Methods of Calibration for the Empirical Likelihood Ratio

by

Li (Jenny) Jiang
B.Sc., Peking University, 2003

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

MASTER OF SCIENCE

in the Department of Mathematics and Statistics

Methods of Calibration for the Empirical Likelihood Ratio

by

Li (Jenny) Jiang
B.Sc., Peking University, 2003

Supervisory Committee

Dr. M. Tsao, (Department of Mathematics and Statistics)

Supervisor

Dr. M. Lesperance, (Department of Mathematics and Statistics)

Departmental Member

Dr. J. Zhou, (Department of Mathematics and Statistics)

Departmental Member

Dr. C.B. Wu, (Department of Statistics and Actuarial Sciences,University of Waterloo)

Outside Member

Supervisory Committee

Dr. M. Tsao, (Department of Mathematics and Statistics)

Supervisor

Dr. M. Lesperance, (Department of Mathematics and Statistics)

Departmental Member

Dr. J. Zhou, (Department of Mathematics and Statistics)

Departmental Member

Dr. C.B. Wu, (Department of Statistics and Actuarial Sciences,University of Waterloo)

Outside Member

ABSTRACT

This thesis provides several new calibration methods for the empirical log-likelihood ratio. The commonly used Chi-square calibration is based on the limiting distribution of this ratio but it constantly suffers from the undercoverage problem. The finite sample distribution of the empirical log-likelihood ratio is recognized to have a mixture structure with a continuous component on $[0, +\infty)$ and a probability mass at $+\infty$. Consequently, new calibration methods are developed to take advantage of this mixture structure; we propose new calibration methods based on the mixture distributions, such as the mixture Chi-square and the mixture Fisher's F distribution. The E distribution introduced in Tsao (2004a) has a natural mixture structure and the calibration method based on this distribution is considered in great details. We also discuss methods of estimating the E distributions.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The empirical likelihood method, introduced by Owen (1988, 1990), is a method for conducting hypothesis test and constructing confidence regions. It is based on the concept of the empirical likelihood, which is an extension of the concept of the parametric likelihood to a nonparametric setting. Similar to the parametric likelihood ratio, the empirical log-likelihood ratio has a limiting Chi-square distribution. This result holds under very weak conditions. In particular, it requires no assumptions about the parametric form of the underlying distribution. Today, the empirical likelihood method has become an important competitor to contemporary techniques such as the parametric method and the bootstrap.

This chapter provides an introduction to the empirical likelihood method with emphasis on the methods of calibration. In Section 1.1, we give a systematic review

of the empirical likelihood method, first for the population mean and then for other parameters that can be expressed as a smooth function of the population mean or the solution of an unbiased estimating equation. In Section 1.2, we discuss the advantages and further applications of the empirical likelihood approach. In Section 1.3, we discuss existing methods of calibration and the associated undercoverage problem. We conclude this chapter with an overview of this thesis.

1.1 The Empirical Likelihood Method

1.1.1 The Nonparametric Maximum Likelihood

In this part, we begin by defining the empirical cumulative distribution function and showing that this distribution function is a nonparametric maximum likelihood estimate (*NPMLE*) of the underlying distribution F_0.

For a random variable $X \in \mathcal{R}^1$, the cumulative distribution function (*CDF*) is the function $F(x) = P(X \leq x)$, for $-\infty < x < +\infty$. We use $F(x-)$ to denote $P(X < x)$ and so $P(X = x) = F(x) - F(x-)$. In the following definition, the notation $1_{A(x)}$ represents the value 1 if the assertion $A(x)$ is true, and 0 otherwise.

Definition 1.1 *Empirical Cumulative Distribution Function (ECDF)*

Let $X_1, ..., X_n$ be a random sample in $\mathcal{R}^1$ from some distribution F_0. The empirical

distribution function (ECDF) for the sample is

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1_{X_i \leq x},$$

for $-\infty < x < +\infty$.

The nonparametric likelihood ratio is defined below by taking very literally the notion that likelihood is the probability of observing the actual data values at hand.

Definition 1.2 *Nonparametric Likelihood*

Let $X_1, ..., X_n$ *be a random sample in* $\mathcal{R}^1$ *from some distribution* F_0. *The nonparametric likelihood of the cumulative distribution function (CDF)* F *is*

$$L(F) = \prod_{i=1}^{n} (F(X_i) - F(X_i-)).$$

The fact that F_n is the *NPMLE* was apparently first noticed by Kiefer and Wolfowitz (1956).

Theorem 1.1 *Let* $X_1, ..., X_n$ *be a random sample in* $\mathcal{R}^1$ *from some distribution* F_0. *Let* F_n *be their ECDF and let* F *be any CDF on* $\mathcal{R}^1$. *If* $F \neq F_n$, *then* $L(F) \leq L(F_n)$.

Proof of Theorem 1.1:

Let $Y_1, ..., Y_k$ be the distinct values observed among $X_1, \ldots, X_n$. The number of times that Y_i appeared in this sample is denoted as n_i ($n_i \geq 1$). Let $p_i = F(Y_i) - F(Y_i-)$ and put $\hat{p}_i = n_i/n$. By $\log(z) \leq z - 1$ and $\sum_{i=1}^{k} p_i \leq 1$, we have

$$\begin{aligned}
\log\left(\frac{L(F)}{L(F_n)}\right) &= \sum_{i=1}^{k} n_i \log\left(\frac{p_i}{\hat{p}_i}\right) \\
&\leq n\sum_{i=1}^{k} \hat{p}_i \log\left(\frac{p_i}{\hat{p}_i}\right) \\
&\leq n\sum_{i=1}^{k} \hat{p}_i \left(\frac{p_i}{\hat{p}_i} - 1\right) \\
&\leq 0,
\end{aligned}$$

and so $L(F) \leq L(F_n)$.

□

Nonparametric maximum likelihood estimators of parameters can be found through the nonparametric maximum likelihood estimate (*NPMLE*) of F_0. If the true unknown parameter is $\theta_0 = T(F_0(x))$, where T is a real-valued function of the distribution function, and $\hat{F}$ is a *NPMLE* of F_0, such as F_n in the *i.i.d.* case, then $\hat{\theta} = T(\hat{F})$ is the *NPMLE* of θ. For example, the population mean is $\mu_0 = \int x dF_0(x)$. The *NPMLE* of μ_0 is the mean of F_n, $\int x dF_n(x) = (1/n)\sum_{i=1}^{n} X_i$, which is just $\bar{X}$.

The concept of *NPMLE* can be generalized to multivariate settings. See Section 3.1 in Owen (2001) for details.

1.1.2 The Empirical Likelihood Ratio

In parametric likelihood theory, the hypothesis test and the confidence region are based on the parametric likelihood ratio whose asymptotic Chi-square distribution is employed to decide the critical value for calibration. Similarly, a empirical likelihood ratio can be constructed and used as the basis for empirical likelihood inferences.

Definition 1.3 *Empirical Likelihood Ratio*

For a distribution function F, define empirical likelihood ratio as:

$$R(F) = \frac{L(F)}{L(F_n)},$$

where F_n is the ECDF.

The empirical likelihood ratio of parameters is defined through the profile empirical likelihood ratio function

$$\mathcal{R}(\theta) = \sup\{R(F) \mid T(F) = \theta, F(X_1, \ldots, X_n) = 1\},$$

where $F(X_1, \ldots, X_n) = 1$ is any distribution function supported on the sample.

Consider the population mean, μ. Let w_i be the weight that F places on the observation X_i, where $w_i > 0$ and $\sum_{i=1}^{n} w_i = 1$. For the empirical distribution function F_n, w_i is equal to $1/n$ for $i = 1, 2, \ldots, n$. The profile empirical likelihood ratio for the mean can be written as:

$$\mathcal{R}(\mu) = \max\{\prod_{i=1}^{n} nw_i \mid \sum_{i=1}^{n} w_i X_i = \mu, w_i \geq 0, \sum_{i=1}^{n} w_i = 1\}. \tag{1.1}$$

Correspondingly, the empirical likelihood confidence region for the mean μ is in the form of

$$\begin{aligned} C_{r,n} &= \{\mu : R(\mu) \geq r_0\} \\ &= \Big\{ \sum_{i=1}^{n} w_i X_i \mid \prod_{i=1}^{n} n w_i \geq r_0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \Big\}, \end{aligned} \tag{1.2}$$

where r_0 is a constant that depends on the confidence level and method of calibration.

The following theorem gives the asymptotic distribution of the empirical log-likelihood ratio.

Theorem 1.2 *Empirical Likelihood Theorem (ELT)(Owen 2001)*

Let $X_1, \ldots, X_n$ be independent random vectors in R^k with a common distribution F_0 having mean μ_0 and finite variance covariance matrix V_0 of rank $q > 0$. Then the empirical log-likelihood ratio $-2\log(R(\mu_0))$ converges in distribution to a χ^2_q random variable as $n \to \infty$.

We can use this asymptotic Chi-square distribution for calibration; based on Theorem 1.2, at level α, the null hypothesis $H_0 : \mu = \mu_0$ should be rejected when $-2\log(R(\mu_0)) > \chi^2_{q,1-\alpha}$, and the r_0 in (1.2) for a $100(1-\alpha)\%$ confidence region is $\exp(-\frac{1}{2}\chi^2_{q,1-\alpha})$.

1.1.3 Extensions

In the Section 1.1.2, the empirical likelihood methods for the mean was discussed. Here we describe two extensions. The first extension is to parameters that can be written as smooth functions of finite-dimensional vector of means. For example, since $\sigma^2 = E(X^2) - [E(X)]^2$, σ^2 is a smooth function of the mean of the random vector (X, X^2). Similarly, the correlation coefficient between two random variables X and Y,

$$\rho = \frac{E(XY) - E(X)E(Y)}{\sqrt{Var(X)Var(y)}},$$

is a smooth function of the mean of (X, Y, X^2, Y^2, XY).

In general, for a parameter $\theta_0 = h(\mu_0)$, where $\mu_0 = E(X)$ and h is a smooth function from R^k to R^q for $1 \leq q \leq k$, $\hat{\theta} = h(\bar{X})$ is the *NPMLE* for θ_0 and the profile empirical likelihood ratio function can be formulated as,

$$\mathcal{R}(\theta) = max\Big\{ \prod_{i=1}^{n} nw_i \mid h\Big(\sum_{i=1}^{n} w_i X_i \Big) = \theta, w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \Big\}. \qquad (1.3)$$

The empirical likelihood theorem also applies in this case through the following theorem.

Theorem 1.3 *(Owen 2001) Suppose that $X_1, \ldots, X_n$ are independent random vectors with a common distribution F_0 having mean μ_0 and variance covariance matrix V_0 of full rank k. Let h be a smooth function from R^k to R^p for $1 \leq p \leq k$, and*

suppose that the $k \times p$ matrix G of partial derivatives of $h(\mu)$ with respect to the components of μ has rank $q > 0$ at $\mu = \mu_0$. Define $\theta_0 = h(\mu_0)$,

$$\begin{aligned} C_{r,n}^{(1)} &= \Big\{ \sum_{i=1}^{n} w_i X_i \mid \prod_{i=1}^{n} n w_i \geq r_0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \Big\}, \\ C_{r,n}^{(2)} &= \{ h(\mu) \mid \mu \in C_{r,n}^{(1)} \}, \\ C_{r,n}^{(3)} &= \{ \theta_0 + G'(\mu - \mu_0) \mid \mu \in C_{r,n}^{(1)} \}. \end{aligned}$$

Then, as $n \to \infty$

$$Pr(\theta \in C_{r,n}^{(3)}) \to Pr(\chi^2_{(q)} \leq -2\log(r)),$$

and

$$\sup_{\mu \in C_{r,n}^{(1)}} \| h(\mu) - \theta_0 - G'(\mu - \mu_0) \| = o_p(n^{-1/2}).$$

The theorem does not say directly that the confidence region $C_{r,n}^{(2)}$, which is of the most interest, has an asymptotic Chi-square calibration. But the confidence region $C_{r,n}^{(3)}$ formed by linearizing $h(\mu)$ around $h(\mu_0)$ has a χ^2_q calibration, where q is the rank of the $\partial h(\mu_0)/\partial \mu$. The difference, $o_p(n^{-1/2})$, between these two confidence regions is an asymptotically negligible fraction of the diameter $O(n^{-1/2})$ of the confidence regions. And the stronger result $Pr(\theta \in C_{r,n}^{(2)}) \to Pr(\chi^2_{(q)} \leq -2\log(r))$ holds under somewhat stronger conditions (DiCiccio, Hall and Romano (1991)).

Another generalization is for parameters defined implicitly through estimating equations. For a random variable $X \in R^k$, a parameter of X, $\theta \in R^p$, and a vector-valued estimating function $m(X, \theta) \in R^s$, suppose $E(m(X, \theta)) = 0$. Then, θ may be

estimated by solving

$$\frac{1}{n}\sum_{i=1}^{n} m(X_i, \hat{\theta}) = 0.$$

This equation is called estimating equation. The empirical likelihood ratio function for θ is defined by

$$\mathcal{R}(\theta) = \max\Big\{ \prod_{i=1}^{n} nw_i \mid \sum_{i=1}^{n} w_i m(X_i, \theta) = 0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \Big\}. \qquad (1.4)$$

Theorem 1.4 *(Owen 2001) Let $X_1, \ldots, X_n$ be independent random vectors in R^k with a common distribution F_0. For $\theta \in \Theta \subseteq R^p$, $X \in R^k$, and $m(X, \theta) \in R^s$, let $\theta_0 \in \Theta$ be such that $Var(m(X_i, \theta_0))$ is finite and has rank $q > 0$. If θ_0 satisfies $E(m(X, \theta_0)) = 0$, then $-2\log(R(\theta_0)) \to \chi^2_q$ in distribution as $n \to \infty$.*

The above theorem holds under very weak conditions, which does not even require that $\hat{\theta} = T(F_n)$ to be a good estimate of $\theta = T(F_0)$. Qin and Lawless (1994) proved a sharper version under stronger conditions that ensure $T(F_n)$ to be a good estimator.

1.1.4 Computation Algorithms

In this section, we will discuss algorithms for evaluating the empirical likelihood ratio of the mean. More often, we will use the log-likelihood ratio:

$$\begin{aligned} l(\mu) &= -2\log(R(\mu)) \\ &= \max\{ -2\sum_{i=1}^{n} \log(nw_i) \mid \sum_{i=1}^{n} w_i X_i = \mu, w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \}. \end{aligned} \qquad (1.5)$$

By (1.5), to compute the empirical log-likelihood ratio at a given μ is to maximize $-2\sum_{i=1}^{n}\log(nw_i)$ subject to constrains $\sum_{i=1}^{n} w_i X_i = \mu$, $w_i \geq 0$, and $\sum_{i=1}^{n} w_i = 1$. This optimization problem can be solved by Lagrange multipliers method as follows:

Let

$$M = \sum_{i=1}^{n}\log(nw_i) - n\lambda'(\sum_{i=1}^{n} w_i(X_i - \mu)) + \gamma(\sum_{i=1}^{n} w_i - 1).$$

It can be shown that

$$w_i = \frac{1}{n}\frac{1}{1+\lambda'(X_i-\mu)}, \tag{1.6}$$

and $\lambda \in R^k$ has to satisfy k equations given by

$$0 = \frac{1}{n}\sum_{i=1}^{n}\frac{X_i - \mu}{1+\lambda'(X_i-\mu)}. \tag{1.7}$$

Then, $\log R(\mu)$, can be written as

$$\log R(\mu) = -\sum_{i=1}^{n}\log(1+\lambda'(X_i-\mu)). \tag{1.8}$$

Let $\mathcal{H} = \mathcal{H}(X_1, \ldots, X_n)$ be the convex hull of X_i. When μ is an interior point of $\mathcal{H}$, equation (1.7) has a unique solution (Owen 1990). A trivial case arises when $\mu \notin \mathcal{H}$. In this case (1.7) does not have solutions and by convention, we define $R(\mu) = 0$ or $l(\mu) = +\infty$. Under mild conditions, however, $P(\mu_0 \notin \mathcal{H}) \rightarrow 0$ as $n \rightarrow \infty$, where μ_0 is the true mean value.

The solution λ for (1.7) could be found by a multivariate Newton's algorithm, which is commonly used for computing numerical solutions of nonlinear equations.

To briefly describe this algorithm, let $x = (x_{(1)}, ..., x_{(m)})^T$ be an m-vector, and suppose we wish to solve a system of equations.

$$f_i(x) = 0 \quad (1 \leq i \leq m),$$

where $f_1, ..., f_m$ are known functions. Let $f = (f_1, ...f_m)^T$, $H^{ij} = \partial f_i / \partial x_{(j)}$ and $H = (H^{ij})$ (an $m \times m$ matrix). Given an initial approximation $x_{(1)}$ to the solution $x_{(0)}$ of $f(x) = 0$, develop successive approximations $x_{(j)}$ by iteration:

$$x_{(j+1)} = x_{(j)} - H(x_{(j)})^{-1} f(x_{(j)}), \quad j \geq 1.$$

If the initial solution $x_{(1)}$ is sufficiently close to $x_{(0)}$, and H is continuous in a neighborhood of $x_{(0)}$ and nonsingular at $x_{(0)}$, then $x_{(j)} \to x_{(0)}$ as $j \to \infty$.

The solution to (1.7), λ, may be found by other algorithms, some of which are listed in Owen (1990). Algorithms for more complicated cases are discussed in Hall and La Scala (1990) and Wood, Do and Broom (1994).

For computations and simulations in this thesis, we will be using a S-plus code for computing the empirical likelihood ratio developed by Owen. The code is based on the algorithm described above.

1.1.5 A Numerical Example

As an example of the empirical likelihood confidence regions for the mean, we construct the confidence regions for the mean vector of a " bi-variate" Chi-square random

vector, $X = (X_1, X_2)$, where X_1 and X_2 are independent χ^2_1 random variables.

A sample of 20 values of X were simulated, which is showed in Table 1.1. The confidence regions based on this sample for nominal confidence levels 0.5, 0.9, 0.95 and 0.99 are plotted in Figure 1.1.

i	(X_1, X_2)	i	(X_1, X_2)
1	(0.08497, 1.28724)	11	(0.01779, 2.34564)
2	(0.84066, 0.09249)	12	(2.67303, 0.31460)
3	(0.53594, 2.39307)	13	(0.14290, 1.00220)
4	(1.39065, 4.27176)	14	(1.62709, 0.16095)
5	(0.21680, 0.04723)	15	(0.02112, 0.65333)
6	(1.15714, 0.24674)	16	(0.00552, 1.42659)
7	(0.00013, 0.85608)	17	(1.32200, 1.84956)
8	(3.37054, 0.99508)	18	(3.27765, 0.71281)
9	(0.67170, 0.02843)	19	(0.44006, 0.76136)
10	(0.02440, 3.15762)	20	(1.43427, 0.09708)

Table 1.1: Simulated "bivariate" χ^2_1 sample of size $n = 20$.

The boundary of each confidence region is given by: $\{\mu : -2\log R(\mu) = c\}$, where, by the Chi-square calibration, c is $\chi^2_{2,1-\alpha}$ for $\alpha = 0.5$, 0.1, 0.05 and 0.01. The log-likelihood ratio, $-2\log R(\mu)$, is evaluated using the S-plus code by Owen.

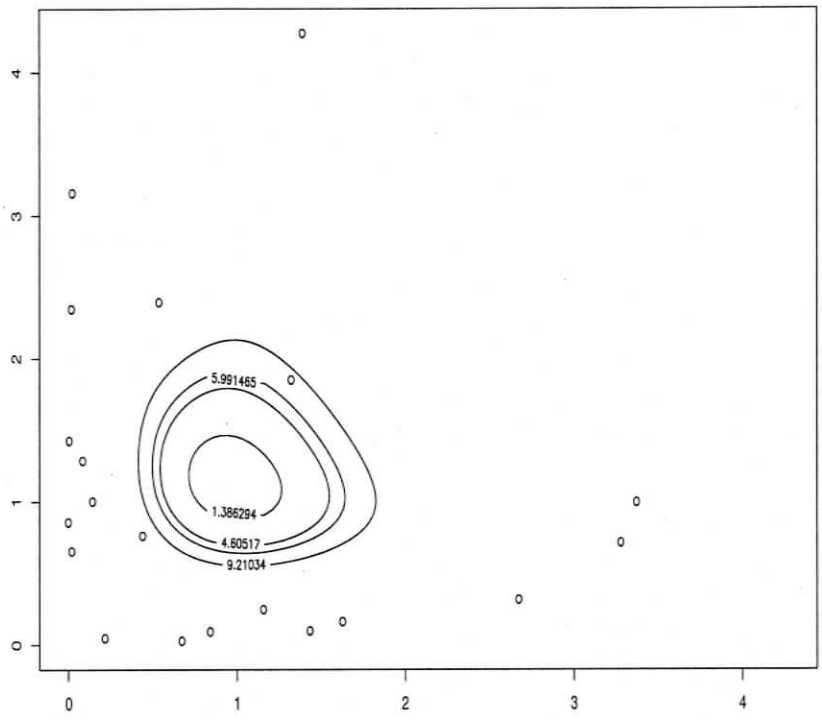

Figure 1.1: The empirical likelihood confidence regions for the mean of a "bivariate" χ^2_1 distribution.

