) - V_2\right\}\zeta^1, \\
C_{32}^{*} &= -\frac{f_m^2}{m}\sum\left\{(\zeta^1)^T g(x_i,\theta_x^1)\right\}^2 g(x_i,\theta_x^1), \\
C_{42} &= \frac{f_m^3}{m}\sum\left\{(\zeta^0)^T g(x_i,\theta_x^0)\right\}^3 g(x_i,\theta_x^0).
\end{aligned}
$$

We have the following higher order expansions of λ^*,

$$
\begin{aligned}
\lambda^* &= V_1^{-1}(C_{11} + C_{21}^* + C_{31}^* + C_{41}) + O_p(n^{-2}), \qquad (4.31) \\
\lambda^* &= -V_2^{-1}(C_{12} + C_{22}^* + C_{32}^* + C_{42}) + O_p(n^{-2}).
\end{aligned}
$$

Thus

$$
\begin{aligned}
\zeta^2 &= V^{-1}\left\{(C_{11} - C_{12}) + (C_{21}^* - C_{22}^*) + (C_{31}^* - C_{32}^*) + (C_{41} - C_{42})\right\} \qquad (4.32) \\
&= V^{-1}(D_1 + D_2^* + D_3^* + D_4).
\end{aligned}
$$

Then, the Taylor expansion for $l(\pi)/N$ can be expressed as

$$\begin{aligned}
\frac{l(\pi)}{N} = \quad & \frac{2}{n}\sum(\zeta^2)^T g(y_j,\theta_y^0) - \frac{2}{m}\sum(\zeta^2)^T g(x_i,\theta_x^0) \\
& -\frac{f_n}{n}\sum(\zeta^2)^T g(y_j,\theta_y^1)g(y_j,\theta_y^1)^T\zeta_2 - \frac{f_m}{m}\sum(\zeta^2)^T g(x_i,\theta_x^1)g(x_i,\theta_x^1)^T\zeta_2 \\
& +\frac{2f_n^2}{3n}\sum\left\{(\zeta^1)^T g(y_j,\theta_y^1)\right\}^3 - \frac{2f_m^2}{3m}\sum\left\{(\zeta^1)^T g(x_i,\theta_x^1)\right\}^3 \\
& +\frac{f_n^3}{2n}\sum\left\{(\zeta^0)^T g(y_j,\theta_y^0)\right\}^4 - \frac{f_m^3}{2m}\sum\left\{(\zeta^0)^T g(x_i,\theta_x^0)\right\}^4 + O_p(n^{-5/2}) \\
= \quad & 2I_1 - I_2 + \frac{2}{3}I_3 - \frac{1}{2}I_4 + O_p(n^{-5/2}),
\end{aligned} \tag{4.33}$$

where

$$\begin{aligned}
I_1 &= \zeta^2(C_{11}-C_{12}) = D_1^T V^{-1}(D_1+D_2^*+D_3^*+D_4), \qquad (4.34)\\
I_2 &= \zeta^2(C_{11}-C_{12}+C_{31}^*-C_{32}^*+C_{41}-C_{42}) = (D_1+D_3^*+D_4)V^{-1}(D_1+D_2^*+D_3^*+D_4),\\
I_3 &= \zeta^1(C_{31}^*-C_{32}^*) = (D_3^*)V^{-1}(D_1+D_2+D_3),\\
I_4 &= -\zeta^0(C_{41}-C_{42}) = -D_1V^{-1}(D_4).
\end{aligned}$$

Noting that

$$D_1 = O_p(n^{-1/2}),\quad D_2 = O_p(n^{-1}),\quad D_2^* = O_p(n^{-1}),\quad D_2 - D_2^* = O_p(n^{-3/2}),$$

$$D_3 = O_p(n^{-1}),\quad D_3^* = O_p(n^{-1}),\quad D_3 - D_3^* = O_p(n^{-3/2}),\quad D_4 = O_p(n^{-3/2}),$$