1.2 Advantages and Extensions of the Empirical Likelihood Method

1.2.1 The Shape of Empirical Likelihood Confidence Regions

Now we review the construction of the parametric likelihood confidence region for the mean of a bivariate normal distribution $MVN(\mu_0, \Sigma)$ where Σ is unknown. Suppose $X_1, \ldots, X_n$ is a random sample from this distribution. Let $\bar{X}$ be the mean and S be the variance covariance matrix of the sample. Then a $100(1-\alpha)\%$ confidence region for μ_0 based on the normal assumption is

$$\{\mu : n(\bar{X}-\mu)^T S^{-1}(\bar{X}-\mu) < T^2(1-\alpha)\}, \tag{1.9}$$

where T^2 is the Hotelling's T^2 distribution whose $(1-\alpha)th$ quantile is determined by that of the F distribution through equation

$$T^2(1-\alpha) = \frac{k(n-1)}{(n-k)} F_{k,n-k}(1-\alpha).$$

The normal assumption based confidence regions (1.9) are often used even when the underlying distribution is unknown, and for such cases (1.9) is justified asymptotically.

Also, recall that the empirical likelihood confidence region is:

$$\{\mu : -2\log R(\mu) < \chi^2_{k,1-\alpha}\} \tag{1.10}$$

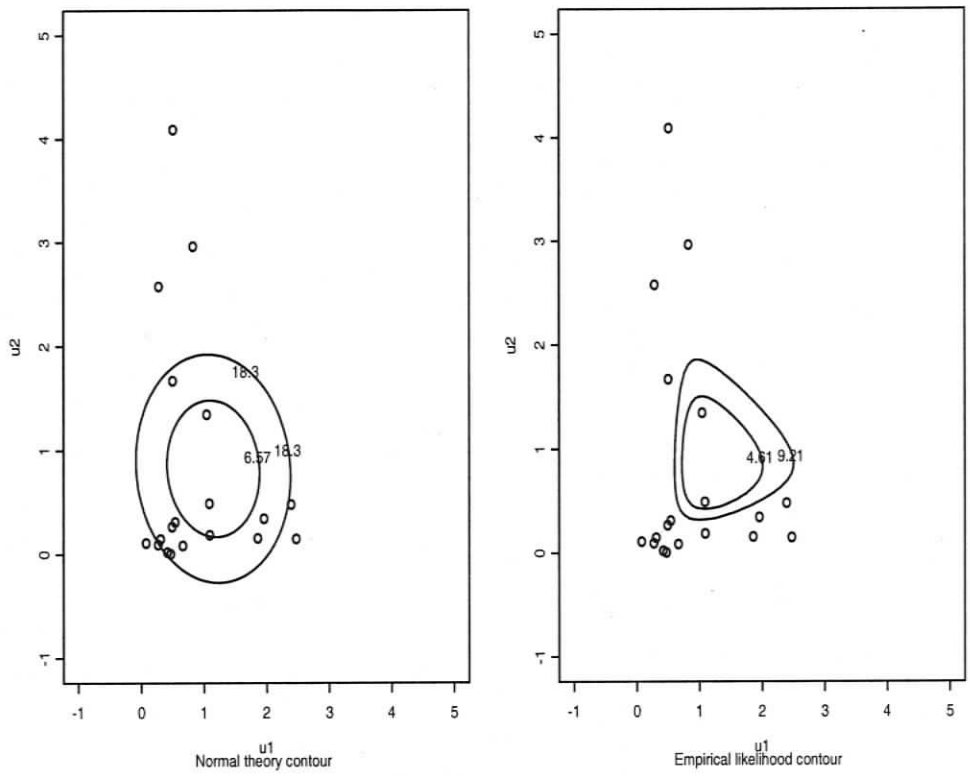

Figure 1.2: The normal theory confidence regions (Left) and the empirical likelihood confidence regions (Right) for the mean of the "bivariate" χ^2_1 distribution.

Figure 1.2 shows both empirical likelihood and normal likelihood confidence regions based on a random sample of 20 "bivariate" χ^2_1 observations. The figure on the left column represents normal theory based confidence regions $\{\mu : n(\bar{X}-\mu)^T S^{-1}(\bar{X}-\mu) < c\}$ where c is decided by $P(38/18F_{2,18} < c) = 0.90$ and 0.99, respectively. The figure on the right column represents the empirical likelihood confidence regions $\{\mu : -2\log R(\mu) < c\}$ with c calculated by $P(\chi^2_2 < c) = 0.90$ and 0.99, respectively.

Note that the shape of empirical likelihood confidence regions is quite interesting because it is determined by the data automatically. The regions reflect the emphasis in the observed data set. This attractive feature is not shared by the normal theory based regions, which are in a predefined ellipsoidal shape that implies a nonexistent ellipsoidal symmetry. In practice, only relatively complex bootstrap methods can give regions which emulate this property of "letting the data determine the shape of the region" (Hall and La Scala 1990).

1.2.2 Other Advantages and Extensions

Besides the shape of confidence regions, another key advantage of the empirical likelihood ratio is that it is Bartlett correctable. That is, a simple mean correction to the likelihood ratio reduces the coverage error from order n^{-1} to order n^{-2}. On the other hand, similar improvement of the coverage accuracy of other methods such as bootstrap call for large increase in iterations, which requires expensive computations.

Empirical likelihood confidence regions are also range preserving and transformation respecting. For variances, the confidence regions contain no negative values; for probabilities, they are between 0 and 1; and for correlations, they fall into interval $(-1, 1)$. For transformations in the form of $\eta = g(\theta)$, if the empirical likelihood confidence region for θ is (θ_1, θ_2), then that for η is $\{\eta : \eta = g(\theta), \theta \in (\theta_1, \theta_2)\}$. The empirical likelihood method requires less conditions to apply. It does not require a pivotal statistic. On the other hand, the parametric method is not applicable to the circumstances where pivotalness is difficult to achieve. Lastly and most importantly, the empirical likelihood inferences enjoy comparable accuracy as those based on other nonparametric approaches, such as the delta method, the jackknife and the simpler bootstrap methods.

Owen (1991) studied cases where the random variables are not independently and identically distributed, and gave a modified empirical likelihood theorem that applies to such cases. Owen (1992) also considered estimating equations derived from certain projection pursuit models. Hall and Owen (1993) provided an empirical likelihood theorem for kernel density estimates. This can be used to form pointwise confidence intervals and simultaneous confidence bands for density functions. Owen (1995) constructed empirical likelihood based confidence bands for the univariate empirical cumulative distribution function.

Chen and Hall (1993) considered empirical likelihood for quantiles, and discussed

Bartlett correctability and smoothing for this problem. Qin and Lawless (1994) studied the case where the number of estimating equations exceeds the number of parameters. This can be thought of as knowing a number of prior constraints that the parameters must satisfy. Qin (1994) considered problems in which parametric likelihoods on one sample may be combined with empirical likelihood on another.

Empirical likelihood method can also be extended to the problem such as regression, smoothing, biased sampling, censored and truncated data, time series, etc. Major developments have been summarized in the recent book by Owen (2001).

1.3 Calibration Methods of the Empirical Likelihood Ratio

1.3.1 The Chi-square Calibration and the Undercoverage Problem

Since the exact distribution of the empirical log-likelihood ratio is usually unknown, to construct an empirical likelihood confidence region, we need to approximate quantiles of this unknown distribution. This is typically handled by using the Chi-square calibration which refers to the method of approximating quantiles of the exact finite sample distribution of the empirical log-likelihood ratio with that of its limiting Chi-

square distribution, much like what we do in parametric likelihood inferences.

However, the Chi-square calibration tends to suffer from an undercoverage problem in that the actual coverage level of Chi-square calibrated empirical likelihood confidence region tends to be lower than the nominal level. Numerical evidence of this problem can be seen through many examples in the literature. See, for example, Owen (1988), Hall and La Scala (1990) and Qin and Lawless (1994).

By definition, the finite sample distribution of the empirical log-likelihood ratio is in general a mixture distribution with a continuous component supported on $[0, +\infty)$ and an atom at $+\infty$. The later is the probability that the convex hall of the sample $\mathcal{H}$ does not contain the true mean. As the sample size increases, the atom will vanish and the continuous component will approach the Chi-square distribution. However, when the sample size is not large, the atom can be substantial. That the Chi-square distribution cannot accommodate an atom at the tail can have serious impact on the accuracy of the Chi-square approximation. Because of the atom, the upper percentiles of the finite sample distribution tend to be larger than the corresponding percentiles of the Chi-square distribution and consequently, in practice, the Chi-square calibrated confidence regions usually suffer from an undercoverage problem.

As an example from Tsao (2004a), Figure 1.3 shows the shape of the estimated cumulative distribution of $-2\log(\mathcal{R}(\mu_0))$ based on random samples of size $n = 20$ and 30 from a $MVN(\mu_0, \Sigma)$, where $\mu_0 \in R^{10}$ and Σ is the 10×10 positive definite

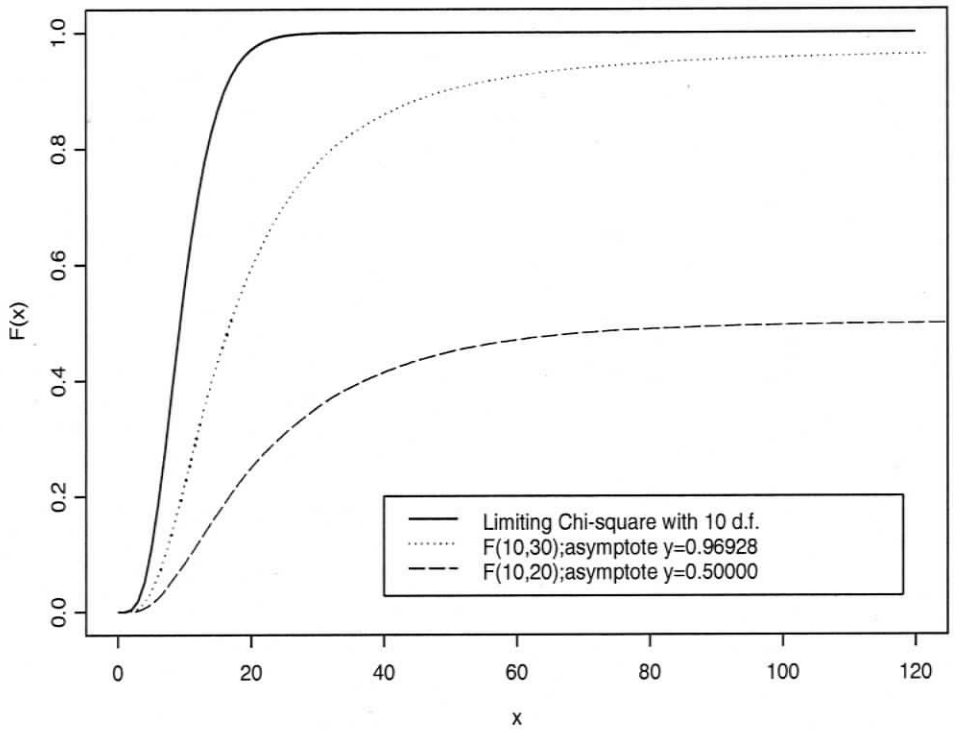

Figure 1.3: The undercoverage problem

matrix. $\hat{F}_{10,20}(x)$ is the estimated cumulative distribution function for the empirical log-likelihood ratio $-2\log(R(\mu_0))$ for the case where $n = 20$ and $\hat{F}_{10,30}(x)$ is that for the case where $n = 30$. The cumulative distribution function $F_{10}(x)$ of the corresponding limiting distribution χ^2_{10} is also shown. Here, $\hat{F}_{10,20}(x)$ and $\hat{F}_{10,30}(x)$ are estimated by the quantiles of large samples of size $m = 500,000$ from the corresponding distributions, respectively. The asymptotes of the $\hat{F}_{10,20}(x)$ and $\hat{F}_{10,30}(x)$ are $y = 0.5$ and $y = 0.96928$, respectively and that of $F_{10}(x)$ is $y = 1$. The distance between each asymptote and $y = 1$ is the value of the atom. Not surprisingly, $\hat{F}_{10,30}(x)$ is closer to the limiting $F_{10}(x)$ than $\hat{F}_{10,20}(x)$ and they both position substantially lower than $F_{10}(x)$. It suggests, for example, using the 0.9th quantile of $F_{10}(x)$ will result in a coverage level significantly less than 90%.

1.3.2 Other Calibration Methods

There are other methods of calibration based on Fisher's F distribution, the Bartlett correction and the bootstrap. In the present section, we describe these methods and discuss their advantages and disadvantages.

Details of the proof of the empirical likelihood theorem suggest that the critical value $\chi^2_{k,1-\alpha}$ should perhaps be replaced by,

$$\frac{(n-1)k}{(n-k)} F_{k,n-k,1-\alpha}, \tag{1.11}$$

which represents the critical values of the Fisher's F distribution.

To see this, through Taylor's expansion,

$$-2\log\mathcal{R}(\mu_0)\doteq n(\bar{X}-\mu_0)'\hat{V}^{-1}(\bar{X}-\mu_0)=T^2+O_p\Big(\frac{1}{n}\Big), \tag{1.12}$$

where

$$\hat{V}=1/n\sum_{i=1}^{n}(X_i-\mu_0)(X_i-\mu_0)'.$$

This shows the leading term converges to Hotelling's T^2 distribution. Based on this asymptotic equivalence, it would seem to be appropriate to calibrate empirical log-likelihood ratios with critical values of Hotelling's T^2. Since

$$T^2_{k,n}(1-\alpha)=\frac{(n-1)k}{(n-k)}F_{k,n-k}(1-\alpha),$$

the quantiles of T^2 distribution can be easily computed using that of the F distribution and hence (1.11). As n goes to $+\infty$ the Chi-square and Fisher's F calibrations become equivalent and the F-based calibration usually gives better coverage levels in simulations.

Another calibration method employs a Bartlett correction, which replaces $\chi^2_{k,1-\alpha}$ by

$$\Big(1+\frac{a}{n}\Big)^{-1}\chi^2_{k,1-\alpha},$$

where the appropriate value for a dependents on higher moments of the underlying distribution. Bartlett correction can reduce the coverage error to $O(n^{-2})$ from the typical level of order $O(n^{-1})$ of the Chi-square or F calibration. The parameter a

can be replaced by its sample version calculated through sample moments, and the asymptotic rate $O(n^{-2})$ still holds.

Bootstrap is another effective calibration method. For example, we sample n observations from the empirical distribution F_n for $B = 1000$ times and record the n observations as $X_1^{*b}, \ldots, X_n^{*b}$, for $b = 1, \ldots, B$. Compute $C^{*b} = -2\log(\mathcal{R}^{*b}(\bar{X}))$ where

$$\mathcal{R}^{*b}(\bar{X}) = \max\{\prod_{i=1}^{n} nw_i \mid \sum_{i=1}^{n} w_i(X_i^{*b} - \bar{X}) = 0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1\}.$$

The $\mathcal{R}^{*b}(\bar{X})$ is the empirical likelihood ratio calculated using the b^{th} bootstrap sample at the sample mean. Define the order statistics of C: $C^{(1)} \leq C^{(2)} \leq \ldots \leq C^{(B)}$. Then the bootstrap-calibrated 95% confidence region is given by $\{\theta \mid -2\log\mathcal{R}(\theta) \leq C^{(950)}\}$. See Owen (2001) for further discussions on these alternative methods of calibration.

Simulation shows these alternatives are more accurate than the Chi-square calibration. However, these methods still have their own drawbacks. For the method based on Fisher's F, it is easy to implement but the improvement is limited. For Bartlett correction, it can be easily applied to cases like the mean, but it can be too complicated for other parameters because the correction factor has to be derived for each individual case. The bootstrap provides the most accurate numerical results compared with the other two since it takes into consideration the atom of the finite sample distribution, but the coverage properties of the bootstrap calibration are not

well understood and the process is very time consuming.

1.3.3 Overview of the Thesis

A distribution which can better approximate the finite sample distribution of the empirical log-likelihood ratio may need to have a similar mixture structure with an atom at the infinity. There are few known mixture distributions of this kind but we can define some new ones by combining a proper point mass at the infinity with distributions such as the Chi-square and F distribution. We also have identified one with the mixture structure naturally. It is the finite sample distribution of the empirical log-likelihood ratio for the multivariate normal mean, which is called E distribution by Tsao (2004a). The rest of this paper is organized as follows: in Chapter 2, the mixture distributions based on a χ^2 and an F distribution will be examined. Numerical results will be provided to show the improvement of the calibration methods associated with these new distributions. In Chapter 3, we will revisit the E distribution and E calibration. Some new convergence properties of E random variables will be discussed. We will then compare several calibration methods numerically in this chapter. In Chapter 4, we will discuss methods of estimating the quantiles of the E distribution.

Chapter 2

New Calibration Methods Based on Mixture Distributions

There is a need for more studies on new calibration methods since the current methods are inaccurate (e.g. the Chi-square calibration) or inefficient (e.g. the bootstrap method). In this chapter, we propose two methods which will improve the accuracy of the Chi-square calibration yet still retain its computational simplicity. We have noted that the inaccuracy of the Chi-square calibration is mainly due to its inability to accommodate an atom at the infinity. Based on the fact that the distribution of the empirical log-likelihood ratio is a mixture distribution with a probability mass at the positive infinity, we define two new mixture distributions, E_C and E_F, by moving a certain amount of probability mass from $[0, +\infty)$ of the Chi-square distribution

or the Fisher's F distribution to $+\infty$. The resulting mixture distribution should provide better approximations to the finite sample distribution of the empirical log-likelihood ratio. Details such as computing the critical values of the new distributions and the properties of the new distributions will be presented. Simulation results are provided to show that calibration methods based on these mixture distributions are more accurate than the Chi-square calibration, especially for small sample situations.

In Section 2.1, we discuss atoms of the empirical log-likelihood ratio. In Section 2.2, we first define new distributions, E_C and E_F, by incorporating an atom of a proper size into the Chi-square and Fisher's F distributions, and then study properties of these new distributions. In Section 2.3, calibration methods based on these two new distributions will be developed and their accuracy will be examined by both theoretical considerations and several simulation examples.

2.1 The Atom of the Distribution of the Empirical Log-likelihood Ratio

We have commented in Section 1.1.4 that a trivial case exists when computing $l(\mu)$. Recall equations (1.5) and (1.7). If $\mu \notin \mathcal{H}$, then by convention, we define $\mathcal{R}(\mu) = 0$ or $l(\mu) = +\infty$. Moreover, if μ is on the boundary of $\mathcal{H}$, some w_i will be 0 and thus $\mathcal{R}(\mu) = 0, l(\mu) = +\infty$ as well. Since there is a positive probability that the true mean,

μ_0, will not be in $\mathcal{H}$, the finite sample distribution of the empirical log-likelihood ratio at the true mean, $l(\mu_0)$, will have an atom at the positive infinity.

Now we consider the empirical log-likelihood ratio defined for a general parameter θ_0 through estimating equations; see equation (1.4). Consider the convex hull of $m(Y_i, \theta_0)$, denoted by $\mathcal{H}(m(Y_1, \theta_0), m(Y_2, \theta_0), ..., m(Y_n, \theta_0))$. Just like the case for a mean μ_0, $l(\theta_0) = -2\log(R(\theta_0))$ will be finite if and only if $\underline{0}$, which is the origin in $\mathcal{R}^k$, is in the interior of the convex hull $\mathcal{H}(m(Y_1, \theta_0), m(Y_2, \theta_0), ..., m(Y_n, \theta_0))$. In other cases where $\underline{0}$ is not in the convex hull, $l(\theta_0)$ is equal to infinity, and this leads an atom at the positive infinity for the distribution of $l(\theta_0)$ as well.

It follows that no matter which distribution the data comes from or which parameter is of interest, the empirical log-likelihood ratio has, in general, a mixture distribution with a continuous part on $[0, +\infty)$ and a probability mass at $+\infty$. Figure 2.1 illustrates the mixture distribution with a continuous density function which has a Chi-square-like shape on $[0, +\infty)$ and an atom at infinity.