thus, we have

$$
\begin{aligned}
\frac{l(\pi)}{N} = \quad & 2D_1^T V^{-1}(D_1 + D_2^* + D_3^* + D_4) - (D_1 + D_3^* + D_4)V^{-1}(D_1 + D_2^* + D_3^* + D_4) \\
& + \frac{2}{3}(D_3^*)V^{-1}(D_1 + D_2 + D_3) + \frac{1}{2}D_1 V^{-1} D_4 + O_p(n^{-5/2}) \\
= \quad & D_1^T V^{-1}(D_1 + D_2^* + D_3^* + D_4) - (D_3^* + D_4)V^{-1}(D_1 + D_2^* + D_3^* + D_4) \\
& + \frac{2}{3}(D_3^*)V^{-1}(D_1 + D_2 + D_3) + \frac{1}{2}D_1 V^{-1} D_4 + O_p(n^{-5/2}) \\
= \quad & \left\{ D_1^T V^{-1} D_1 + D_1^T V^{-1} D_2^* + D_1^T V^{-1} D_3^* + D_1^T V^{-1} D_4 \right\} \\
& - \left\{ D_3^{*T} V^{-1} D_1 + D_3^{*T} V^{-1} D_2^* + D_3^{*T} V^{-1} D_3^* + D_4^T V^{-1} D_1 + O_p(n^{-5/2}) \right\} \\
& + \frac{2}{3}\left\{ D_3^{*T} V^{-1} D_1 + D_3^{*T} V^{-1} D_2 + D_3^{*T} V^{-1} D_3 \right\} + \frac{1}{2} D_1^T V^{-1} D_4 + O_p(n^{-5/2}) \\
= \quad & D_1^T V^{-1} D_1 + D_1^T V^{-1} D_2^* + \frac{2}{3} D_1^T V^{-1} D_3^* - \frac{1}{3} D_2^T V^{-1} D_3 \\
& - \frac{1}{3} D_3^T V^{-1} D_3 + \frac{1}{2} D_1^T V^{-1} D_4 + O_p(n^{-5/2}).
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\frac{l(\pi)}{N} \quad = \quad & S^\tau S^\tau - S^{\tau\upsilon} S^\tau S^\upsilon + \frac{2}{3} s^{\tau\alpha\beta} S^\tau S^\alpha S^\beta + S^{\tau\upsilon} S^{\upsilon\omega} S^\tau S^\omega + \frac{2}{3} S^{\tau\alpha\beta} S^\tau S^\alpha S^\beta, \\
& -2 s^{\tau\upsilon\omega} S^{\tau\alpha} S^\upsilon S^\alpha S^\upsilon + s^{\tau\upsilon\omega} s^{\tau\alpha\beta} S^\upsilon S^\omega S^\alpha S^\beta - \frac{1}{2} s^{\tau\upsilon\omega\alpha} S^\tau S^\upsilon S^\omega S^\alpha \\
& + (S^{\tau\upsilon} - S_1^{\tau\upsilon}) S^\tau S^\upsilon + \frac{2}{3}(S_1^{\tau\alpha\beta} - S^{\tau\alpha\beta}) S^\tau S^\alpha S^\beta + O_p(n^{-5/2}),
\end{aligned}
$$

where

$$D_1^T V^{-1} D_1 = S^\tau S^\tau,$$

$$D_1^T V^{-1} D_2^* = -S_1^{\tau\upsilon} S^\tau S^\upsilon + S^{\tau\upsilon} S^{\upsilon\omega} S^\tau S^\omega - s^{\omega\alpha\beta} S^{\tau\upsilon} S^{\upsilon\omega} S^\alpha S^\beta + O_p(n^{-5/2}),$$

$$D_1^T V^{-1} D_3^* = (S_1^{\tau\alpha\beta} - S^{\tau\alpha\beta}) S^\tau S^\alpha S^\beta, \quad D_2^T V^{-1} D_3 = -s^{\tau\upsilon\omega} S^{\tau\alpha} S^\upsilon S^\alpha S^\upsilon,$$

$$D_3^T V^{-1} D_3 = s^{\tau\upsilon\omega} s^{\tau\alpha\beta} S^\upsilon S^\omega S^\alpha S^\beta, \quad D_1^T V^{-1} D_4 = s^{\tau\upsilon\omega\alpha} S^\tau S^\upsilon S^\omega S^\alpha,$$

which proves Lemma 4.1. □

With Lemma 4.1, the proof of Theorem 4.2 follows from that for the second order result in DiCiccio, Hall and Romano (1991). See DiCiccio, Hall and Romano (1988) for details.