Although it is in general difficult to compute the size of the atom of the empirical log-likelihood ratio, for important cases, such as the empirical log-likelihood ratio for the mean of a multivariate normal distribution, a formula for the atom is available. It can be shown that the value of this atom equals the probability that a random sample of size n from the uniform distribution on the unit sphere all lie on some hemisphere. Denote this probability by $a(k, n)$. Wendel (1962) proved the following formula, for

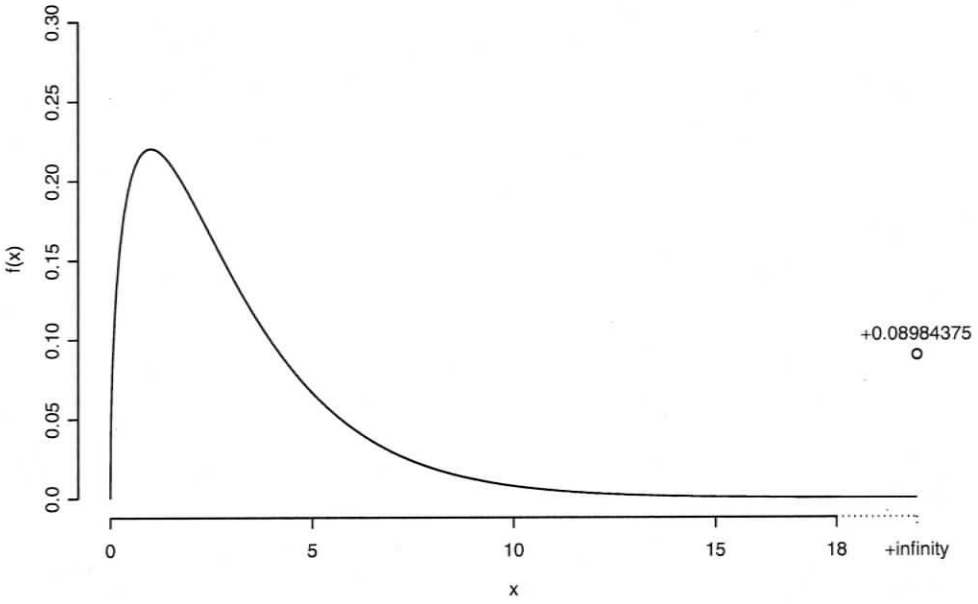

Figure 2.1: Illustration plot of the density function and the atom of the empirical log-likelihood ratio of a standard multivariate normal mean. Dimension $k = 3$ and sample size $n = 10$.

any $n > k$,

$$a(k,n) = \left\{ \binom{n-1}{0} + \binom{n-1}{1} + \cdots + \binom{n-1}{k-1} \right\} 2^{-(n-1)}. \tag{2.1}$$

Formula (2.1) holds for all cases where the underlying distribution are symmetric. Tsao (2004b) also suggested that the atom of the empirical log-likelihood ratio for the multivariate normal mean, $a(k,n)$, is the smallest among all members of $\Lambda_{k,n}$, the collection of distributions of all possible $l(\theta_0)$ based on estimating equations (1.4).

2.2 The E_C and E_F Distributions

The mixture nature of the distribution of the empirical log-likelihood ratio motivates us to consider the following mixture distributions, E_C and E_F.

Definition 2.1 *Denote by χ^2_k a Chi-square random variable with k degrees of freedom. An E_C random variable with degrees of freedom k and n is:*

$$E_C(k,n) = \begin{cases} \chi^2_k & \textit{with probability } 1 - a(k,n), \\ +\infty & \textit{with probability } a(k,n). \end{cases} \tag{2.2}$$

The above construction is quite natural since the Chi-square distribution is the limiting distribution of the empirical log-likelihood ratio. Also, recall from Chapter 1 that Fisher's F distribution provides an alternative method of calibration and it is usually more accurate than the Chi-square calibration. Therefore, we could replace

χ^2_k in E_C with Fisher's F, or more precisely, $k(n-1)/(n-k)F_{k,n-k}$, according to (1.11). This leads a new mixture distribution, E_F.

Definition 2.2 *Denote by $F_{k,n}$ an F random variable with k and n degrees of freedom. An E_F random variable with degrees of freedom k and n is:*

$$E_F(k,n) = \begin{cases} \frac{k(n-1)}{n-k}F_{k,n-k} & \text{with probability } 1-a(k,n), \\ +\infty & \text{with probability } a(k,n). \end{cases} \tag{2.3}$$

Both E_C and E_F depend on the sample size n and the dimension of the random vector k. By convention, we call k the first degree of freedom and n the second degree of freedom. For convenience, the notation E_C and E_F are used to represent both the random variables and the corresponding distribution functions, and where and when they represent the random variables or the distribution functions should be clear from the context.

For any random vector Y and $\alpha \in (0,1)$, denote by $Y(\alpha)$ the α critical value ((1-α)th quantile) of Y. The following theorem provide an easy way for computing critical values of $E_C(k,n)$ and $E_F(k,n)$ through that of χ^2_k and $F_{k,n-k}$.

Theorem 2.1 *For any $\alpha \in (a(k,n),1)$,*

$$E_C(k,n,\alpha) = \chi^2_k\left(\frac{1-\alpha}{1-a(k,n)}\right), \tag{2.4}$$

$$E_F(k,n,\alpha) = \frac{k(n-1)}{n-k}F_{k,n-k}\left(\frac{1-\alpha}{1-a(k,n)}\right). \tag{2.5}$$

Proof of the Theorem 2.1:

Note that, by law of totally probability, at any $x \in [0, +\infty)$,

$$\begin{aligned} P[E_C \leq x] &= P[\chi_k^2 \leq x] \times P[E_C = \chi_k^2] + P[+\infty \leq x] \times P[E_C = +\infty], \\ &= P[\chi_k^2 \leq x] \times P[E_C = \chi_k^2], \\ &= (1 - a(k,n))P[\chi_k^2 \leq x]. \end{aligned}$$

To find the α critical value of E_C, $E_C(k,n,\alpha)$, for some $a(k,n) < \alpha < 1$, by the above equation, we have

$$1 - \alpha = P[E_C \leq E_C(k,n,\alpha)] = (1 - a(k,n))P[\chi_k^2 \leq E_C(k,n,\alpha)].$$

Thus

$$P[\chi_k^2 \leq E_C(k,n,\alpha)] = \frac{1-\alpha}{1-a(k,n)}.$$

This gives the relation between the critical values of E_C and that of the χ_k^2:

$$E_C(k,n,\alpha) = \chi_k^2 \left(\frac{1-\alpha}{1-a(k,n)} \right).$$

It can be derived similarly that the critical values of the E_F distribution relates to that of the Fisher's F distribution through:

$$E_F(k,n,\alpha) = \frac{k(n-1)}{n-k} F_{k,n-k} \left(\frac{1-\alpha}{1-a(k,n)} \right).$$

□

	Dimension $k=1$				Dimension $k=2$				Dimension $k=3$			
$n\backslash\alpha$	0.20	0.10	0.05	0.01	0.20	0.10	0.05	0.01	0.20	0.10	0.05	0.01
5	2.11	4.22	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$
10	1.65	2.73	3.91	7.02	3.38	5.00	6.94	$+\infty$	5.81	11.11	$+\infty$	$+\infty$
15	1.64	2.71	3.84	6.64	3.23	4.62	6.03	9.40	4.70	6.39	8.11	13.57
20	1.64	2.71	3.84	6.64	3.22	4.61	5.99	9.22	4.65	6.26	7.83	11.42
30	1.64	2.71	3.84	6.63	3.22	4.61	5.99	9.21	4.64	6.25	7.81	11.35
χ^2_k	1.64	2.71	3.84	6.63	3.22	4.61	5.99	9.21	4.64	6.25	7.81	11.35

Table 2.1: Some critical values of the E_C distributions

	Dimension $k=1$				Dimension $k=2$				Dimension $k=3$			
$n\backslash\alpha$	0.20	0.10	0.05	0.01	0.20	0.10	0.05	0.01	0.20	0.10	0.05	0.01
5	2.95	7.60	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$
10	1.90	3.32	5.06	10.87	4.53	7.30	11.28	$+\infty$	9.56	24.41	$+\infty$	$+\infty$
15	1.80	3.07	4.55	8.70	3.88	5.83	7.99	12.36	6.22	8.95	12.06	24.62
20	1.76	2.97	4.35	8.10	3.69	5.47	7.38	12.71	5.69	7.99	10.41	16.71
30	1.72	2.88	4.17	7.56	3.52	5.16	6.87	11.17	5.30	7.33	9.42	14.53
χ^2_k	1.64	2.71	3.84	6.63	3.22	4.61	5.99	9.21	4.64	6.25	7.81	11.35

Table 2.2: Some critical values of the E_F distributions

Equations (2.4) and (2.5) give the computation methods for the critical values of E_C and E_F. Following these formulae, we generate tables of critical values which is presented in Table 2.1 and Table 2.2. At fixed α and k, the critical values of E_C and E_F converge to that of the limiting χ^2_k as n approaches infinity.

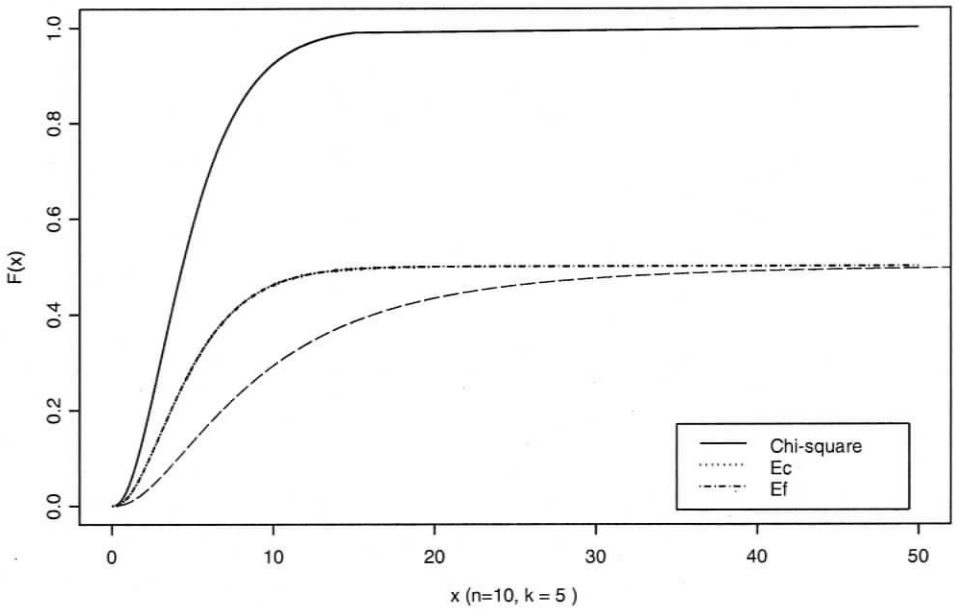

Figure 2.2: Plots of the cumulative distribution functions of Chi-square, E_C and E_F. Dimension $k = 5$ and sample size $n = 10$.

To see how E_C and E_F look like in relation to the Chi-square distribution, we plot the cumulative distribution functions of the three distributions for $k = 5$ and $n = 10$ in Figure 2.2. The shapes of the cumulative distribution functions of E_C and

E_F are quite similar to that of the Chi-square distribution but they are lower due to their atoms. The cumulative distribution function of the E_F is even lower than that of the E_C, especially when x is not large. When x approaches infinity, E_C and E_F begin to merge and coverage to the same value of $1 - a(5, 10) = 0.5$. If n goes to infinity, the atom will vanish and the continuous parts of E_C and E_F will move up and converge to the limiting Chi-square distribution. The observations that the cumulative distribution function of E_F is lower than that of E_C has been confirmed by numerical computations at different x values and n and k combinations. But a theoretical proof of this point has eluded us so far.

2.3 The E_C and E_F Calibrations

2.3.1 Theoretical Considerations

The E_C and E_F calibrations refer to the method of using critical values of E_C and E_F distributions to approximate that of the distribution of the empirical log-likelihood ratio, respectively.

As approximations of the distribution of empirical log-likelihood ratio, E_F and E_C may not be always accurate. Nevertheless, because of their mixture structure, they should be more accurate than the Chi-square or Fisher's F based approximations, and should lead to better empirical likelihood inferences. First of all, the asymptotic

behaviors of E_C and E_F agree with that of the distribution of empirical log-likelihood ratio, all approaching the Chi-square distribution as n approaches infinity. Consequently, these two new calibrations will provide consistant coverage probabilities just like the Chi-square calibration. Secondly, the critical values are enlarged by the correction made through incorporating the atom at the infinity, and the larger critical values improve the coverage level. For example in Table 2.1, for $\alpha = 0.2$ and $k = 3$, $\chi^2_{3,0.2} = 4.64$. However, $E_C(3, 10, 0.2) = 12.71$, which is almost three times as large. It is clear that an 80% empirical likelihood confidence interval calibrated with $E_C(3, 10, 0.2)$, $\{\theta : -2\log(R(\theta)) \leq 12.71\}$, will provide a higher coverage probability than that calibrated by $\chi^2_{3,0.2}$, $\{\theta : -2\log(R(\theta)) \leq 4.64\}$. Thirdly, the size of the atom incorporated is the true size for the distribution of $l(\mu_0)$ for the case of symmetric distributions. Therefore, in the case of symmetric distributions, the distribution function of E_C, E_F and $l(\mu_0)$ are expected to behave alike at least at the tail, so E_C and E_F calibrations are expected to give accurate coverage probability for high significant levels. Lastly, the atom of E_C and E_F are the smallest among all distributions of empirical log-likelihood ratio with the same dimension and the same sample size. This implies that the E_C and E_F calibrations provide the minimum adjustment to the Chi-square calibration as far as the atom is considered. So E_C and E_F calibrations tend to undercover when the underlying distributions are not symmetric. These points will be illustrated in the simulation results in the next section.

2.3.2 Simulations Results

Figure 2.3 shows the cumulative distribution functions of five distributions together: the Chi-square, Fisher's F, E_C, E_F and the estimated finite sample distribution of the empirical log-likelihood ratio of the mean for the multivariate normal distribution, the last of which will be referred to as the distribution of ELR for now. They all have the first degree of freedom $k = 5$ and the second degree of freedom $n = 10$. The left-hand plot shows the functions over the range of $(0, 50)$, whereas the right-hand shows the functions over the range of $(0, 20)$,which gives a closer look of the intersection in the plot. The size of atoms of three mixture distributions, E_C, E_F and ELR are all 0.5. According to formula (2.1), the size of the atom is always 0.5 when the sample size is twice as the dimension. The atom is indicated by the difference between the asymptote, $y = 0.5$, of the mixture distributions, and the asymptote, $y = 1$, of the Chi-square or Fisher's F distribution. When x becomes large, the Chi-square and Fisher's F converge to 1; E_C, E_F and ELR converge to 0.5. Due to the way that E_C and E_F are constructed, E_C inherits the shape of the Chi-square distribution and E_F inherits the shape of the Fisher's F distribution, but both are lower than their respective "parent" distribution. At the same probability level, E_C and E_F provide larger critical values which implies a better coverage for the E_C and E_F based calibration methods.

The closer the candidate distributions are to the true distribution of the empirical

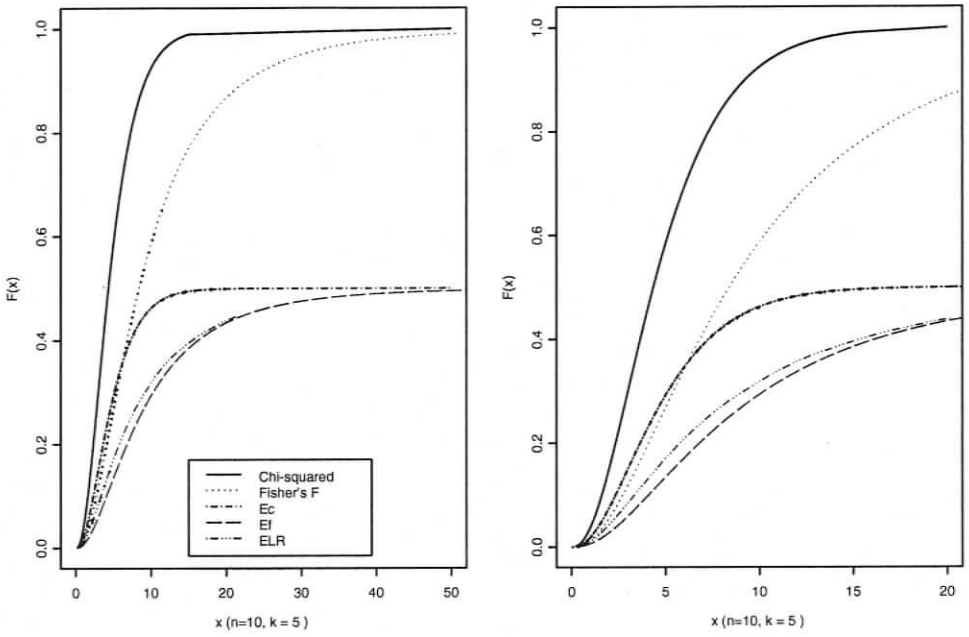

Figure 2.3: Plots of the cumulative distribution functions of Chi-square, Fisher's F, E_C, E_F and ELR. Dimension $k = 5$ and sample size $n = 10$.

log-likelihood ratio, the better the corresponding calibrations will be. Base on this observation, the best calibration is the one based on the E_F distribution. The Chi-square calibration is the worst since the cumulative distribution function of the Chi-square is the furthest from that of the distribution of ELR. E_C and Fisher's F are ranked in the middle. Two interesting points are showed in Figure 2.3. First, for small x, E_F is even lower than ELR which indicates the possibility of overcoverage by E_F calibration. But this does not contradict to our previous observations that the E_F calibration tends to undercover in general. This overcoverage does not usually occur in practice when the ratio of n/k is large. It is also interesting that E_C and Fisher's F intersect each other at some point between 5 and 10 as seen from the right plot in Figure 2.3. Before the intersection, E_C is closer to the ELR distribution. However, after the intersection, the Fisher's F is closer. So there is no general conclusion on the advantage of E_C over Fisher's F, which depends on the range of coverage level considered and has to be decided case by case. Fortunately, the intersection can be obtained by numerical computations since the distribution functions for E_C and Fisher's F are not difficult to evaluate. For practical applications, E_F based calibration should be better since E_F provides a more accurate approximation to the distribution of ELR at the tail, where is of the most interest in making inferences.

Figure 2.4 shows the cumulative distribution function of E_C and E_F with different combinations of degrees of freedom. Each single plot contains both E_C and E_F at the

same degrees of freedom. When n is increasing (see plots in the same column), they are getting closer to each other and both moving closer to the Chi-square distribution. As the ratio of n/k becomes smaller, like the case $k = 3$ and $n = 10$, the difference of the shape of cumulative distribution function between E_F and E_C are significant for x not large. As x goes to $+\infty$, they eventually merge to one another and approach the common horizontal asymptote. Not surprisingly, the curve of E_F is always lower than that of E_C, thus the E_F based calibration provides a consistently larger coverage probability than the E_C based calibration.

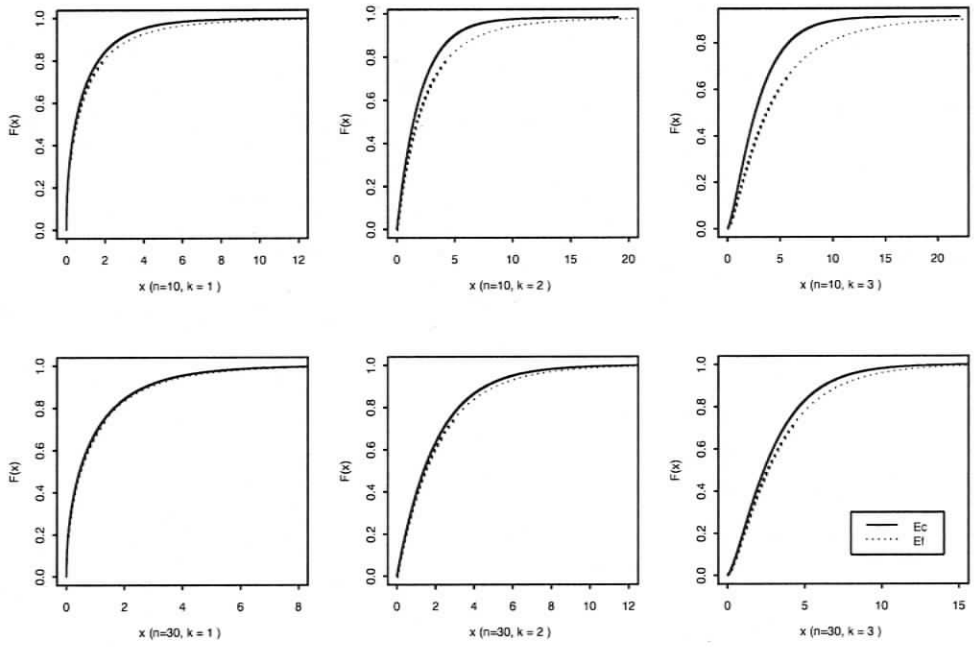

Figure 2.4: Plots of the cumulative distribution functions of E_C and E_F

Figure 2.5, Figure 2.6 and Figure 2.7 report the simulated coverage probability of calibration methods based on the E_C and E_F distribution along with methods based on the Chi-square and Fisher's F distribution for different underlying distributions. The coverage probability is calculated as follows: First, generate a random sample from a underlying distribution of interest and evaluate the empirical log-likelihood ratio at the true mean by Owen's S-plus code. Then, repeat this process for $m = 5000$ times, record the number of times m' that the evaluated empirical log-likelihood ratio is smaller than the critical value from the candidate method. Then, m'/m is the simulated coverage probability which is reported in these plots. In each plot, we display both the coverage probability at selected sample size and the trendline formed by connecting neighboring points. The coverage probability is only observed for sample sizes $n = 10, 15, 20, 25$ and 30. The trendline gives an indication of coverage probabilities for sample sizes between these numbers. All coverage probabilities obtained by different methods for the same predefined nominal level $1 - \alpha$ and fixed dimension k are reported in the same plot. The three plots in each row correspond to nominal levels $0.8, 0.9, 0.95$, respectively, and the levels are indicated by the horizontal lines above all trendlines. Each column contains plots for different underlying distribution: $N(0, 1)$, χ^2_1 and T_5, respectively, from top to bottom. Each figure contains coverage results for samples with the same dimension. In Figure 2.7, when $k = 3$ and $n = 10$, the atom is greater than 0.05 and the calibration value is not available for E_C and

E_F, So the trend lines start at $n = 15$.

In each plot, all coverage probabilities approach the nominal level as n increases, including that of E_C and E_F calibrations. Thus the E_C and E_F based methods are equivalent to the Chi-square method for large sample sizes. In general, E_F and Fisher's F have better coverage probabilities than E_C and the Chi-square method since their trendlines are constantly higher. E_F appears to be the best method so far, because it shows the largest coverage probabilities for any combination of dimension and sample sizes. E_F and Fisher's F give almost the similar coverage levels for sample sizes larger than 15, but for small sample sizes like 10, E_F and E_C based methods have substantially better accuracy. As we expected, all methods perform the best for samples from Normal distribution, the second best for samples from the symmetric T_5 distribution and perform the worst for samples from Chi-square distribution.

There is a strange trend in Figure 2.7 where $k = 3$. The coverage probabilities of E_F and E_C for $n = 10$ are close to or even bigger than that for $n = 15$, which seems to be inconsistent with our intuition that the larger sample size will lead to better accuracy. This abnormality may be caused by the large size of the atom. For the Chi-square method, the critical value does not change with n. When n increases, the empirical log-likelihood ratio is more Chi-square like and the proportion of simulated empirical log-likelihood ratio value falling into

$$\{\theta : -2\log(R(\theta)) \leq \chi^2_{k,\alpha}\}$$

increased and it approaches $(1-\alpha)100\%$ when $n \to +\infty$. But for intervals based on E_C and E_F calibration,

$$\{\theta : -2\log(R(\theta)) \leq C^*_{n,k}(\alpha)\} \tag{2.6}$$

where $C^*_{n,k}(\alpha)$ is the appropriate critical values of E_C or E_F distribution. $C^*_{n,k}(\alpha)$ changes with n. From $n = 10$ to $n = 15$ in Figure 2.7, the drop in C^* value is too large resulting in a smaller portion of the simulated empirical log-likelihood ratio values falling into confidence region (2.6). For example, the $E_C(3, 10, 0.9) = 31.28$ but the $E_C(3, 15, 0.9) = 12.84$, which is significantly smaller. This explains the abnormal behavior of the plot of E_C and E_F in Figure 2.7. But this abnormality is of little importance because it only happens at small n/k which is rarely the case.

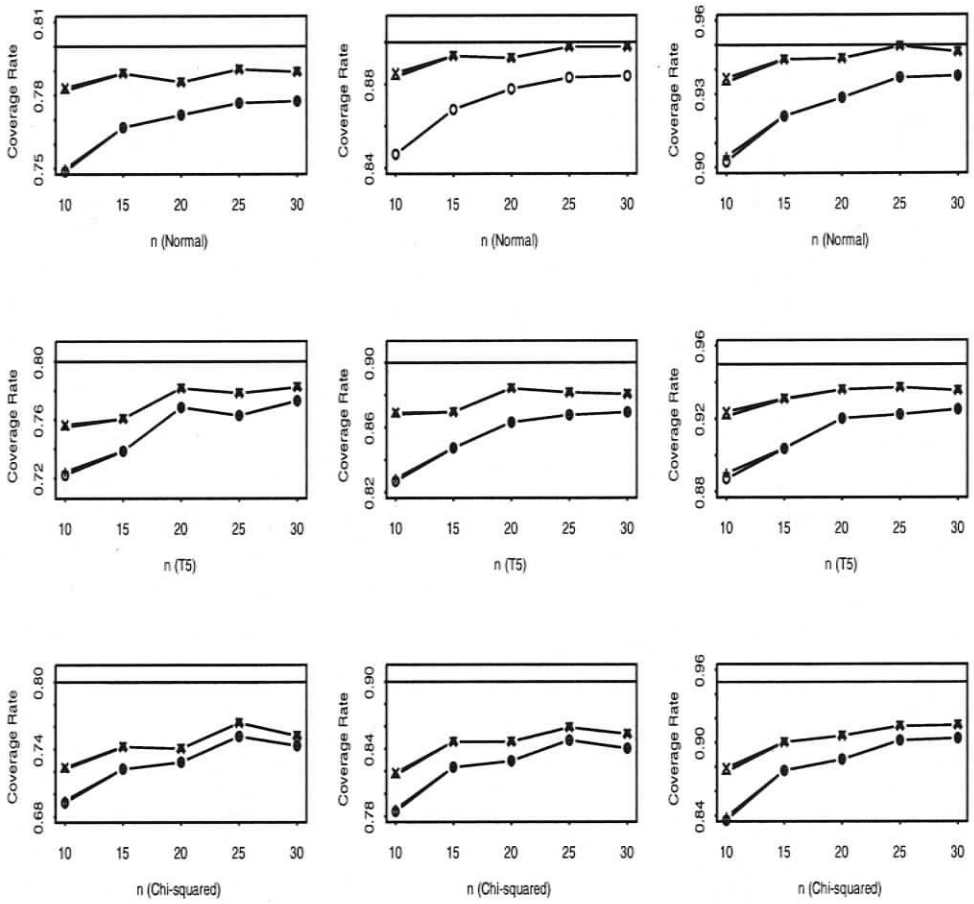

Figure 2.5: Coverage probabilities of different calibration methods for 1-dimensional samples. The underlying distributions considered from the top row to the bottom row are Normal, Chi-square, and T_5, respectively. O represents results by the Chi-square calibration method, Δ represents results by the Fisher's F calibration method, $+$ represents results by the E_C calibration method and $\times$ represents results by the E_F calibration method.

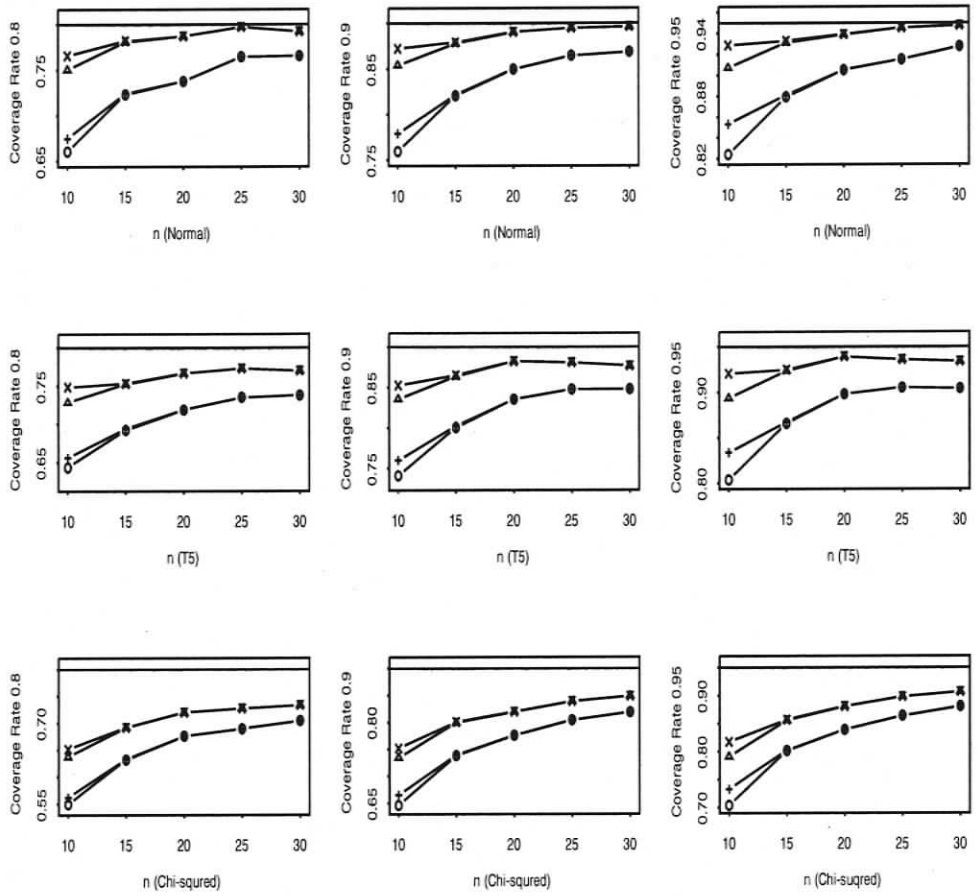

Figure 2.6: Coverage probabilities of different calibration methods for 2-dimensional samples. The underlying distributions considered from the top row to the bottom row are Normal, Chi-square, and T_5, respectively. O represents results by the Chi-square calibration method, Δ represents results by the Fisher's F calibration method, $+$ represents results by the E_C calibration method and $\times$ represents results by the E_F calibration method.

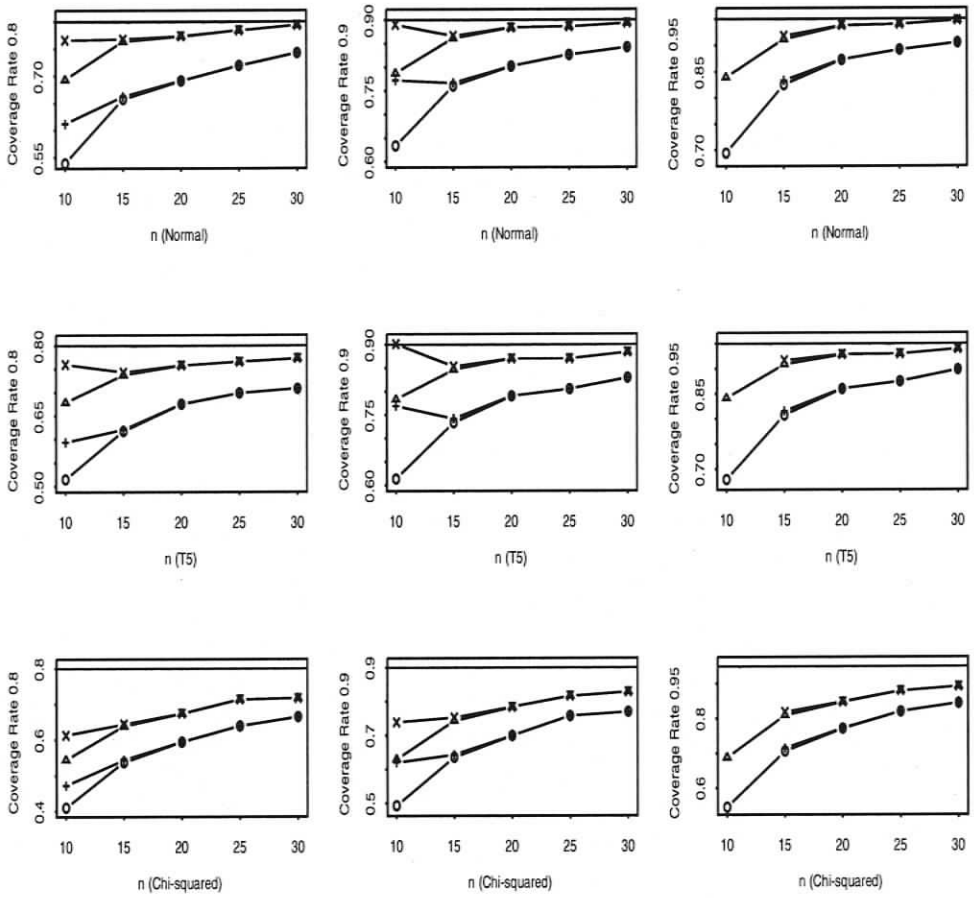

Figure 2.7: Coverage probabilities of different calibration methods for 3-dimensional samples. The underlying distributions considered from the top row to the bottom row are Normal, Chi-square, and T_5, respectively. O represents results by the Chi-square calibration method, Δ represents results by the Fisher's F calibration method, $+$ represents results by the E_C calibration method and $\times$ represents results by the E_F calibration method.

Chapter 3

The E Distribution and the E Calibration

The calibration methods discussed up to this point are based on either the limiting Chi-square distribution, such as the Chi-square and the E_C calibration, or based on the asymptotic considerations, such as the Bartlett correction, the Fisher's F or E_F calibrations and the Bootstrap calibration. In this chapter, we will study the exact finite sample distribution of the empirical log-likelihood ratio. In particular, we will study the E distribution (Tsao 2004a), which is the distribution of the empirical log-likelihood ratio for the multivariate normal mean. We will also study the E calibration, a method of calibration based on the E distribution. The E distribution depends on only the sample size n and the dimension k of the underlying multivariate

normal distribution. It is a mixture distribution, consisting of a continuous component on $[0, +\infty)$ and an atom at the positive infinity. The E calibration provides accurate empirical likelihood confidence regions for the normal mean. Moreover, it can be applied to calibrate general empirical log-likelihood ratios and to provide robust confidence regions against the undercoverage problem associated with the Chi-square calibration.

In Section 3.1, we discuss the definition and properties of the E distribution. Section 3.1.1 gives the definition of the E distribution and presents several known facts about this distribution. Section 3.1.2 introduces several properties related to the E distribution and shows that the E random variables converge to the limiting χ^2 random variable at a rate of $O_p(n^{-1})$. Using a result by Diciccio, Hall and Romano (1991), we also derive the rate of convergence of the E quantiles to the Chi-square quantiles. Details on the application of the E calibration, its advantages and disadvantages are discussed in Section 3.2.

3.1 The E Distribution Family

The general empirical log-likelihood ratio is a statistic of a random sample. Nevertheless, for any given sample, the log-likelihood ratio has to be computed numerically and cannot be expressed as an analytic function of the sample. Thus the exact finite sample distribution of the empirical log-likelihood ratio is not available, even when

the underlying distribution of the population is given. It is in general difficult to study the finite sample distribution of the empirical likelihood ratio. For the multivariate normal mean, however, the problem of the lack of the analytic expression for its cumulative distribution function can be bypassed by generating random samples from the finite sample distribution of the empirical likelihood ratio. With such large random samples, the cumulative distribution function or quantiles can be approximated accurately.

3.1.1 Definition of the E Distribution

The E random variable is the empirical log-likelihood ratio for the true mean of the standard multivariate normal distribution.

Definition 3.1 *The E random variable*

Let $X_1, X_2, \cdots, X_n$ be independent copies of a random vector X in R^k, where $n > k \geq 1$ and X has a standard multivariate normal distribution, $MVN(\underline{0}, I)$. Here I denotes the $k \times k$ identity matrix and $\underline{0}$ denotes the origin in $\mathcal{R}^k$. Then

$$E = l(\underline{0}) = 2\sum_{i=1}^{n} \log\{1 + \lambda^T X_i\}, \tag{3.1}$$

where λ is determined by the non-linear system:

$$\sum_{i=1}^{n} (1 + \lambda^T X_i)^{-1} X_i = \underline{0}. \tag{3.2}$$

The E random variable is univariate, positive and continuous on $[0, +\infty)$. We will refer to k and n as the first and the second degree of freedom, respectively. Also, the condition $n > k$ is necessary. Otherwise the sample $X_1, X_2, \cdots, X_n$ does not form a convex hull in R^k and the empirical likelihood is not defined.

As a special case of the general empirical log-likelihood ratio random variables, the E random variable has the atom $a(k, n)$ at the positive infinity. Because of this, none of the moments of E exists. Equation (2.1) gives an analytic expression for calculating this atom. Based on this expression, we can show that $a(k, n)$ is a strictly monotonically decreasing function of n and a strictly monotonically increasing function of k.

For convenience, we denote the E distribution with first degree of freedom k and second degree of freedom n as $E_{k,n}$. We may also use $E_{k,n}$ to represent the E random variable when we wish to emphasis the degrees of freedom. We also use $F_{k,n}(x)$ to denote the cumulative distribution of the $E_{k,n}$ random variable. That is, for any $x \in (0, +\infty)$, $F_{k,n} = P(E \leq x)$. In view of the atom of $E_{k,n}$ at the positive infinity, we have,

$$\lim_{x \to +\infty} F_{k,n}(x) + a(k, n) = 1, \quad \text{or} \quad \lim_{x \to +\infty} F_{k,n}(x) = 1 - a(k, n).$$

The line $y = 1 - a(k, n)$ is a horizontal asymptote for $F_{k,n}(x)$, since $F_{k,n}(x)$ is a strictly increasing function.

The pth quantile of $E_{k,n}$, $\xi(p,k,n)$, is given by,

$$\xi(p,k,n) = \begin{cases} \inf\{x : p \leq F_{k,n}(x)\} & \text{for any } p \in (0, 1-a(k,n)), \\ +\infty & \text{for } p \geq 1-a(k,n). \end{cases}$$

For any $\alpha \in [0,1]$, $e(\alpha,k,n)$ is used to represent the $100\alpha\%$ critical value of the $E_{k,n}$. We have

$$e(\alpha,k,n) = \xi(1-\alpha,k,n).$$

The pth quantile $q(p,k,n)$ of a random sample of size m from the $E_{k,n}$ distribution may be used as the (sample) quantile estimate of the critical value $e(\alpha,k,n)$. Table 3.1 contains estimated critical values $\hat{e}(\alpha,k,n)$ for $k \leq 3$ at selected n values. The value 3.81 in the second row and second column is the estimated value $\hat{e}(0.1,1,10)$. This value is the $0.90th$ quantile $q(0.90,1,10)$ of a random sample of size 50000 from the $E_{1,10}$ distribution. The estimated value $\hat{e}(0.01,2,10)$ at the first row and eighth column is $+\infty$, indicating that the atom of the $E_{2,10}$ distribution is more than 1%. The last row contains the corresponding Chi-square critical values. We make two observations from this table: [1] The estimated critical values $\hat{e}(\alpha,k,n)$ in each column converge to the value of χ^2_k at the bottom. [2] The estimated values converge monotonically in that they are strictly decreasing as n increases.

	Dimension $k=1$				Dimension $k=2$				Dimension $k=3$			
$n\backslash\alpha$	0.20	0.10	0.05	0.01	0.20	0.10	0.05	0.01	0.20	0.10	0.05	0.01
5	3.47	9.06	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$
10	2.10	3.81	6.02	14.74	5.49	9.69	17.16	$+\infty$	12.74	36.14	$+\infty$	$+\infty$
15	1.91	3.26	4.81	9.59	4.25	6.53	9.42	19.43	7.26	11.22	16.57	46.97
20	1.84	3.10	4.47	8.31	3.84	5.74	7.88	14.37	6.15	8.91	12.11	22.38
25	1.78	2.95	4.25	7.78	3.69	5.44	7.34	12.48	5.64	7.98	10.53	18.03
30	1.75	2.95	4.20	7.52	3.56	5.24	6.97	11.53	5.41	7.55	9.74	15.42
χ^2_k	1.64	2.71	3.84	6.63	3.22	4.61	5.99	9.21	4.64	6.25	7.81	11.35

Table 3.1: Some estimated critical values $\hat{e}(\alpha, k, n)$ of the $E_{k,n}$ distributions.

3.1.2 Properties of the E Distribution

By definition, the E distribution is the exact finite sample distribution of $l(\underline{0})$, the empirical log-likelihood ratio for the mean, 0, of the standard multivariate normal distribution. For a general multivariate normal distribution, Tsao (2004a) proved that the empirical likelihood ratio for the mean μ_0, $l(\mu_0)$, is a pivotal quantity with the same distribution as $l(\underline{0})$.