Proof of Lemma 4.2

Differentiating $l(\pi)$ in (4.3) and evaluating the derivative at π_0, we find $J(\pi_0) = \frac{\partial l(\pi)}{\partial \pi}|_{\pi=\pi_0}$ as follows

$$J(\pi_0) = \lambda^T(\pi_0)\left\{\sum \frac{f_n g'(y_j,\theta_{y0})\frac{\partial\theta_y}{\partial\pi}|_{\pi=\pi_0}}{1+f_n\lambda^T(\pi_0)g(y_j,\theta_{y0})} - \sum \frac{f_m g'(x_i,\theta_{x0})\frac{\partial\theta_x}{\partial\pi}|_{\pi=\pi_0}}{1-f_m\lambda^T(\pi_0)g(x_i,\theta_{x0})}\right\}, \quad (4.35)$$

where

$$g'(y_j,\theta_{y0})\frac{\partial\theta_y}{\partial\pi}|_{\pi=\pi_0} = \frac{\partial g(y_j,\theta_y)}{\partial\pi}|_{\pi=\pi_0}$$

and

$$g'(x_i,\theta_{x0})\frac{\partial\theta_x}{\partial\pi}|_{\pi=\pi_0} = \frac{\partial g(x_i,\theta_x)}{\partial\pi}|_{\pi=\pi_0}.$$

Under the conditions of the lemma, we can show that $\lambda(\pi_0) = O_p(n^{-1/2})$ and $J(\pi_0) = O_p(n^{1/2})$. Also, applying Taylor expansion to $l(\pi)$ in a small neighbourhood of π_0, $\{\pi : \|\pi-\pi_0\| \le \kappa n^{-1/2}\}$, where κ is some positive constant, we obtain

$$l(\pi) = l\{\pi_0 + (\pi-\pi_0)\} = l(\pi_0) + J(\pi_0)(\pi-\pi_0) + O_p(1). \quad (4.36)$$

By Theorem 4.1, $l(\pi_0) = O_p(1)$. This and (4.36) imply that for a π in that small neighbourhood,

$$l(\pi) = O_p(1). \quad (4.37)$$

To show part (i), since $h_N^C(\pi_0') = \pi_0$, we have

$$\pi_0 - \tilde{\pi} = \gamma\{n, l(\pi_0')\}(\pi_0' - \tilde{\pi}). \tag{4.38}$$

Noting that $\gamma\{N, l(\pi)\} \geq 1$, (4.38) implies that π_0' is on the ray originating from $\tilde{\pi}$ through π_0 and

$$\|\pi_0 - \tilde{\pi}\| \geq \|\pi_0' - \tilde{\pi}\|.$$

Hence, $\pi_0' \in [\tilde{\pi}, \pi_0]$ and part (i) of the lemma 4.2 is proven.

To show part (ii), since $\tilde{\pi}$ is $\sqrt{n}$-consistent and $\pi_0' \in [\tilde{\pi}, \pi_0]$, we have $\pi_0' - \pi_0 = O_p(n^{-1/2})$. It follows from (4.37) that $l(\pi_0') = O_p(1)$. This implies

$$\gamma\{N, l(\pi_0')\} = 1 + \frac{l(\pi_0')}{2N} = 1 + O_p(n^{-1}). \tag{4.39}$$

Adding and subtracting a π_0 on the right-hand side of (4.38) gives

$$\pi_0 - \tilde{\pi} = \gamma\{N, l(\pi_0')\}(\pi_0' - \pi_0 + \pi_0 - \tilde{\pi}).$$

This implies that

$$[1 - \gamma\{N, l(\pi_0')\}]\,(\pi_0 - \tilde{\pi}) = \gamma\{N, l(\pi_0')\}(\pi_0' - \pi_0). \tag{4.40}$$

It follows from (4.39), (4.40) and $\tilde{\pi} - \pi_0 = O_p(n^{-1/2})$ that

$$\pi_0' - \pi_0 = O_p(n^{-3/2}).$$

This proves part (ii) of the lemma 4.1. $\square$

Proof for Theorem 4.3 follows easily from that for Theorem 1 in Tsao and Wu

(2014).

Proof of Theorem 4.4

By (ii) of Lemma 4.2, $\pi_0' - \pi_0 = O_p(n^{-3/2})$. Taylor expansion of $l^*(\pi_0)$ gives

$$l^*(\pi_0) = l(\pi_0') = l\{\pi_0 + (\pi_0' - \pi_0)\} = l(\pi_0) + J(\pi_0)(\pi_0' - \pi_0) + o_p(n^{-3/2}). \tag{4.41}$$

Since $J(\pi_0) = O_p(n^{1/2})$, (4.41) implies that $l^*(\pi_0) = l(\pi_0) + O_p(n^{-1})$. Thus, the extended empirical log-likelihood ratio $l^*(\pi_0)$ has the same limiting χ_q^2 distribution as the original empirical log-likelihood ratio $l(\pi_0)$. □

We need the following lemma for the proof of Theorem 4.5.

Lemma 4.3. *Assume conditions 1, 2, 3 and 4 hold. Under the composite similarity mapping (4.14) with expansion factor $\gamma\{N, l(\pi)\} = \gamma_2\{N, l(\pi)\}$ in (4.22), we have*

$$\pi_0' - \pi_0 = \frac{\eta}{2n}(\tilde{\pi} - \pi_0) + O_p(n^{-2}). \tag{4.42}$$

Proof of Lemma 4.3

It may be verified that under the three conditions and with the composite similarity mapping h_N^C defined by (4.14) and (4.22), Theorem 4.1, Lemma 4.2 and Theorem 4.2 all hold. In particular, $\pi_0' - \pi_0 = O_p(n^{-3/2})$ and the extended empirical log-likelihood ratio $l_2^*(\pi_0)$ converges in distribution to a χ_q^2 random variable.

Since $\delta(n) = O(n^{-1/2})$ and $l(\pi_0') = l_2^*(\pi_0)$ which is asymptotically a χ_q^2 variable, we have

$$\{l(\pi_0')\}^{\delta(n)} = 1 + O_p(n^{-1/2}). \tag{4.43}$$

By $h_N^C(\pi_0') = \pi_0$, we have $\pi_0 - \tilde{\pi} = \gamma_2\{N, l(\pi_0')\}(\pi_0' - \tilde{\pi})$. Thus,

$$\pi_0' - \pi_0 = \frac{\eta\{l(\pi_0')\}^{\delta(n)}}{2N}(\tilde{\pi} - \pi_0') = \frac{\eta\{l(\pi_0')\}^{\delta(n)}}{2N}(\tilde{\pi} - \pi_0) + \frac{\eta\{l(\pi_0')\}^{\delta(n)}}{2N}(\pi_0 - \pi_0'). \quad (4.44)$$

It follows from (4.43), (4.44) and $\pi_0' - \pi_0 = O_p(n^{-3/2})$ that

$$\begin{aligned} \pi_0' - \pi_0 &= \frac{\eta\{l(\pi_0')\}^{\delta(n)}}{2N}(\tilde{\pi} - \pi_0) + O_p(n^{-5/2}) \\ &= \frac{\eta}{2N}(\tilde{\pi} - \pi_0) + O_p(n^{-2}), \end{aligned}$$

which proves the lemma. □

Proof of Theorem 4.5

Under conditions 1, 2, 3 and 4, based on the (4.9) we can show that $l(\pi_0)$ has the following expansion

$$l(\pi_0) = N(R_1 + R_2 + R_3)^T(R_1 + R_2 + R_3) + N\Delta + O_p(n^{-3/2}), \quad (4.45)$$

where R_i and Δ are functions of $S^{t_1 t_2 \ldots t_l}$ and $S_1^{t_1 t_2 \ldots t_l}$ with

$$\begin{aligned} R_1^\tau &= S^\tau, \quad R_2^\tau = -\frac{1}{2}S^{\tau\upsilon}S^\upsilon + \frac{1}{3}s^{\tau\upsilon\omega}S^\upsilon S^\omega, & (4.46) \\ R_3^\tau &= \frac{3}{8}S^{\tau\upsilon}S^{\upsilon\omega}S^\omega - \frac{5}{12}s^{\tau\upsilon\omega}S^{\omega\alpha}S^\upsilon S^\alpha - \frac{5}{12}s^{\upsilon\omega\alpha}S^{\tau\upsilon}S^\omega S^\alpha \\ &\quad + \frac{4}{9}s^{\tau\upsilon\omega}s^{\omega\alpha\beta}S^\upsilon S^\alpha S^\beta + \frac{1}{3}S^{\tau\upsilon\omega}S^\upsilon S^\omega - \frac{1}{4}s^{\tau\upsilon\omega\alpha}S^\upsilon S^\omega S^\alpha, & (4.47) \\ \Delta &= (S^{\tau\upsilon} - S_1^{\tau\upsilon})S^\tau S^\upsilon + \frac{2}{3}(S_1^{\tau\alpha\beta} - S^{\tau\alpha\beta})S^\tau S^\alpha S^\beta, & (4.48) \end{aligned}$$

where for a vector P, P^r means its rth component. Based on the proofs of (4.9), we have