Theorem 3.1 *(Tsao 2004a)*

If $X \sim MVN(\mu_0, \Sigma)$, where μ_0 is the mean and Σ is a non-singular variance-covariance matrix, then the distribution of $l(\mu_0)$ is $E_{k,n}$, regardless of the values of μ_0 and Σ.

Theorem 3.1 says that the true distribution of $l(\mu_0)$ for a general multivariate normal mean is in fact an E distribution as well. In other words, the distribution of $l(\underline{0})$ for the standard multivariate normal mean and the distribution of $l(\mu_0)$ for the general multivariate normal mean are identical.

At any fixed k, the $E_{k,n}$ random variable converges to χ^2_k in distribution as n goes to infinity. This is a direct consequence of the Empirical Likelihood Theorem (Theorem 1.2) applied to the case of the standard multivariate normal distribution. In fact, this rate of convergence can be proved to be $O_p(n^{-1})$ and we will show this proof in this section. Besides, in order to compute quantiles of the E distribution, we are also interested in finding the rate at which quantiles of the E distribution converge to that of the Chi-square distribution since a better understanding of the rate may lead to better models for computing the quantiles. Using a result from DiCiccio, Hall and Romano, we show that the rate of convergence of quantiles is $O(n^{-1})$.

Theorem 3.2 *Let $E_{k,n}$ be the E random variable with first degree of freedom k and second degree of freedom n, and χ^2_k be the Chi-square random variable with k degrees of freedom. Then,*

$$E_{k,n} = \chi^2_k + O_p\Big(\frac{1}{n}\Big). \tag{3.3}$$

In order to prove Theorem 3.2, we first present some lemmas. For a random vector X in R^k, we use $\|X\|$ to denote the L_2-norm of X.

Lemma 3.1 *Let $X_1, X_2, ..., X_n$ be independent copies of a random vector X in R^k where X is the standard multivariate normal random variable. Then,*

$$n^{-1} \sum_{i=1}^{n} \|X_i\|^3 = O(1). \tag{3.4}$$

Proof of Lemma 3.1.

Let $Y_i = \|X_i\|^3$, then Y_i are independently and identically distributed random variables. Since X_i are multivariate normal random vectors, the Y_i have absolute moments of any order. Denote by μ_Y the mean of Y_i.

By the Strong Law of Large Numbers,

$$\frac{1}{n} \sum_{i=1}^{n} Y_i \xrightarrow{a.s.} \mu_Y.$$

So,

$$\frac{1}{n} \sum_{i=1}^{n} \|X_i\|^3 = \frac{1}{n} \sum_{i=1}^{n} Y_i = O(1).$$

□

Lemma 3.2 *Let $X_1, X_2, ..., X_n$ be independent copies of a random vector X in R^k where X is the standard multivariate normal random variable. Let $Z_n = max_{1 \le i \le n} \|X_i\|$. Then $Z_n = o(n^{1/2})$*

Proof of Lemma 3.2.

Since X_i are multivariate normal vectors, $E(\|X_i\|^2) < \infty$. Lemma 3.2 follows from the Lemma 11.2 in Owen (2001)(Page 218, Chapter 11).

□

Definition 3.2 *Let $X_1, X_2, \ldots, X_n$ be independent copies of X in R^k with common distribution $N(\mu_0, \Sigma)$. Define random variable $L_{k,n}$ as:*

$$L_{k,n} = n(\bar{X} - \mu_0)S^{-1}(\bar{X} - \mu_0),$$

where $\bar{X} = 1/n \sum_{i=1}^{n} X_i$ and $S = 1/n \sum_{i=1}^{n} (X_i - \mu_0)(X_i - \mu_0)'$.

Lemma 3.3 *Let $X_1, X_2, \ldots, X_n$ be independent copies of X in R^k with common distribution $N(\mu_0, \Sigma_k)$. Then the random variable, $L_{k,n}$ converges to χ^2_k with rate $O_p(n^{-1})$, i.e.*

$$L_{k,n} = \chi^2_k + O_p\Big(\frac{1}{n}\Big). \tag{3.5}$$

Proof of Lemma 3.3.

Recall S defined in Definition 3.2:

$$S = \frac{1}{n} \sum_{i=1}^{n} (X_i - \mu_0)'(X_i - \mu_0).$$

Let Σ be the sample variance covariance matrix:

$$\Sigma = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})'(X_i - \bar{X}).$$

We first prove

$$\Sigma S^{-1} = 1 + O_p\Big(\frac{1}{n}\Big). \tag{3.6}$$

Since

$$\begin{aligned} S &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X} + \bar{X} - \mu_0)'(X_i - \bar{X} + \bar{X} - \mu_0) \\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})'(X_i - \bar{X}) + \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})'(\bar{X} - \mu_0) \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}(\bar{X} - \mu_0)'(X_i - \bar{X}) + \frac{1}{n}\sum_{i=1}^{n}(\bar{X} - \mu_0)'(\bar{X} - \mu_0) \qquad (3.7) \\ &= \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})'(X_i - \bar{X}) + \frac{1}{n}\sum_{i=1}^{n}(\bar{X} - \mu_0)'(\bar{X} - \mu_0). \qquad (3.8) \end{aligned}$$

From (3.7) to (3.8) is because the sum of those two cross product terms in (3.7) are in fact 0.

The Central Limit Theory implies:

$$\bar{X} - \mu_0 = O_p\Big(\frac{1}{\sqrt{n}}\Big).$$

So

$$\frac{1}{n}\sum_{i=1}^{n}(\bar{X} - \mu_0)'(\bar{X} - \mu_0) = O_p\Big(\frac{1}{\sqrt{n}}\Big)O_p\Big(\frac{1}{\sqrt{n}}\Big) = O_p\Big(\frac{1}{n}\Big),$$

and therefore

$$S = \Sigma + O_p\Big(\frac{1}{n}\Big),$$

or equivalently,

$$\Sigma S^{-1} = 1 + O_p\Big(\frac{1}{n}\Big).$$

Again by Definition 3.2

$$L_{k,n} = n(\bar{X} - \mu_0)S^{-1}(\bar{X} - \mu_0)$$

$$
\begin{aligned}
&= n(\bar{X}-\mu_0)\Sigma^{-1}\Sigma S^{-1}(\bar{X}-\mu_0) \\
&= n(\bar{X}-\mu_0)\Sigma^{-1}\Big(1+O_p\Big(\frac{1}{n}\Big)\Big)(\bar{X}-\mu_0) \\
&= n(\bar{X}-\mu_0)\Sigma^{-1}(\bar{X}-\mu_0)+O_p\Big(\frac{1}{n}\Big) \\
&= \chi^2_k + O_p\Big(\frac{1}{n}\Big) \qquad (3.9)
\end{aligned}
$$

□

With these lemmas, now we can prove the convergence rate stated in Theorem 3.2.

Proof of Theorem 3.2.

Let $Y_i = \lambda'(X_i-\mu_0)$, where λ is the solution to equation (1.7). Using Y_i, equation (1.7) becomes

$$
\begin{aligned}
0 &= \frac{1}{n}\sum_{i=1}^{n}\frac{X_i-\mu_0}{1+\lambda'(x_i-\mu_0)} = \frac{1}{n}\sum_{1}^{n}\frac{X_i-\mu_0}{1+Y_i} \\
&= \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_0)(1-Y_i+Y_i^2)/(1+Y_i) \\
&= \bar{X}-\mu_0-S\lambda+\frac{1}{n}\sum_{i=1}^{n}\frac{(X_i-\mu_0)Y_i^2}{1+Y_i}. \qquad (3.10)
\end{aligned}
$$

By (3.10) we find:

$$\lambda = S^{-1}(\bar{X}-\mu_0)+\beta,$$

where

$$\beta = S^{-1}\frac{1}{n}\sum_{i=1}^{n}\frac{(X_i-\mu_0)Y_i^2}{1+Y_i}.$$

Because Lemma 3.1,

$$\frac{1}{n}\sum_{i=1}^{n}\|X_i-\mu_0\|^3 = O_p(1).$$

Using standard arguments in Chapter 11 in Owen (2001), we can show that

$$\|\lambda\| = O_p(n^{-1/2}).$$

Thus, β has a norm bounded by

$$\begin{aligned}\beta &= \frac{1}{n}\sum_{i=1}^{n} S^{-1}\|X_i-\mu_0\|^3\|\lambda\|^2|1+Y_i|^{-1} \\ &= O_p(1)O_p(1)O_p(n^{-1})O_p(1) = O_p(n^{-1}) \qquad (3.11)\end{aligned}$$

By Lemma 3.2,

$$\max_{1\le i\le n}|Y_i| = \max_{1\le i\le n}|\lambda'(X_i-\mu_0)| = O_p(n^{-1/2})o_p(n^{1/2}) = o_p(1) \qquad (3.12)$$

Since Y_i is a small quantity, we may write:

$$\log(1+Y_i) = Y_i - \frac{1}{2}Y_i^2 + \eta_i,$$

where for some finite $B > 0$

$$Pr(|\eta_i| \le B|Y_i|^3, 1 \le i \le n) \to 1,$$

as $n \to \infty$.

By (1.5) and (1.6):

$$-2\log R(\mu_0) = -2\sum_{i=1}^{n}\log(nw_i)$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \log(1+Y_i) \\
&= 2\sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} Y_i^2 + 2\sum_{i=1}^{n} \eta_i \\
&= 2n\lambda(\bar{X}-\mu_0) - n\lambda' S\lambda + 2\sum_{i=1}^{n} \eta_i \\
&= n(\bar{X}-\mu_0)'S^{-1}(\bar{X}-\mu_0) - n\beta' S^{-1}\beta + 2\sum_{i=1}^{n} \eta_i.
\end{aligned}
$$

By Definition 3.2, the lead term $n(\bar{X}-\mu_0)'S^{-1}(\bar{X}-\mu_0) = L_{k,n}$. By equation (3.11):

$$n\beta' S^{-1}\beta = nO_p(n^{-1})O_p(1)O_p(n^{-1}) = O_p(n^{-1})$$

and

$$\left|\sum_{i=1}^{n} \eta_i\right| \leq B\,\|\lambda\|^3 \sum_{i=1}^{n} \|X_i-\mu_0\|^3 = O_p(n^{-3/2})O_p(n^{1/2}) = O_p(n^{-1}),$$

where, by the central limit theorem,

$$\sum_{i=1}^{n} \|X_i-\mu_0\|^3 = n(1/n\sum_{i=1}^{n} \|X_i-\mu_0\|^3) = nO_p(n^{-1/2}) = O_p(n^{1/2})$$

It follows that

$$E_{k,n} = L_{k,n} + O_p\Big(\frac{1}{n}\Big) = \chi_k^2 + O_p\Big(\frac{1}{n}\Big).$$

□

To study the convergence of quantiles, we need the following result:

Theorem 3.3 *(DiCiccio, Hall and Romano 1990)*

Let $F_{k,n}(x)$ be the cumulative distribution function of the $E_{k,n}$ distribution and let

$G_k(x)$ be the cumulative distribution function of the χ^2_k distribution. Then,

$$F_{k,n}(x) = G_k(x) + O\Big(\frac{1}{n}\Big) \tag{3.13}$$

Theorem 3.4 *The pth percentile $\xi(p,k,n)$ of $E_{k,n}$ converges to the pth percentile $\chi^2_{k,p}$ of χ^2_k with rate $O(1/n)$ as $n \to \infty$, i.e.*

$$\xi(p,k,n) = \chi^2_{k,p} + O\Big(\frac{1}{n}\Big)$$

Proof of Theorem 3.4.

By Theorem 3.3, we know

$$F_{k,n}(x) = G_k(x) + O\Big(\frac{1}{n}\Big).$$

For simplicity, let $x_0 = \chi^2_{k,p}$, the pth percentile of the Chi-square distribution and $x_n = \xi(p,k,n)$, the pth percentile of the $E_{k,n}$ distribution, then

$$\begin{aligned} G_k(x_0) &= p, \\ E_{k,n}(x_n) &= p, \quad \text{and} \\ G_{k,n}(x_n) + O(1/n) &= p. \end{aligned} \tag{3.14}$$

The Taylor's expansion of right-hand side of the equation (3.14) at x_0 is :

$$\begin{aligned} G_k(x_0) + G'_k(x_0)(x_n - x_0) + G''_k(x_n - x_0)^2/2 + \ldots + O(1/n) &= p, \\ p + G'_k(x_0)(x_n - x_0) + o(x_n - x_0) + O(1/n) &= p, \quad \text{and} \\ G'_k(x_0)(x_n - x_0) + o(x_n - x_0) + O(1/n) &= 0. \end{aligned}$$

For the Chi-square distribution, the density function $G'_k(x_0)$ will not be 0, so

$$\begin{aligned}(x_n - x_0) + o(x_n - x_0) &= O(1/n), \\ (x_n - x_0) &= O(1/n).\end{aligned}$$

Thus

$$x_n = x_0 + O(1/n).$$

□

Another important property about the E distribution is the monotonically decreasing property of the quantiles with respect to n. Although we cannot prove it theoretically, it is supported by both simulation evidences and similar behavior of other distributions. Tsao (2004c) formulated this property with the following conjecture:

Conjecture 3.1 *Monotone Property of Quantiles*

For any $x \in (0, \infty)$ and $\alpha > a(k, n+1)$, it is assumed that the cumulative distribution $F_{k,n}(x)$ and the critical value $e_{\alpha,k,n}$ of the E distribution satisfy that:

$$F_{k,n}(x) < F_{k,n+1}(x), \quad \textit{or equivalently} \quad e_{\alpha,k,n} > e_{\alpha,k,n+1}.$$

3.2 The E Calibration

3.2.1 Theoretical Considerations

The E calibration refers to approximating quantiles of the finite sample distribution of the empirical log-likelihood ratio with that of the E distribution for the purpose of making empirical likelihood inferences. The E calibration can be used to construct empirical likelihood ratio confidence regions or conduct empirical likelihood based hypothesis testing. This method is conceptually different from the Chi-square calibration in that a real finite sample distribution of the empirical log-likelihood ratio is used for calibration. It alleviates the undercoverage problem of the Chi-square calibration.

By Theorem 3.1, when making empirical likelihood based inferences on multivariate normal means, the E calibration is exact because the distribution of the empirical log-likelihood ratio in this case is exactly the E distribution. So the E calibration can be regarded as the exact calibration method for the normal sample means.

For the empirical likelihood inference for the mean, μ, of a symmetric distribution, let $l(\mu)$ be the empirical log-likelihood ratio for μ. It is known that the size of the atom of the E distribution equals that of the atom of $l(\mu)$. Using the E distribution to calibrate the finite sample distribution reduces the calibration error in that the E distribution and the finite sample distribution are the same at the far tail. The only

difference occurs in the continuous component of the distributions and the cumulative distribution functions of the two distributions have the same asymptote (the same size of atom), so they are close to each other at the far right tail. Based on these observations, the E distribution is more attractive than the Chi-square distribution (which cannot accommodate an atom to the infinity) for calibrating $l(\mu)$.

The applications of the E calibration are not limited to the normal or symmetric distributions. Let θ be a parameter of interest for a random variable X. Let θ_0 be the true value and $m(X, \theta)$ be an estimating function such that $E(m(X, \theta_0)) = 0$, consider the empirical log-likelihood ratio:

$$\mathcal{R}(\theta_0) = \max\Big\{ \prod_{i=1}^{n} nw_i \mid \sum_{i=1}^{n} w_i m(X_i, \theta_0) = 0, w_i \geq 0, \sum_{i=1}^{n} w_i = 1 \Big\}. \tag{3.15}$$

Now let $l(\theta_0) = -2 \log R(\theta_0)$, and let $\Lambda_{k,n}$ denotes the collection of distributions of all possible $l(\theta_0)$. Here, θ_0 does not have to be a mean and the underlying distribution of X does not have to be normal. It is clear that $E_{k,n} \in \Lambda_{k,n}$ but $\chi^2_k \notin \Lambda_{k,n}$. As a member of the $\Lambda_{k,n}$ family, E behaves more like the distributions in $\Lambda_{k,n}$ than the Chi-square does. On the other hand, the $E_{k,n}$ distribution has the smallest atom among distributions in $\Lambda_{k,n}$. If we measure the distance between a distribution in $\Lambda_{k,n}$ and the χ^2_k distribution by the size of atom of the former, the $E_{k,n}$ distribution will be the closest to the χ^2_k distribution. Other distributions are further away in that their atoms are larger. So the E calibration, which uses the E distribution to approximate all other distributions in $\Lambda_{k,n}$, provides a "minimum" amount of correction to the

Chi-square calibration. Because of this, the E calibration also tends to have an undercoverage problem but the extent of this problem is far less serious than that of the Chi-square calibration.

E distributions are the empirical likelihood equivalent of Hotelling's T^2 distributions for the parametric likelihood. They are both exact finite sample distributions when the underlying distributions are multivariate normal; the E distribution is the exact finite sample distribution of the empirical likelihood ratio whereas the Hotelling's T^2 distribution is the exact finite sample distribution of the parametric likelihood ratio. From this standpoint, the E calibration is very much an extension of T^2 calibration (parametric inference) in the context of the empirical likelihood inference.

3.2.2 Simulation Results

To compare the E calibration with other methods of calibration, from Figure 3.1 to Figure 3.4, we show the plot of cumulative distribution functions of E, E_C, E_F, Fisher's F and Chi-square for $n = 10$ and $k = 1, 3, 4, 5$. Recall that the undercoverage problem of the Chi-square calibration is caused by the under estimation of the quantiles of the finite sample distribution of the empirical log-likelihood ratio. To correct this problem, the estimated quantiles need to be larger. When $k = 1$, the advantage of E distribution is not obvious since all curves (cumulative distribution functions) are

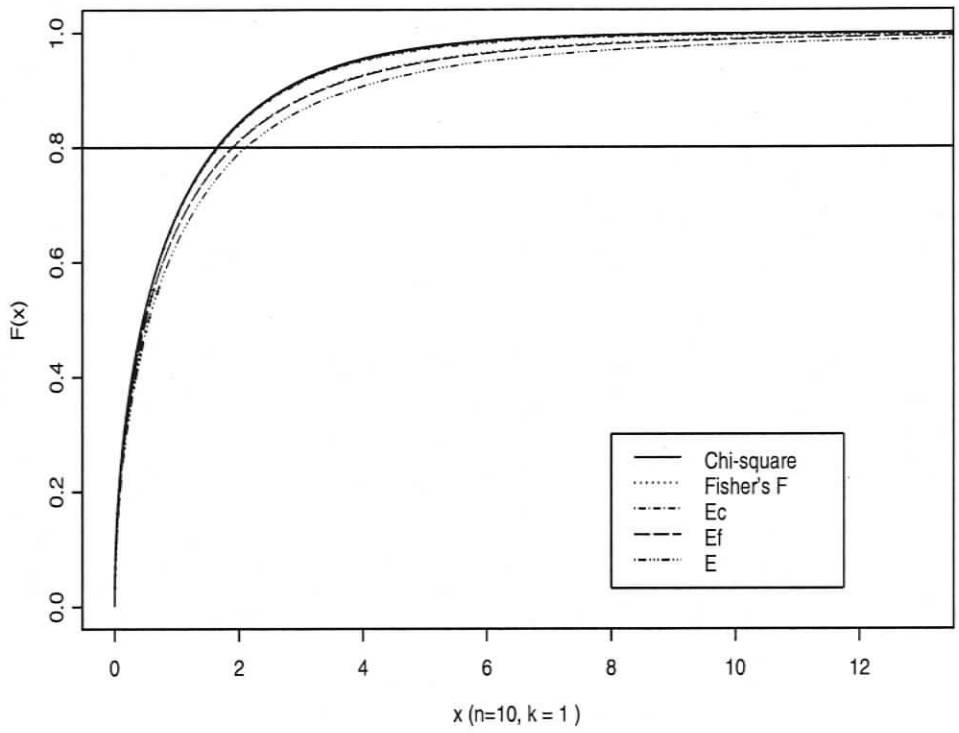

Figure 3.1: Plots of the cumulative distribution functions of E, Chi-square, Fisher's F, E_C and E_F. Dimension $k = 1$ and sample size $n = 10$.