$$
\begin{aligned}
&(i) \quad R_j = O_p(n^{-j/2}) \quad \text{for } j = 1, 2, 3, && (4.49)\\
&(ii) \quad D_1 = \frac{1}{n}\sum g(y_j, \theta_y^0) - \frac{1}{m}\sum g(x_i, \theta_x^0) = O_p(n^{-1/2}), && (4.50)\\
&(iii) \quad \lambda(\pi_0) = V^{-1}D_1 + O_p(n^{-1}), && (4.51)\\
&(iv) \quad R_1^T R_1 = D_1^T V^{-1} D_1, && (4.52)\\
&(v) \quad \Delta = O_p(n^{-3/2}). && (4.53)
\end{aligned}
$$

It may be verified that Lemma 4.2, Theorem 4.1 and Theorem 4.2 all hold under $\gamma_2(N, l(\pi))$. Hence, the limiting distribution of $l_2^*(\pi_0)$ is also χ_q^2. This and the condition that $\delta(n) = O(n^{-1/2})$ imply

$$
[l(\pi_0')]^{\delta(n)} = 1 + O_p(n^{-1/2}). \tag{4.54}
$$

Since $h_N^C(\pi_0') = \pi_0$, by (4.14) and (4.22), we have

$$
\begin{aligned}
\pi_0' - \pi_0 &= \frac{\eta\,\{l(\pi_0')\}^{\delta(n)}}{2N}(\tilde{\pi} - \pi_0')\\
&= \frac{\eta\,\{l(\pi_0')\}^{\delta(n)}}{2N}(\tilde{\pi} - \pi_0) + \frac{\eta[l(\pi_0')]^{\delta(n)}}{2N}(\pi_0 - \pi_0'). \qquad (4.55)
\end{aligned}
$$

By the part (ii) of Lemma 4.2, (4.54) and (4.55), we find that

$$
\begin{aligned}
\pi_0' - \pi_0 &= \frac{\eta\,\{l(\pi_0')\}^{\delta(n)}}{2N}(\tilde{\pi} - \pi_0) + O_p(n^{-5/2})\\
&= \frac{\eta}{2N}(\tilde{\pi} - \pi_0) + O_p(n^{-2}). \qquad (4.56)
\end{aligned}
$$

By (4.42) from Lemma 4.3 and Taylor expansion (4.41), we have

$$\begin{aligned} l_2^*(\pi_0) &= l(\pi_0) + J(\pi_0)(\pi_0' - \pi_0) + o_p(n^{-3/2}) \\ &= l(\pi_0) + \frac{\eta}{2n} J(\pi_0)(\tilde{\pi} - \pi_0) + O_p(n^{-3/2}), \end{aligned} \tag{4.57}$$

where $J(\pi_0)$ is given by (4.35). Define

$$G(\pi) = \frac{1}{n} \sum g(y_j, \theta_y) - \frac{1}{m} \sum g(x_i, \theta_x). \tag{4.58}$$

Under condition 2, Taylor expansion of $G(\tilde{\pi})$ at π_0 gives

$$\begin{aligned} G(\tilde{\pi}) = \quad & G(\pi_0) + G'(\pi_0)(\tilde{\pi} - \pi_0) + O_p(\|\pi_0 - \tilde{\pi}\|^2) \\ = \quad & \left\{ \frac{1}{n} \sum g(y_j, \theta_{y0}) - \frac{1}{m} \sum g(x_i, \theta_{x0}) \right\} + \\ & \left\{ \frac{1}{n} \sum g'(y_j, \theta_{y0}) \frac{\partial \theta_y}{\partial \pi}|_{\pi_0} - \frac{1}{m} \sum g'(x_i, \theta_{x0}) \frac{\partial \theta_x}{\partial \pi}|_{\pi_0} \right\} \\ & (\tilde{\pi} - \pi_0) + O_p(n^{-1}). \end{aligned}$$

Under the condition that the estimating equations are just-determined, $G(\tilde{\pi}) = 0$. Hence, the above expansion implies

$$\begin{aligned} & \left\{ \frac{1}{n} \sum g'(y_j, \theta_{y0}) \frac{\partial \theta_y}{\partial \pi}|_{\pi_0} - \frac{1}{m} \sum g'(x_i, \theta_{x0}) \frac{\partial \theta_x}{\partial \pi}|_{\pi_0} \right\} (\pi_0 - \tilde{\pi}) \\ & = \left\{ \frac{1}{n} \sum g(y_j, \theta_{y0}) - \frac{1}{m} \sum g(x_i, \theta_{x0}) \right\} + O_p(n^{-1}). \end{aligned} \tag{4.59}$$

Noting that $\lambda(\pi_0) = O_p(n^{-1/2})$ and $\pi_0 - \tilde{\pi} = O_p(n^{-1/2})$, we can show

$$\begin{aligned} & \left\{ \frac{1}{n} \sum \frac{g'(y_j, \theta_{y0}) \frac{\partial \theta_y}{\partial \pi}|_{\pi=\pi_0}}{1 + f_n \lambda^T(\pi_0) g(y_j, \theta_{y0})} - \frac{1}{m} \sum \frac{g'(x_i, \theta_{x0}) \frac{\partial \theta_x}{\partial \pi}|_{\pi=\pi_0}}{1 - f_m \lambda^T(\pi_0) g(x_i, \theta_{x0})} \right\} (\pi_0 - \tilde{\pi}) \\ & = \left\{ \frac{1}{n} \sum g'(y_j, \theta_{y0}) \frac{\partial \theta_y}{\partial \pi}|_{\pi_0} - \frac{1}{m} \sum g'(x_i, \theta_{x0}) \frac{\partial \theta_x}{\partial \pi}|_{\pi_0} \right\} (\pi_0 - \tilde{\pi}) + O_p(n^{-1}). \end{aligned} \tag{4.60}$$

It follows from (4.59) and (4.60) that

$$\left\{\frac{1}{n}\sum\frac{g'(y_j,\theta_{y0})\frac{\partial\theta_y}{\partial\pi}|_{\pi=\pi_0}}{1+f_n\lambda^T(\pi_0)g(y_j,\theta_{y0})}-\frac{1}{m}\sum\frac{g'(x_i,\theta_{x0})\frac{\partial\theta_x}{\partial\pi}|_{\pi=\pi_0}}{1-f_m\lambda^T(\pi_0)g(x_i,\theta_{x0})}\right\}(\pi_0-\tilde{\pi})$$
$$=\left\{\frac{1}{n}\sum g(y_j,\theta_{y0})-\frac{1}{m}\sum g(x_i,\theta_{x0})\right\}+O_p(n^{-1}). \tag{4.61}$$

By (4.57), (4.35) and (4.61), we have

$$\begin{aligned}
l_2^*(\pi_0) &= l(\pi_0)+\frac{\eta}{2N}J(\pi_0)(\tilde{\pi}-\pi_0)+O_p(n^{-3/2})\\
&= l(\pi_0)-\frac{\eta}{2N}2N\lambda^T(\pi_0)\frac{1}{N}\\
&\quad\left\{\sum\frac{f_ng'(y_j,\theta_{y0})\frac{\partial\theta_y}{\partial\pi}|_{\pi=\pi_0}}{1+f_n\lambda^T(\pi_0)g(y_j,\theta_{y0})}-\sum\frac{f_mg'(x_i,\theta_{x0})\frac{\partial\theta_x}{\partial\pi}|_{\pi=\pi_0}}{1-f_m\lambda^T(\pi_0)g(x_i,\theta_{x0})}\right\}(\tilde{\pi}-\pi_0)+O_p(n^{-3/2})\\
&= l(\pi_0)-\frac{\eta}{N}N\lambda^T(\pi_0)\left\{\frac{1}{n}\sum g(y_j,\theta_{y0})-\frac{1}{m}\sum g(x_i,\theta_{x0})\right\}+O_p(n^{-3/2}).
\end{aligned}\tag{4.62}$$

Finally, by(4.45), (4.62), and from (4.49) to (4.53), we have

$$\begin{aligned}
l_2^*(\pi_0) &= l(\pi_0)-\frac{\eta}{N}\left\{N(V^{-1}D_1+O_p(n^{-1}))^T(D_1+O_p(n^{-1}))\right\}+O_p(n^{-3/2})\\
&= l(\pi_0)-\frac{\eta}{N}NR_1^TR_1+O_p(n^{-3/2})\\
&= l(\pi_0)-\frac{\eta}{N}\{N(R_1+R_2+R_3)^T(R_1+R_2+R_3)+N\Delta\}+O_p(n^{-3/2})\\
&= l(\pi_0)-\frac{\eta}{N}l(\pi_0)+O_p(n^{-3/2})\\
&= l(\pi_0)\left\{1-\frac{\eta}{N}+O_p(n^{-3/2})\right\},
\end{aligned}$$