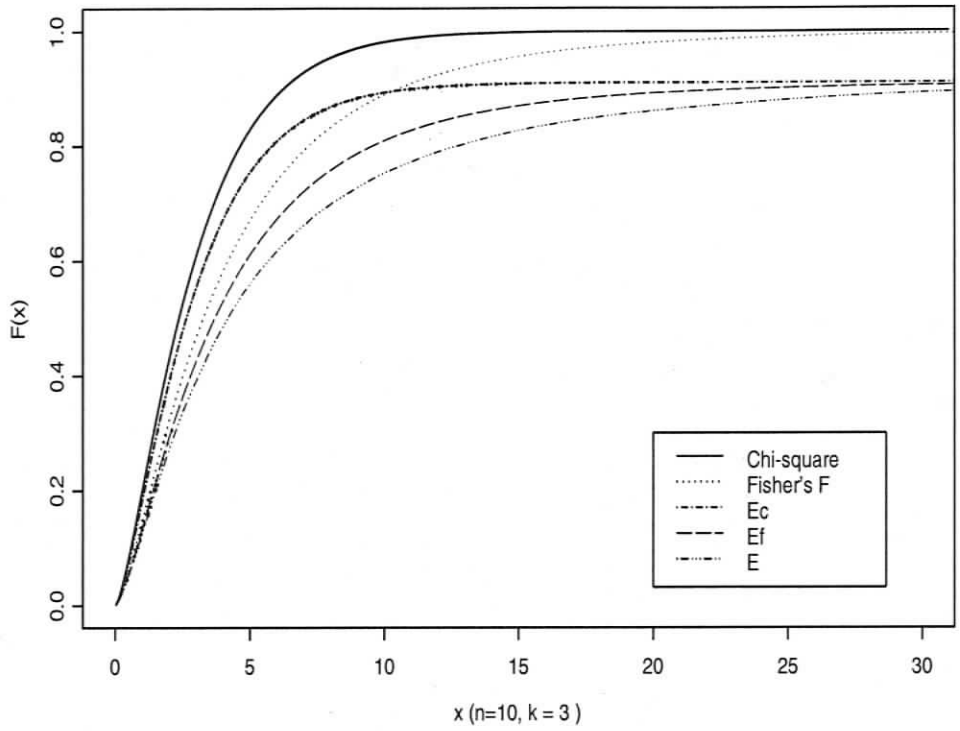

Figure 3.2: Plots of the cumulative distribution functions of E, Chi-square, Fisher's F, E_C and E_F. Dimension $k = 3$ and sample size $n = 10$.

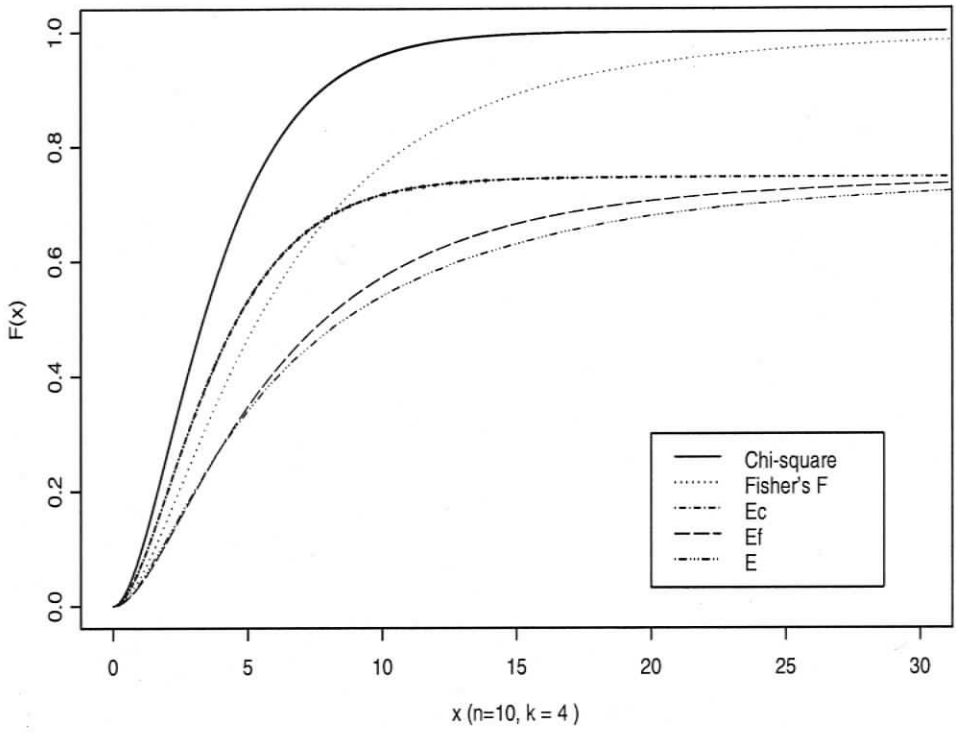

Figure 3.3: Plots of the cumulative distribution functions of E, Chi-square, Fisher's F, E_C and E_F. Dimension $k = 4$ and sample size $n = 10$.

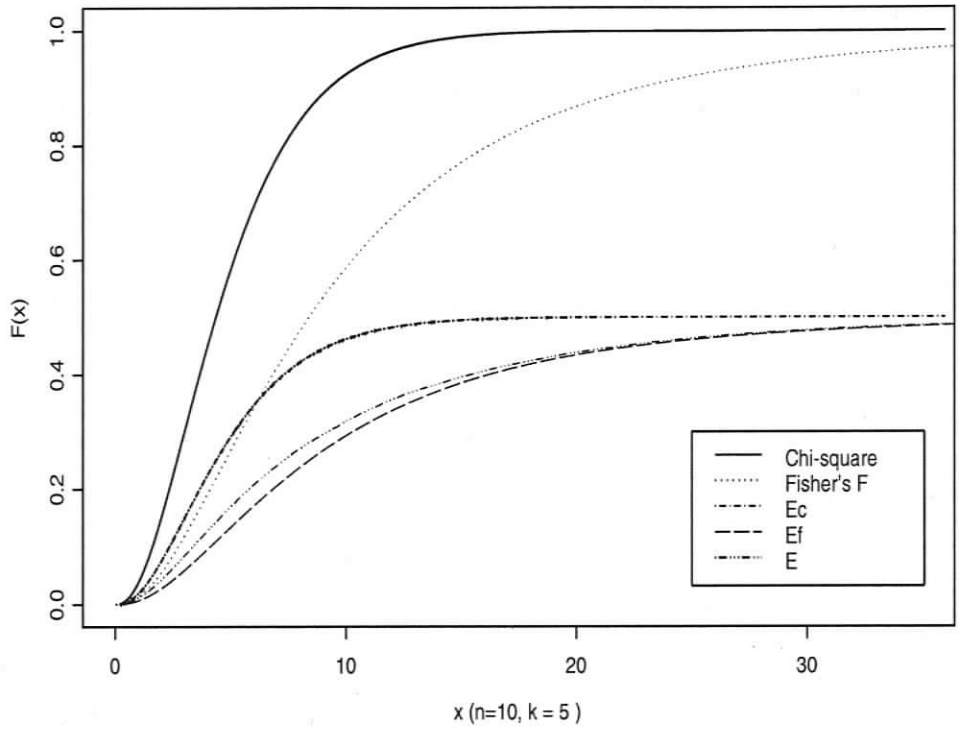

Figure 3.4: Plots of the cumulative distribution functions of E, Chi-square, Fisher's F, E_C and E_F. Dimension $k = 5$ and sample size $n = 10$.

very close. The quantile estimates based on the four methods will also be very close. For example, the horizontal line in the first plot in Figure 3.1 is $y = F(x) = 0.8$. The intersections between this line and the four cumulative distribution curves represent the 0.8th quantile estimates provided by these four methods, and they are not much different from each other. When k is large, for instance, $k = 3, 4, 5$, the difference is significant between E and the Chi-square because of the large atom. For $k = 1, 3, 4$, E is the lowest curve in each plot so that it provides the largest quantile estimate at any level. It reveals that the E calibration is not only better than the Chi-square calibration, but also better than E_C, E_F and Fisher's F calibrations. As k increases, the cumulative distribution of E_F distribution becomes lower. When $k = 5$, it moves under the cumulative distribution of E. So in this case, it is possible that E_F based calibration method may perform better than the E calibration.

Comparing the five calibration methods, the Chi-square calibration stands at one extreme in that it does not take into consideration the atom of the distribution of the empirical log-likelihood ratio. From the cumulative distribution plots, we see that it consistently gives smaller estimates for the quantiles of the distribution of the empirical log-likelihood ratio. At the other extreme we have the E_F and E calibration, which take into consideration the atom of the distribution of empirical log-likelihood ratio and give consistently larger estimates of the quantiles than other methods. The E_C and Fisher's F calibrations are methods between the two extremes,

and their corresponding cumulative distribution functions are located between that of the Chi-square and E or E_F distribution.

	Dimension $k=1$			Dimension $k=2$			Dimension $k=3$		
$n=10$	80%	90%	95%	80%	90%	95%	80%	90%	95%
χ^2	0.7488	0.8466	0.9022	0.6602	0.7594	0.8240	0.5384	0.6334	0.6940
E_F	0.7830	0.8854	0.9366	0.7656	0.8722	0.9288	0.7662	0.8900	NA
E	0.8052	0.9016	0.9494	0.8018	0.9026	0.9530	0.7988	0.8986	NA
$n=20$	80%	90%	95%	80%	90%	95%	80%	90%	95%
χ^2	0.7720	0.8778	0.9286	0.7376	0.8496	0.9052	0.6910	0.8022	0.8728
E_F	0.7854	0.8926	0.9446	0.7870	0.8902	0.9392	0.7734	0.8840	0.9382
E	0.7944	0.8968	0.9462	0.7972	0.9000	0.9472	0.7956	0.9020	0.9546
$n=30$	80%	90%	95%	80%	90%	95%	80%	90%	95%
χ^2	0.7776	0.8840	0.9374	0.7654	0.8686	0.9276	0.7428	0.8426	0.9054
E_F	0.7896	0.8978	0.9472	0.7920	0.8960	0.9476	0.7938	0.8932	0.9476
E	0.7930	0.9000	0.9478	0.7960	0.8972	0.9488	0.7960	0.8956	0.9506

Table 3.2: Simulated coverage probabilities of the Chi-square calibrated, E_F calibrated and E calibrated empirical likelihood confidence intervals for the mean of the $N(0,1)$ distributions.

To evaluate the performance of the E calibration in terms of coverage probabilities, we conducted a simulation study on confidence intervals based on the Chi-square calibration, the E calibration and the E_F calibration. We only consider this three

	Dimension $k=1$			Dimension $k=2$			Dimension $k=3$		
$n=10$	80%	90%	95%	80%	90%	95%	80%	90%	95%
χ^2	0.6924	0.7842	0.8362	0.5494	0.6458	0.7036	0.4094	0.4934	0.5448
E_F	0.7236	0.8186	0.8792	0.6518	0.7518	0.8172	0.6132	0.7392	NA
E	0.7424	0.8358	0.8944	0.6842	0.7860	0.8488	0.6424	0.7492	NA
$n=20$	80%	90%	95%	80%	90%	95%	80%	90%	95%
χ^2	0.7284	0.8292	0.8862	0.6760	0.7756	0.8390	0.5940	0.6990	0.7712
E_F	0.7406	0.8462	0.9056	0.7202	0.8196	0.8808	0.6740	0.7838	0.8482
E	0.7480	0.8508	0.9084	0.7294	0.8302	0.8894	0.7294	0.8302	0.8894
$n=30$	80%	90%	95%	80%	90%	95%	80%	90%	95%
χ^2	0.7430	0.8402	0.9036	0.7044	0.8188	0.8806	0.6642	0.7690	0.8446
E_F	0.7518	0.8530	0.9144	0.7334	0.8484	0.9066	0.7168	0.8278	0.8922
E	0.7548	0.8572	0.9146	0.7380	0.8490	0.9076	0.7380	0.8490	0.9076

Table 3.3: Simulated coverage probabilities of the Chi-square calibrated, E_F calibrated and E calibrated empirical likelihood ratio confidence intervals for the mean of the χ^2_1 distributions.

	Dimension $k=1$			Dimension $k=2$			Dimension $k=3$		
$n=10$	80%	90%	95%	80%	90%	95%	80%	90%	95%
χ^2	0.7220	0.8268	0.8870	0.6436	0.7408	0.8038	0.5148	0.6138	0.6792
E_F	0.7566	0.8692	0.9238	0.7484	0.8528	0.9208	0.7598	0.9010	NA
E	0.7788	0.8868	0.9406	0.7826	0.8892	0.9480	0.7948	0.9096	NA
$n=20$	80%	90%	95%	80%	90%	95%	80%	90%	95%
χ^2	0.7686	0.8632	0.9202	0.7188	0.8354	0.8982	0.6752	0.7900	0.8600
E_F	0.7818	0.8842	0.9360	0.7664	0.8824	0.9392	0.7576	0.8700	0.9286
E	0.7894	0.8894	0.9384	0.7798	0.8910	0.9472	0.7842	0.8944	0.9454
$n=30$	80%	90%	95%	80%	90%	95%	80%	90%	95%
χ^2	0.7730	0.8694	0.9252	0.7378	0.8482	0.9044	0.7082	0.8292	0.8982
E_F	0.7824	0.8806	0.9356	0.7698	0.8768	0.9340	0.7726	0.8836	0.9390
E	0.7868	0.8824	0.9358	0.7718	0.8792	0.9346	0.7762	0.8870	0.9436

Table 3.4: Simulated coverage probabilities of the Chi-square calibrated, E_F calibrated and E calibrated empirical likelihood ratio confidence intervals for the mean of the T_5 distributions.

because from the above discussion, we know that the coverage probabilities of the E_C calibration and the F calibration will be somewhat in between that of the Chi-square and the E or E_F calibration. E quantiles used in the calibration are obtained from the E Table in Tsao (2004c). Table 3.2, Table 3.3 and Table 3.4 show the coverage probability for the mean of the $N(0,1)$, χ^2_1 and T_5 distribution, respectively. Not surprisingly, for all sample size and dimension combinations we considered, the E calibration offers better coverage probabilities than the Chi-square and the E_F calibration. For example, in the normal case (Table 3.2) when $k = 1$ and $n = 30$, the simulated coverage probability based on the Chi-square method of the 90% confidence interval is 0.8840. The E_F method improves this probability to 0.8978, and the E calibration gives the exact probability of 0.9. The same phenomena happens at the 90% confidence interval of $k = 2$ and $n = 20$ for the normal case. Naturally, for the case of the χ^2_1 and the T_5 distribution, the coverage probabilities offered by the E calibration are not as accurate as those for the case of the normal distribution, but they are still much more satisfactory than the other two methods. For example, in the case of $k = 3$ and $n = 10$, for the 90% confidence interval for the mean of the χ^2_1 distribuion, the Chi-square method only yields a coverage probability of 0.4934, but E produces a coverage probability of 0.7492, which represents an increase of 0.25. Overall, the improvement brought by the E calibration (to the Chi-square calibration) ranges from 0.0104 to 0.2958. However, the improvement in the coverage probability

when compared to the E_F calibration is not substantial in general but can be as large as 0.0554 (the case of $k = 3$ and $n = 20$ for 80% confidence interval in Table 3.3).

3.2.3 Problems of the E Calibration

It is clear from Table 3.2, Table 3.3 and Table 3.4 that the E calibration also suffers from an undercoveage problem. There are three possible causes for this. [1] The underlying distribution is not symmetric such as χ^2_1 so the atom of the distribution of the empirical log-likelihood ratio will be larger than that of the E distribution. [2] Even when the underlying distribution is symmetric, the continuous part of $l(\mu_0)$ could be different from that of the E distribution. [3] The empirical log-likelihood ratio is not for the mean but for other parameters, such as the correlation coefficient. The first two cases are already seen from the above tables. To see the last one, we revisit an example first considered in Tsao (2004a).

Consider the distribution of $l(\theta_0)$ where θ_0 is the parameter vector of a bivariate normal random vector $(X, Y)^T$; $\theta_0 = (\mu_X, \mu_Y, \sigma^2_X, \sigma^2_Y, \sigma_{XY})^T$. Let $\theta_0 = (0, 0, 10, 17, 7)^T$ and formulate $R(\theta_0)$ through an estimating function which may be found on page 42 in Owen (2001). Table 3.5 gives the simulated critical values of the distribution of $l(\theta_0)$, the corresponding critical values of the E distributions and the corresponding critical values of the limiting Chi-square distributions. At $n = 20, \alpha = 0.25$, the simulated critical value of $l(\theta_0)$ is 23.78, which is still much larger than the quantile

n	Distribution	α=0.5	0.4	0.3	0.25	0.2	0.15	0.1	0.05	0.01
15	$l(\theta_0)$	16.21	3.38	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$
	$E_{5,15}$	8.32	10.69	14.31	17.45	22.33	31.48	62.31	$+\infty$	$+\infty$
20	$l(\theta_0)$	9.5	12.98	18.59	23.78	31.82	50.70	$+\infty$	$+\infty$	$+\infty$
	$E_{5,20}$	6.52	7.77	9.64	10.90	12.68	15.23	19.20	28.20	$+\infty$
30	$l(\theta_0)$	6.30	7.86	9.66	11.01	12.43	15.50	19.52	30.23	88.81
	$E_{5,30}$	5.25	6.33	7.71	8.60	9.63	11.08	13.02	17.12	26.78
$+\infty$	χ^2_5	4.35	5.13	6.06	6.63	7.29	8.12	9.24	11.07	15.09

Table 3.5: The critical values of the finite sample distribution of $l(\theta_0)$, the critical values of the corresponding $E_{k,n}$ distribution and the critical values of the asymptotic χ^2_k distribution.

of $E_{5,20} = 10.90$. The undercoverage is obvious for other n and α combinations as well.

Incidentally, the E calibration based empirical log-likelihood ratio confidence interval may overcover as well in that the actual coverage level is higher than the nominal level. Consider the empirical log-likelihood ratio at the true mean of $X_1, \ldots, X_{10}$, where X_i has an univariate uniform distribution, $U(-1, 1)$. Denote this empirical log-likelihood ratio by $H_{1,10}$. Figure 3.5 shows that the cumulative distribution function of $H_{1,10}$ is lower than that of $E_{1,10}$. In this case, using the critical value of the $E_{1,10}$ distribution will result in an overcoverage problem. It may happen to other cases as well. Consider the empirical log-likelihood ratio at the true mean of $X_1, \ldots, X_{10}$,

where X_i has a conditional normal distribution on $[-2, +2]$; that is the density function of this distribution is $\frac{\phi(x)}{\Phi(2)-\Phi(-2)}$, where $\phi(x)$ and $\Phi(x)$ are the density and the *CDF* of the standard normal distribution, respectively. Denote this empirical likelihood ratio by $T_{1,10}$. Figure 3.6 shows that the cumulative distribution function of the $T_{1,10}$ is also lower than that of the $E_{1,10}$. The $E_{1,10}$ calibration will overcover in this case as well.

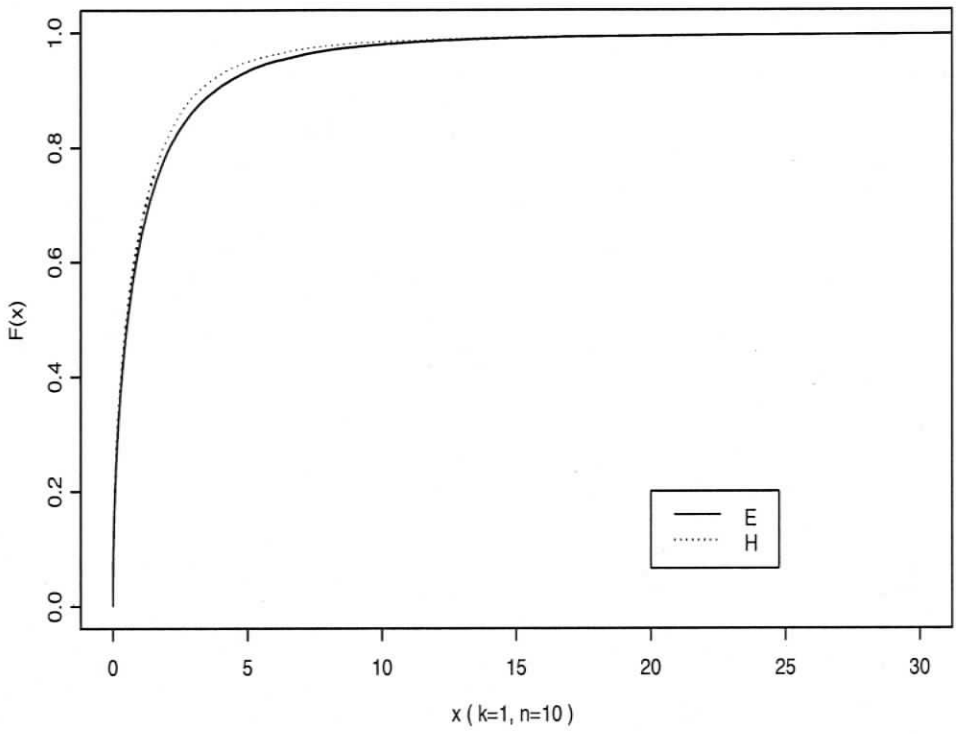

Figure 3.5: Overcoverage Example 1: Plots of the cumulative distribution functions of the $E_{1,10}$ and the $H_{1,10}$.

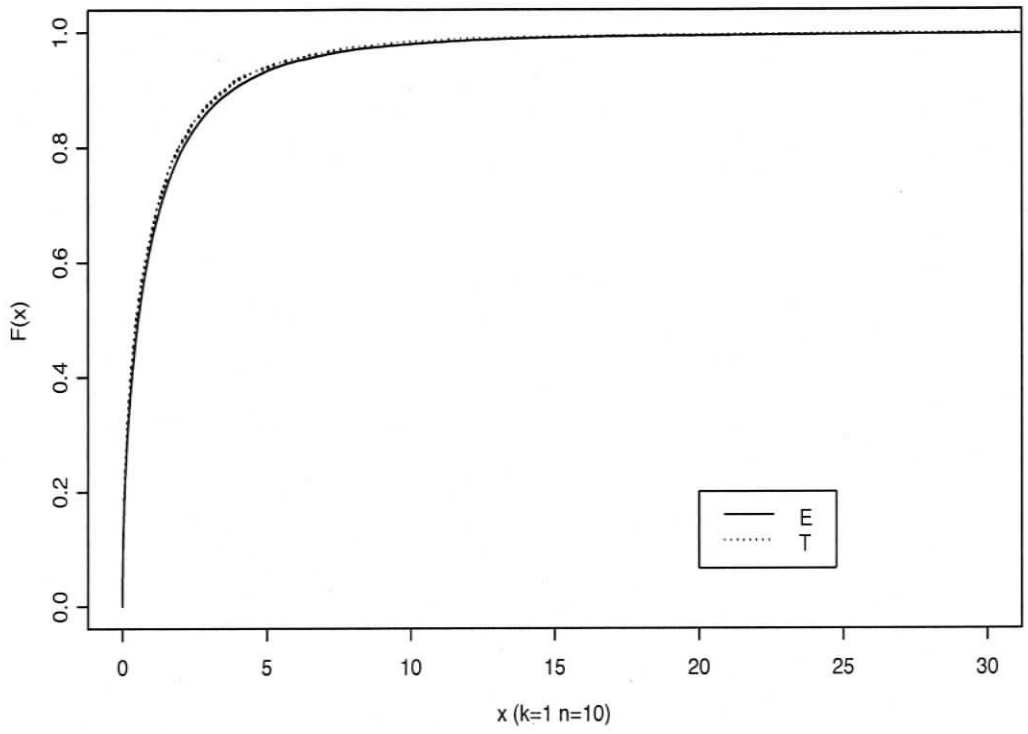

Figure 3.6: Overcoverage Example 2: Plots of the cumulative distribution functions of the $E_{1,10}$ and the $T_{1,10}$.

Chapter 4

Estimation of the E Distribution

Although the E calibration is an attractive method, in order to apply it we need to have its quantiles. The definition of the E random variable involves a set of nonlinear equations (3.2), which has to be solved numerically. Consequently, there is no analytic expression of the E random variable in terms of the random sample from $MVN(\underline{0}, I)$, and this makes it impossible to find the exact cumulative distribution function of the E distribution. Nevertheless, large random samples from an E distribution can be easily generated and such random samples can be used to approximate the quantiles of the E distributions.

Building on our discussions in Chapter 3, in the present chapter, we consider four different methods for estimating (the quantiles of) the E distribution. The four methods that will be considered are: [1] Sample quantile estimation; [2] Nonparametric

smoothing technique; [3] Linear model; [4] Linear model combined with nonparametric smoothing. This chapter consists of four sections, each of which covers one of the four methods.

4.1 Sample Quantile Estimation

The most straightforward way to estimate quantiles of the E distribution is to use the sample quantiles of the large random sample. To obtain random samples from an E distribution, first generate a random sample of n independent observations from $MVN(\underline{0}, I)$ in R^k. Then compute λ by solving equation (3.2) numerically and calculate $l(\underline{0})$ in equation (3.1). We record this $l(\underline{0})$ as E_1, and E_1 is then a random observation from $E_{k,n}$. Repeat this process for $i = 1, 2, ..., m$ and each time record $l(\underline{0})$ as E_i. Let $E^{(1)} \leq E^{(2)} \leq \ldots, \leq E^{(m)}$ be the order statistics of the E_i and $E^{([mp])}$ be the pth sample quantile, where $[mp]$ is the smallest integer greater than or equal to mp. The method of *sample quantile estimation* uses $E^{([mp])}$ as the estimate for $\xi(p, k, n)$. For example, for $m = 10000$, $E^{(9500)}$ is the 0.95 sample quantile and is the estimation of $\xi(0.95, k, n)$.

Glivenko-Cantelli Theorem implies that sample quantiles will be close to the true quantiles if the sample size m is large enough. Recall $F_{k,n}(x)$ is the cumulative distribution of the continuous component of the $E_{k,n}$ distribution and define $F^m_{k,n}$ as

the corresponding empirical distribution:

$$F_{k,n}^{m}(x) = \frac{1}{m}\sum_{i=1}^{m} I_{[Y_i,+\infty)}(x),$$

then

$$P\{\sup_{x} \left|F_{k,n}^{m}(x) - F_{k,n}(x)\right| \to 0\} = 1,$$

as $m \to +\infty$.

This indicates that, for large m, the sample quantile estimates should be accurate. Nevertheless, the sample quantile estimation method suffers an inconsistency problem as pointed out in Tsao (2004c). Recall the monotone property that we discussed at the end of Section 3.1. We expect quantiles of the E distributions, $\xi(1-\alpha, k, n)$ or $e(\alpha, k, n)$, to decrease monotonically for fixed α and k as n increases. In Figure 4.1, the left column contains plots of sample quantile estimation of $\xi(0.5, 1, n)$ and $\xi(0.1, 3, n)$ of the E distributions, respectively. The n value ranges from $(k+1)$ to 200. Both plots show clear overall decreasing trend. The right column contains plots of the tail sections of the corresponding left curves for n between 100 and 200. The downward trend is still visible but the sample quantiles fluctuate and do not decrease monotonically. If we adopt the sample quantiles as our estimates for $\xi(\alpha, k, n)$, then our estimates will not have the monotone property. We refer to this problem as *the inconsistency problem* of the sample quantile estimates. Now, let us see why this inconsistency problem occurs. As n approaches infinity, E distributions are expected to converge to the limiting Chi-square distributions. Therefore the quantiles

of the E distributions are close among themselves for large n values. In such cases, the variances of the quantile estimators are considerable relative to the differences between the values of the quantiles and hence result in the inconsistency problem. Increasing the sample size can alleviate the inconsistency problem by making it less frequent among quantiles for small n but will not solve it for large n values.

Furthermore, the sample quantile method is only applicable to the E distributions from which we have large random samples. And the process of generating these large random samples is quite time consuming. So this method may be appropriate if we want to estimate the critical values of one particular E distribution, but it cannot be easily extended to handle the estimation problem of many E distributions. So the sample quantile method is not suitable for computing comprehensive E quantile tables. This calls for a better method of estimation.

4.2 Nonparametric Smoothing Technique

The most serious problem of the quantile estimates is the inconsistency problem. One way to deal with it is to smooth out the fluctuations of the sample quantiles and thus highlight the decreasing trend. Nonparametric smoothing technique may be appropriate here since it can model the trend by a smooth function without any assumption or prior knowledge of this underlying trend. As in Tsao (2004c), for fixed α and k, we refer to the functional relationship between the critical value $e(\alpha, k, n)$

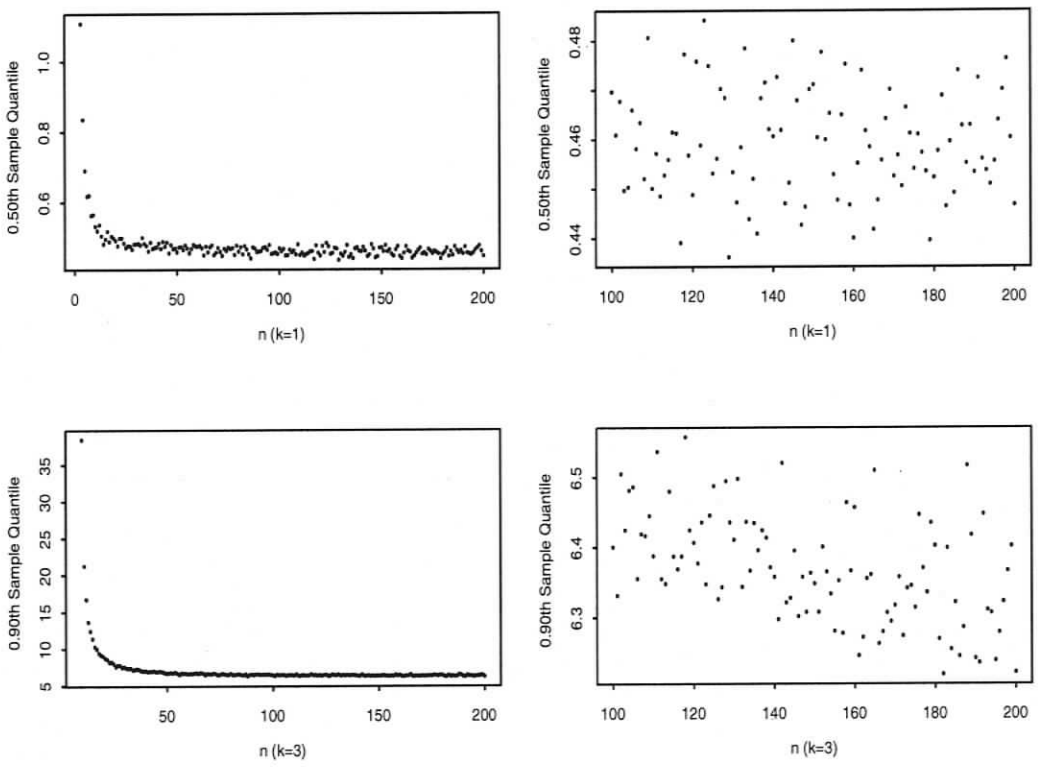

Figure 4.1: The estimated critical value curves $\hat{e}(0.5, 1, n)$ and $\hat{e}(0.1, 3, n)$ by sample quantile estimation method based on samples size of $m = 5 \times 10^4$. Left column: $k + 1 \leq n \leq 200$. Right column: $100 \leq n \leq 200$.

and the second degree of freedom n as the *critical value curve.* We write this curve as $e(\alpha, k, n) = g(n)$. The sample quantiles in Figure 4.1 can be regarded as observations of the true critical value curve plus random noise. The task in this section is to smooth out the noise in a nonparametric way and to recover the underlying critical value curve.

Nonparametric regression is one of the most common nonparametric smoothing techniques. It fits a curve to the data points locally. Any point (x, y) on the smoothed curve is a weighted average of some observations in x's neighborhood and the weights are determined by a kernel function. Mathematically, nonparametric regression is based on the model:

$$y_i = g(x_i) + \epsilon_i, \quad i = 1, .., n,$$

where $g(x)$ is a smooth function and the ϵ_i are independently and identically distributed random errors.

Specifically, we call the estimation of $g(x)$, $\hat{g}(x)$, the nonparametric regression function. Let h $(h > 0)$ be a bandwidth (smoothing parameter). Typically, $\hat{g}(x)$ is the weighted average of observations whose x-coordinates are in $[x - c(h), x + c(h)]$, where $c(h)$ is a function of h. The fitted curve $\hat{g}(x)$ reduces the variability of the original observed responses and $\hat{g}(x)$ is smooth if the kernel function employed is smooth.

The local mean estimator as a simple type of nonparametric regression will be

applied in this section. Here, we will simply provide the functional form of the local mean estimator without reviewing mathematical details (refer to Bowman (1997) to see how it is derived):

$$\hat{g}(x) = \frac{\sum_{i=1}^{n} w(x_i - x; h) y_i}{\sum_{i=1}^{n} w(x_i - x; h)}, \tag{4.1}$$

where the kernel function $w(z, h)$ is a probability density function and the variance of the corresponding distribution is controlled by the bandwidth (smoothing parameter), h. It is often convenient to use for $w(z, h)$ a normal density function with h as the standard deviation. With this choice of $w(z, h)$ and h, observations over an effective range of $4h$ in the covariate axis will contribute to the estimate of $g(x)$ in that observations outside of $[x - 2h, x + 2h]$ will receive little weight.

For the present discussion, let $q(1-\alpha, k, n)$ be the quantile estimate of the critical value curve $e(\alpha, k, n)$. With nonparametric smoothing, we wish to estimate $e(\alpha, k, n)$ through the following model without assuming any functional form,

$$q(1 - \alpha, k, n) = e(\alpha, k, n) + \epsilon_i.$$

First we consider the case of $d = 1$ and $\alpha = 0.5$. Since the underlying relation between n and $e(0.5, 1, n)$ needs not to be assumed, the only step and the key step in model selection is the determination of the smoothing parameter h in (4.1). Smoothing parameter h is of the most importance since it controls the range where the kernel function has effect, and hence the degree of smoothness of the resulting nonparametric fit. If a small smoothing parameter is applied, the fitted curve will

track data closely and it may even inherit the ups and downs of the raw observations (quantile estimates). If a large smoothing parameter is applied, the estimator may lose most fine details of the curvature of the underlying curve and it will only capture the big trend. Clearly, there is a trade-off between the local characteristics and the degree of overall smoothness and trend. An ideal h would allow the fitted curve to capture important local characteristics of the true underlying curve yet achieve a certain degree of smoothness. There are several criteria available in the literature for choosing a suitable h. However, they are only applicable for continuous covariates. In our case, the covariate n is discrete. Thus we follow two simple rules to identify the proper h. The first rule requires that the appropriate h must lead to a small sum of squared residuals (SSR) and a small maximum **absolute** residual (MR). This is to ensure that the resulting curve fits the data well. The second rule requires that the resulting curve is strictly decreasing. This ensures that the monotone property of the critical value curve is not violated.

The top plot in Figure 4.2 shows the result of fitting 0.50th sample quantiles for $k = 1$ and $n \leq 200$ using the local mean estimator (4.1). Note that the curve starts at $n_s = 3$, the smallest n for which $e(\alpha, 1, n)$ is not infinity. The value $h = 21$ is the smallest h to yield a monotonically decreasing fitted curve $\hat{e}(\alpha, k, n)$. Unfortunately, for small n, the fitted curve $\hat{e}(\alpha, k, n)$ is not very satisfactory. The plot shows the resulting curve does not follow the first 3 observations at all and it overestimates for

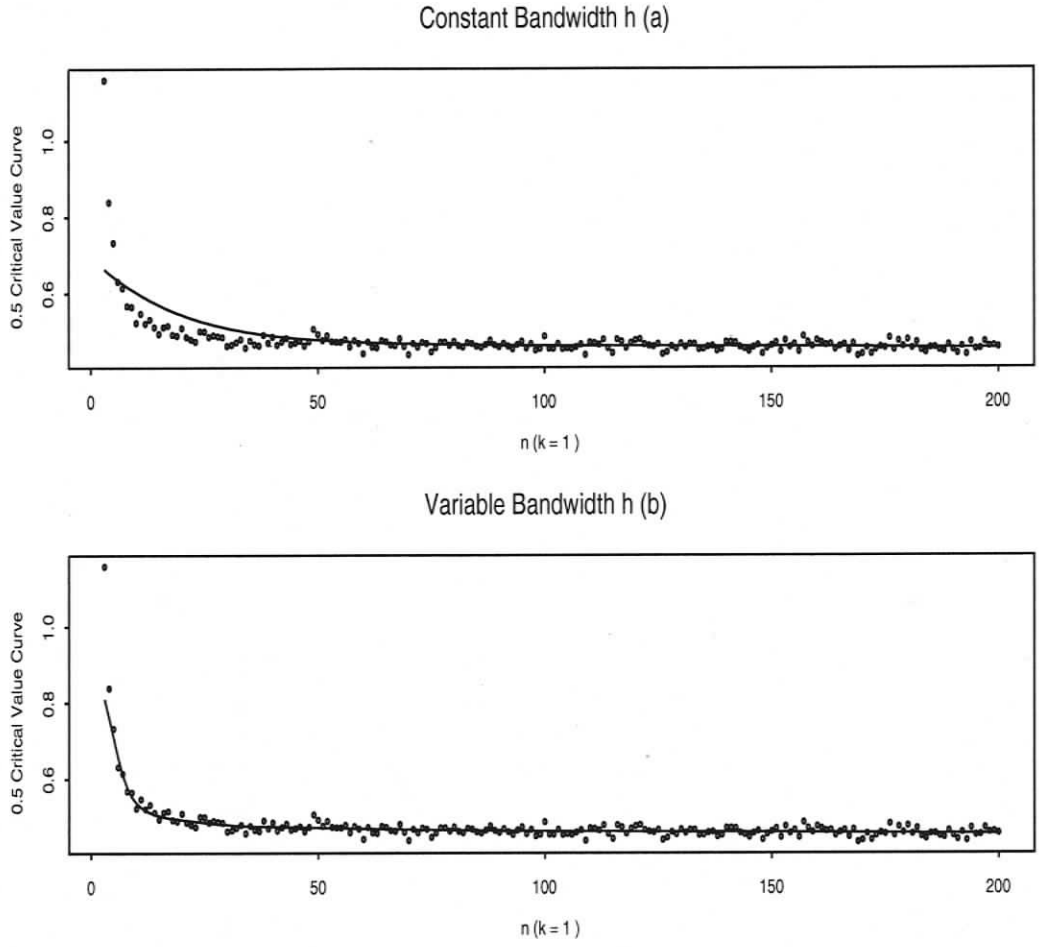

Figure 4.2: The estimated critical value curves $\hat{e}(0.5, 1, n)$ by nonparametric smoothing with constant bandwidth (a) and variable bandwidth (b).

Type	$\underline{h}$	*Seg*	MR	SSR
Fixed h	21	198	0.4955	0.3742
Variable h	(2,5,21)	(10,20,168)	0.3940	0.1942

Table 4.1: Nonparametric smoothing by constant and variable bandwidth for $k = 1$, $\alpha = 0.5$ and $3 \leq n \leq 200$

the range of $4 \leq n \leq 50$. This is because the relatively large smoothing parameter h ($h = 21$) has been used for the entire data range. Note that a spike appears for small n in the sample quantile data in that the values of observations for small n decrease significantly and the first few of them are obviously larger than those observations for large n. For example, the first observation $q(0.5, 1, 3) = 1.1584581$ and the fiftieth observation $q(0.5, 1, 50) = 0.4924798$. Taking average of a large number of observations will smooth out the spike at the beginning. So for small n, we do not prefer a large smoothing parameter. But h cannot be less than 21 either. Otherwise, the monotone property will be violated. This suggests that the constant bandwidth is not desirable and variable bandwidth should be applied to improve the estimation, where we use a small bandwidth for small n to preserve the spike at the beginning and a large bandwidth for large n to ensure the fitted curve is monotonically decreasing.

Based on the shape of the sample quantile plots, the underlying curve can be separated into three segments, the steep drop, the concave part and the flat tail. The

first segment calls for relatively small bandwidth. A small value of h like 2 would be enough to obtain the smooth downward curve. The second segment needs a moderate h like 5 to catch the trend and also remove the roughness. The rest of the points form the tail with a slow decreasing pattern, and a larger h suffices to clarify the decreasing trend. To summarize, the following steps will be taken to apply local mean estimator with variable bandwidth:

1. Apply the local mean estimator for the entire data range and find the smallest smoothing parameter h that will give a monotonically decreasing fit. Record this h as h_0 and it will be used for the smoothing at the tail.

2. Divide the data from n_s to 200 into three segments, and write the numbers of observations contained in three segments as $Seg = (n_1, n_2, n_3)$.

3. Apply the local mean estimator for the three segments with bandwidth $2, 5, h_0$ respectively. For convenience, we use $\underline{h} = (2, 5, h_0)$ to represent the set of bandwidth used. In the case of the constant bandwidth, it is understood that $\underline{h}$ is a scaler, for example, $\underline{h} = 21$ in the case of $k = 1$ and $\alpha = 0.5$.