which proves Theorem 4.5. □

Remark. The second-order result of Theorem 4.5 holds only for the just-determined case as the proof above used the condition that $G(\tilde{\pi})=0$ to obtain (4.59). For over-determined cases, a weaker condition $G(\tilde{\pi})=O_p(n^{-1})$ would also allow us to get (4.59). However, it is not clear that outside of the just-determined cases when this weaker condition would hold. When this weaker condition does not hold, the extended empirical log-likelihood ratio $l_2^*(\pi)$ defined in Theorem 4.5 reduces to a first-order extended empirical log-likelihood ratio as Theorem 4.4 is still valid for $l_2^*(\pi)$.

Chapter 5

Concluding Remarks

The literature on the empirical likelihood method has been growing at a remarkable speed over the last twenty-five years. There is no sign that this growth is slowing down as more and more applications of this method are appearing in many areas of statistical research.

The common theme of the three papers in this thesis is to overcome a fundamental problem of the empirical likelihood, namely, the mismatch of its domain to the parameter space. Through the extended empirical likelihood, these papers have successfully resolved the mismatch issue for two large families of empirical likelihoods covering that for parameters defined by one-sample and two-sample estimating equations. These results add to the theoretical foundation of this powerful method. They also bring about substantially improved accuracy to the empirical likelihood inference, especially for small sample and high dimension situations. They will be very useful to practitioners of the empirical likelihood method.

Due to space constraints for journal papers, we have not included much details about computation of the extended empirical likelihood in the preceding chapters. We conclude this thesis with the following discussion on the computation. Let $l(\theta)$ be the original empirical log-likelihood ratio of a value $\theta \in \mathbb{R}^d$ and let $l^*(\theta)$ be the extended empirical log-likelihood ratio. Use the usual notation $h_n^C(\theta) : \Theta_n \to \mathbb{R}^d$ for the composite similarity mapping where Θ_n is the domain of the original empirical log-likelihood ratio. Then,

$$l^*(\theta) = l(\theta')$$

where θ' satisfies

$$h_n^C(\theta') = \theta. \tag{5.1}$$

The main computational effort for evaluating the extended empirical likelihood $l^*(\theta)$ at a given θ value is in computing the corresponding θ', that is, in finding the solution θ' to equation (5.1).

To see why this can be potentially time consuming, as the analytic expression of $l(\theta)$ and hence that of $h_n^C(\theta)$ (which involves $l(\theta)$ through the expansion factor γ) are not available, equation (5.1) has to be solved numerically through Newton's method for finding roots. This iterative method involves evaluating $l(\theta)$ repeatedly, which can be time consuming, especially for multi-dimensional parameters where $d \geq 2$. Further, if the initial value of Newton's iteration is not properly chosen, then its convergence is not assured.

The key to develop an efficient and reliable algorithm to compute the extended empirical likelihood is to take advantage of the theoretical result that shows the location of θ_0' and its distance to true parameter value θ_0; in Chapter 2 this is Lemma 2.1, in Chapter 3 this is Lemma 3.2, and in Chapter 4 this is Lemma 4.2. Specifically, these lemmas show that

(i) $\theta_0' \in [\tilde{\theta}, \theta_0]$ and (ii) $\theta_0' - \theta_0 = O_p(n^{-3/2})$

where $\tilde{\theta}$ is the maximum empirical likelihood estimator of θ_0. A careful reading of the proofs of these lemmas reveals that (i) actually holds for any θ value; that is, for any $\theta \in \mathbb{R}^d$ we have $\theta' \in [\tilde{\theta}, \theta]$. Since $\tilde{\theta}$ is available, for a given θ this means the corresponding θ' is in the known interval $[\tilde{\theta}, \theta]$. Consequently, θ' can be expressed as

$$\theta' = w\theta + (1 - w)\tilde{\theta}$$

for some $w \in [0, 1]$. Hence, the problem of finding solution θ' to (5.1) is equivalent to finding the w value in $[0, 1]$ that satisfies

$$h_n^C(w\theta + (1 - w)\tilde{\theta}) = \theta, \tag{5.2}$$

which is a much easier univariate root finding problem over the short interval $[0, 1]$. Further, part (ii) of the lemmas also suggests that θ would be a good initial value for Newton's iteration, which is equivalent to using $w = 1$ as the initial value. To see this, the width of the interval $[\tilde{\theta}, \theta_0]$ is typically $O_p(n^{-1/2})$. So part (ii) of the lemmas suggests that θ_0' is a point in the interval that is closer to the end-point θ_0 than to the other end-point $\tilde{\theta}$. Hence, we also expect θ' to be closer to θ than to $\tilde{\theta}$. Alternatively, one could simply use the mid-point, $(\theta + \tilde{\theta})/2$ or $w = 0.5$ as the initial value.

We have used the above observations to reformulate the computational problem for the extended empirical likelihood and to select initial values for Newton's iterations. We developed an efficient and reliable R code based on these. We hope to develop an R package for computing the extended empirical likelihood and make it available through the R software library.

Bibliography

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

[2] BEE, ADAM (2012). Household Income Inequality Within U.S. Counties: 2006-2010. *Census Bureau, U.S. Department of Commerce.*

[3] CHEN, J. AND HUANG, Y. (2012). Finite-sample properties of the adjusted empirical likelihood. *J. Nonparametric Statistics*, **25**, 147–159.

[4] CHEN, J., VARIYATH, A. M. AND ABRAHAM, B. (2008). Adjusted empirical likelihood and its properties. *J. Comput. Graph. Statist*, **3**, 426–443.

[5] CHEN, S. X. AND CUI, H. J. (2007). On the second-order properties of empirical likelihood with moment restrictions. *J. Econometrics*. **141**, 492–516.

[6] CHEN, WEN-HAO (2009). Cross-national Differences in Income Mobility: Evidence from Canada, the United States, Great Britain and Germany. *Review of Income and Wealth*, **55** (1): 75–100

[7] DOMEIJ, DAVID AND FLODéN, MARTIN (2010). Inequality trends in Sweden 1978-2004. *Review of Economic Dynamics*, Volume 13, Issue 1, January 2010, Pages 179-208

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

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

[10] Fan, J., Leeuwner, W. and Hunter, L. (1997). Compatibility of Outer and Fusible Interlining Fabrics in Tailored Garments. *Textile Research Journal.*, 67:137.

[11] Gini, C. (1936). On the Measure of Concentration with Special Reference to Income and Statistics. Colorado College Publication, *General Series No. 208*, 73-79.

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

[13] Jing, B. Y. (1995). Two-sample empirical likelihood method. *Statistics & Probability Letters.*, **24**, 315–319.

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

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

[16] Liu,Y. K., Zou, C. L. and Zhang, R. C. (2008). Empirical likelihood for two-sample mean problem. *Statistics & Probability Letters*, **78**, 548–556.

[17] Liu, Y. K. and Yu, C. W. (2010). Bartlett correctable two-sample adjusted empirical likelihood. *Journal of Multivariate Analysis*, **101**, 1701-1711.

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

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

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

[21] Peng, Liang (2011). Empirical Likelihood methods for the Gini index. *Australian & New Zealand Journal of Statistics*, **53(2)**, 131–139.

[22] Qin, J. (1994). Semi-parametric likelihood ratio confidence intervals for the difference of two sample means. *Ann. Insti. Statist. Math.*, **46**, 117–126.

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

[24] Qin, Yongsong, Rao, J.N.K. and Wu, Changbao (2010). Empirical likelihood confidence intervals for the Gini measure of income inequality. *Economic Modelling*, **27**, 1429-1435.

[25] Qin, Y. and Zhao, L. (2000). Empirical likelihood ratio confidence intervals for various differences of two populations. *Syst. Sci. Math. Sci.*, **13**, 23-30.

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

[27] Tsao, M. (2013). Extending the empirical likelihood by domain expansion. *Canadian J. Statist.*, **41**, 257–274.

[28] Tsao, M. and Wu, F. (2013). Empirical likelihood on the full parameter space. *Ann. Statist.*, **41**, 2176–2196.

[29] Tsao, M. and Wu, F. (2014). Extended empirical likelihood for estimating equations. *Biometrika*, **101**, 703–710.

[30] Wu, F. and Tsao, M. (2014). Two-sample extended empirical likelihood. *Stat Probab Lett*, **84**, 81–87.

[31] Wu, C. and Yan, Y. (2012). Empirical likelihood inference for two-sample problems. *Stat. Interface.*, **5**, 345–354.

[32] Zi, X. M., Zou, C. L. and Liu, Y. K.(2012). Two-sample empirical likelihood method for difference between coefficients in linear regression model. *Statistical Papers, Springer*, vol. 53(1), pages 83-93.