Consider the case of $k = 1$ and $\alpha = 0.5$ again. In this case, the three segments will contain 10, 20, and 168 $(200 - n_s - 10 - 20)$ observations (quantile estimates), respectively. So with $\underline{h} = (2, 5, 21)$, we would use a bandwidth of 2 for the first 10 observations and 5 for the next 20, and 21 for the remaining 168. The fitted curves

α	n_s	h	*Seg*	MR	SSR	MR(n_s+2)	SSR(n_s+2)
0.5	3	(2,5,21)	(10,20,168)	0.3940	0.1942	0.0334	0.0216
0.8	4	(2,5,18)	(15,25,157)	1.6060	2.9357	0.0755	0.1540
0.9	5	(2,5,24)	(20,30,146)	2.2931	6.2841	0.1070	0.3705
0.95	6	(2,5,25)	(25,35,135)	3.7449	16.2722	0.1776	0.9216
0.99	8	(2,5,45)	(30,40,123)	17.2266	314.2057	0.5570	6.2616

Table 4.2: Nonparametric smoothing summary by variable bandwidth for $k = 1$.

are shown at the bottom of Figure 4.2. The maximum absolute residual (MR) and the sum of squared residual (SSR) are reported in Table 4.1 together with those of the constant bandwidth fitting as a comparison. The curve now fits most of the data well except for the first two observations. In particular, for most small n, the fit is quite improved. Although the largest residual has not decreased significantly (0.4955 for the constant bandwidth vs 0.3940 for the variable bandwidth), the sum of squared residuals has decreased a lot from 0.3742 to 0.1942, which indicates an overall improvement.

Tables 4.2 and 4.3 show the fitted results by local mean estimator with variable bandwidth for the case of $k = 1$ and $k = 3$. Several critical value curves are studied such as $\alpha = 0.50, 0.20, 0.10, 0.05, 0.01$. Again, n_s represents the smallest n for which the critical value is not infinity, h is the bandwidth combination applied and *Seg* indicates the data ranges that different bandwidths are applied to. Maximum absolute

α	n_s	h	Seg	MR	SSR	MR(n_s+2)	SSR(n_s+2)
0.5	7	(2,5,11)	(10,20,164)	0.2420	0.1412	0.0843	0.1217
0.8	9	(2,5,13)	(15,25,152)	4.9381	27.5377	0.1254	0.4329
0.9	10	(2,5,9)	(20,30,141)	5.1076	72.4308	0.2569	0.9465
0.95	12	(2,5,9)	(25,35,129)	5.2158	67.3522	0.2933	1.8877
0.99	15	(2,5,14)	(30,35,121)	5.9807	109.0530	0.7285	9.3171

Table 4.3: Nonparametric smoothing summary by variable bandwidth for $k = 3$.

residual (MR) and sum of squared residuals (SSR) are calculated twice; once by using the whole data range and another time by excluding the first two observations, *i.e.* starting at $n_s + 2$.

Although the values of the maximum absolute residual (MR) and the sum of squared residuals (SSR) for the entire data range are relatively large, for example $SSR = 09.0530$ for $k = 3, \alpha = 0.99$, the first two data points are responsible for a large portion of these errors. Removing the first two points, we see that the maximum absolute residual MR$(n_s + 2)$ and the sum of squared residuals SSR$(n_s + 2)$ are quite reasonable, for example SSR reduces to 9.3171 in the last case. This implies that the smoothed curves fit the data well except for the first two points. Consider the number of data points involved, which is 183 in this case $(k = 3, \alpha = 0.99)$, this SSR$(n = 2)$ is relatively small and represents a good fit.

We now consider the selection of the combination of bandwidth $\underline{h}$ and Seg. In our

method, the first and second components 2 and 5 are the pre-chosen numbers and the third component are different depending on the critical value curve considered. However, the division of the corresponding segments can have many possible choices and hence the combination of $\underline{h}$ and Seg. To keep it simple and to lessen the degree of arbitrariness, the number of observations assigned for each segment is chosen to be a multiple of 5 only (such as $5, 15, 20, 25, ...$). Consequently, Seg can only be, for example, $(5, 10, r)$ or $(10, 30, r)$, where r is the remaining of observations. Different fitting divisions were tried and some were eliminated since they did not yield strictly decreasing curves. Others were compared using the maximum absolute residual (MR) and the sum of squared residuals (SSR) and the best combinations are reported in Tables 4.2 and 4.3 here.

In this section, we performed the nonparametric regression for estimating the critical value curve, using the quantile estimates as the data. The smoothing method with variable bandwidth is of the most interest. Because the nonlinearity of the critical value curve is not yet fully understood, a parametric model is not available. The nonparametric regression with variable bandwidth provides a valuable means in such situations by displaying the underlying trend of the data and producing a smooth monotonically decreasing fit without the need for a parametric model.

However, this nonparametric regression method has several problems. First of all, in despite of providing good fit to most data points, for the first several points,

in particular, the first two points, the method under-fit in that the fitted values tend to be smaller than the observations. A way to improve this is to analyze the data more carefully. We can divide the data range into more intervals and apply a very small h like 0.25 for the first 2 or 3 data points. Then gradually increase h for the remaining data points. Secondly, the selection of the bandwidth vector $\underline{h}$ and the data ranges are based on mostly empirical considerations and is, to a certain degree, subjective. Different bandwidth and/or different segment selections will lead to different fit. There can be a large number of possible combinations of bandwidth and segments, and the optimal combination could be very complicated to identify. Thirdly, we examined the monotone property for the fitted curve through calculating the difference between the adjacent fitted values, *i.e.*, $\hat{g}(i+1) - \hat{g}(i)$ for $i = n_s, n_s + 1, \ldots, 200$. If all the differences are negative, then we regard the fit as a monotone decreasing fit. However, the fitted curve $\hat{g}(t)$ is also defined at non-integer t values and it is possible that $\hat{g}(t)$ is not strictly monotonically decreasing but this is of little concern as we are only interested in the monotone property of the set $\{\hat{g}(i), i = n_s, n_s + 1, \ldots\}$. Finally, there are few diagnostic and inference tools available for this nonparametric method. Also since our covariate n is not a continuous variable, it is more difficult to do model-checking.

4.3 Linear Models

Since the information of the convergence rate is available for E quantiles, it is also possible to build parametric models for sample quantiles to obtain better estimations. As proved in Chapter 3 that the quantiles of the $E_{k,n}$ distributions, *i.e.*, $e(\alpha, k, n)$, converges to the corresponding quantiles of $\chi^2_{k,1-\alpha}$ at the rate of $1/n$. It suggests that for large n values, E quantiles can be modeled in the following form:

$$e(\alpha, k, n) = \chi^2_{k,1-\alpha} + \frac{b}{n} + \epsilon, \quad n \geq n_0, \tag{4.2}$$

where b and n_0 are the parameters of the nonlinear model and the last term ϵ is a small random error normally distributed with mean 0 and a constant variance not depending on n. The suitable value for n_0 will depend on α and k. For some combination of α and k, the asymptotic effect as represented by model (4.2) applies to n as small as 10, in which case $n_0 = 10$. For other combinations, the asymptotic effect is not present until much later, in which case n_0 is much larger.

Let $x = 1/n$ and $x_0 = 1/n_0$, model (4.2) can be transformed into the linear form:

$$\hat{e}(\alpha, k, x) = \chi^2_{k,1-\alpha} + b \times x + \epsilon, \quad x \leq x_0. \tag{4.3}$$

. Note that the intercept term $\chi^2_{k,1-\alpha}$ is the limiting value of $e(\alpha, k, n)$ as $n \to +\infty$. Therefore, as $n \to +\infty$ or equivalently as $x \to 0$, the model conforms to $E(\hat{e}(\alpha, k, n)) \to \chi^2_{k,1-\alpha}$. Thus the model has taken into consideration the limiting behavior of $e(\alpha, k, n)$. Further, we also require that $b > 0$. With this condition,

$\chi^2_{k,1-\alpha} + b/n$ is monotonically decreasing with respect to n and thus the model (4.2) has also incorporated the monotone property.

Consider the same case used in the last section, which is $k = 1$ and $\alpha = 0.5$. The resulting linear model for $e(0.5, 1, n)$ is:

$$\hat{e}(0.50, 1, n) = 0.8024 \times n^{-1} + 0.4549 \quad \text{for } n \geq 10 \tag{4.4}$$

Note that the model starts at 10 where the asymptotic effect starts. This is verified by looking at if the constant variance assumption and the linear structure are appropriate. For the above case, the constant variance assumption of (4.3) cannot be met if the observations for $n \leq 10$ are included. The value of n_0 is usually larger than the original n_s for the data and we will describe the determination of n_0 later.

Figure 4.3 shows different plots for model checking. The residual plot and QQ-plot show that the equal variance and normal assumptions of the linear model are reasonable. These suggest that $n_0 = 10$ is a proper starting value. The left plot in the second row shows the fitted regression line superimposed on the scatter plot of the transformed data, i.e. $e(\alpha, k, x)$ against x. Data points are aggregated to the left, because most of them has small x-coordinates after the $1/n$ transformation. Nevertheless, the overall strong linear relationship is visible. In right plot of the same row, data points are spread evenly because the x-coordinates changes to n from $1/n$. The fitted curve is seen to have provided a good fit to the data.

The standard summary statistics for the regression model are also quite satis-

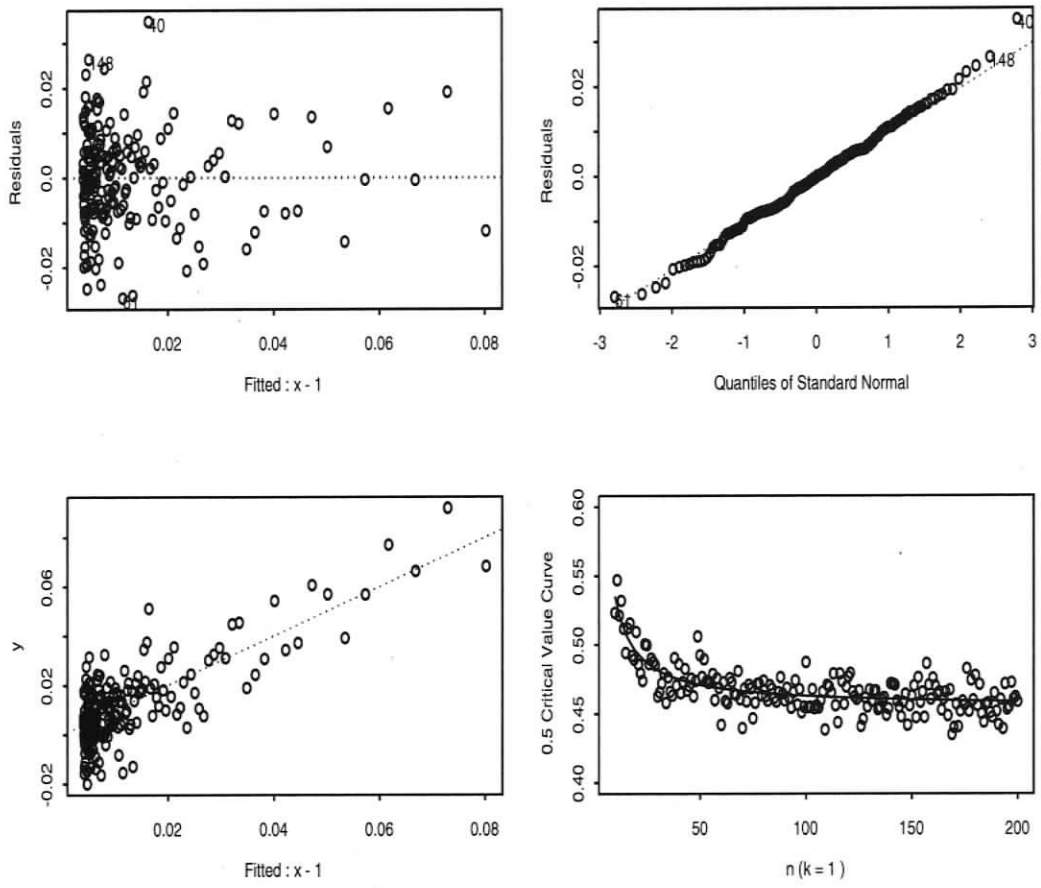

Figure 4.3: Linear model for $\hat{e}(0.5, 1, n)$. The left top plot is the residual plot. The right top plot is the QQ plot. The bottom left plot is the transformed data with the linear regression line and the bottom right plot is the sample quantile data with the fitted curve obtained through the conversion of the regression line.

factory. The R^2 statistics is 0.7428. In other words, by using $1/n$ as a predictor, we have explained nearly 74.28% of the variability in $q(0.5, 1, n)$. And as we would expect with such a high R-squared value, the estimated value of the coefficient b is significantly different from zero: its t-statistic is greater than 23, with 189 degree of freedom, which leads to a p-value of essentially 0. The residual sum of square of the model is 0.0223, which is very small considering that the fit involves 190 data points. Also, the maximum and minimum values of the residuals are also small. These are 0.0348 and -0.0269, respectively.

We now discuss the step of determining n_0. The n_0 has to be determined for every α and k to make sure the model (4.3) is appropriate for the particular combination. For each combination, we begin by setting n_0 to 10 and fit a linear model. By examining the residual plot and the goodness-of-fit of this model, we can determine if $n_0 = 10$ gives a good fit. If it does not, we repeat the above process for $n_0 = 20, 30, \ldots, 100$ to look for the proper n_0. The search is not as time consuming as it might seem because in all cases examined thus far the appropriate n_0 values are all less than 100. We only need to examine a small number values to find the final value n_0 so that model (4.3) fits the sample quantiles well.

Tables 4.4 and 4.5 show the fitted results for $k = 1, 3$ at $\alpha = 0.5, 0.8, 0.9, 0.95, 0.99$. The tables contain the estimated equations, the domain for the equations, maximum absolute residual (MR), sum of squared residuals (SSR) and the R^2 values. Based on

α	Estimated Critical Value Curve	Domain	MR	SSR	R^2
0.50	$e(\alpha, 1, n) = 0.8024 \times n^{-1} + 0.4549$	$n \geq 10$	0.0348	0.0223	0.7428
0.20	$e(\alpha, 1, n) = 2.8808 \times n^{-1} + 1.6423$	$n \geq 30$	0.0682	0.1381	0.6347
0.10	$e(\alpha, 1, n) = 4.9297 \times n^{-1} + 2.7055$	$n \geq 30$	0.1087	0.3240	0.6844
0.05	$e(\alpha, 1, n) = 7.4049 \times n^{-1} + 3.8415$	$n \geq 30$	0.1868	0.8608	0.6481
0.01	$e(\alpha, 1, n) = 16.4009 \times n^{-1} + 6.6349$	$n \geq 40$	0.4565	5.5754	0.4951

Table 4.4: Linear model for $k = 1$

α	Estimated Critical Value Curve	Domain	MR	SSR	R^2
0.50	$e(\alpha, 1, n) = 7.1179 \times n^{-1} + 2.3660$	$n \geq 30$	0.0826	0.1133	0.9282
0.20	$e(\alpha, 1, n) = 14.0136 \times n^{-1} + 4.6416$	$n \geq 60$	0.1314	0.3496	0.8691
0.10	$e(\alpha, 1, n) = 19.7654 \times n^{-1} + 6.2514$	$n \geq 70$	0.1790	0.6761	0.8445
0.05	$e(\alpha, 1, n) = 27.5035 \times n^{-1} + 7.8147$	$n \geq 70$	0.2572	1.4655	0.8291
0.01	$e(\alpha, 1, n) = 50.2390 \times n^{-1} + 11.3449$	$n \geq 70$	0.6580	6.3774	0.7882

Table 4.5: Linear model for $k = 3$

the small MR, SSR and large R^2 values, model (4.3) works quite well. As α increases, the value of R^2 reported in the table tends to decrease. This is because when α is larger, the asymptotic behavior is present later, *i.e.*, for a larger n. If we simply increase the corresponding value of n_0, the value of R^2 will rise up. The linear model (4.3) incorporates important information such as monotone property, the limiting value and the convergence rate of $e(\alpha, k, n)$, and produce more efficient estimations of the E quantiles. Note that we could have fitted the non-linear model (4.2) as in Tsao (2004c). The advantage of the linear model (4.3) is that computation and inference for model parameters are easier to handle. One drawback of both (4.2) and (4.3) is that they cannot be used for estimation of $e(\alpha, k, n)$ for small n, where $n < n_0$.

4.4 Combining Linear Model with Nonparametric Smoothing

Neither nonparametric smoothing technique nor linear model is a problem-free solution for estimating E quantiles. For the nonparametric smoothing method, we have to deal with the selection of the smoothing parameters. And being a nonparametric method, it does not make use of information such as the limiting value and convergence rate of the quantiles. On the other hand, the linear model does not have these problems but it cannot be used for small n values. To draw strength from both

methods, we can combine these two ways together to produce better estimations for the whole range of n, both small and large. That is, we use linear model to obtain estimations for the tail first and then applying nonparametric method to smooth observations not at the tail.

The following three steps may be used to build final E quantile estimations:

1. Fit the linear model (4.3) for a properly chosen large n and keep the fitted values.

2. Replace the sample quantiles with the fitted values (for those where the fitted values are available) produced by the linear model (4.3). The new dataset now consists of observations not used in fitting the linear model and the fitted values. This ensures in the next step the smooth estimates are available for the entire data range.

3. Apply nonparametric smoothing to the new dataset.

Again, consider the case of $k = 1$ and $\alpha = 0.5$, the proper n_0 for the linear model is 10. The scatter plot of the new data obtained by following the Step 2 is depicted in Figure 4.4. Now smaller h could be applied since the tail, where a large bandwidth used to be needed, is already smooth. The curve in the plot is obtained by applying $h = 0.5$ for the first 3 observation and $h = 1$ for the rest 194 observations and it visually fits data very well. The maximum residual is 0.0348 and the sum of squared

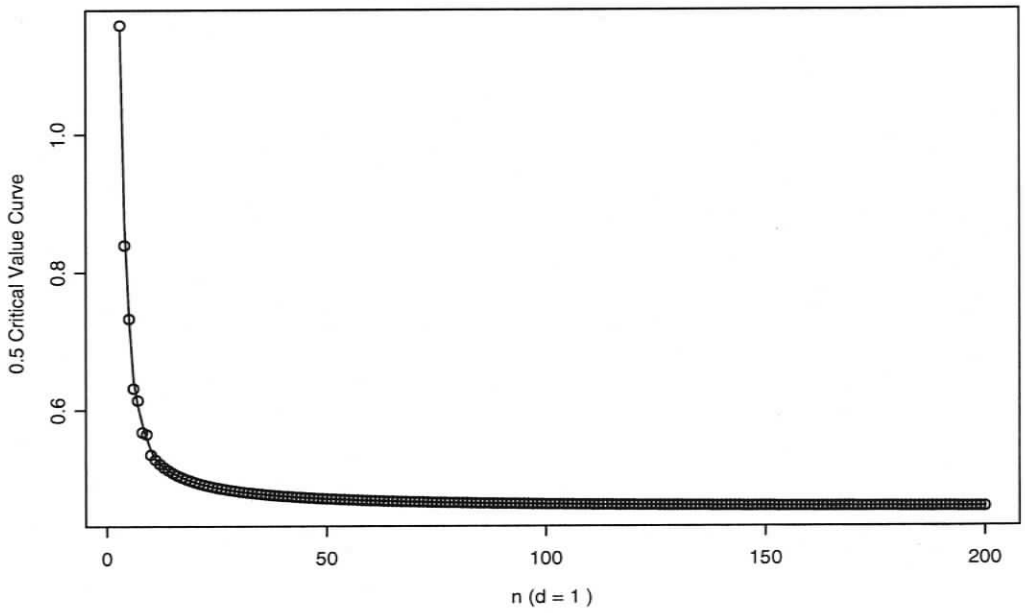

Figure 4.4: Linear model combined with nonparametric smoothing for $\hat{e}(0.5, 1, n)$.

residual is 0.0237. Compared with results in Table 4.2 by the nonparametric method only, they are both reduced substantially.

Tables 4.6 and 4.7 show the fitted results by the linear model combined with nonparametric smoothing. Comparing these with Tables 4.2 and 4.3, the maximum absolute residual (MR) and the sum of squared residuals (SSR) for every case decreased significantly. The estimated critical values by this combined method is also available in Table 1 in the appendix.

α	n_s	n_0	h	Seg	MR	SSR
0.5	3	10	(0.5, 1)	(3, 194)	0.0348	0.0237
0.8	4	30	(0.5, 2)	(3, 193)	0.1682	0.2202
0.9	5	30	(0.5, 2)	(3, 192)	0.2030	0.4447
0.95	6	30	(0.5, 2)	(3, 191)	0.3703	1.2686
0.99	8	40	(0.25, 2)	(3, 189)	0.6076	7.7306

Table 4.6: Linear model combined with nonparametric smoothing for $k = 1$

α	n_s	n_0	h	Seg	MR	SSR
0.5	7	30	(0.5,2)	(3,190)	0.1307	0.1641
0.8	9	60	(0.5,2)	(3,188)	0.4995	1.0226
0.9	10	70	(0.25,4)	(10,180)	0.3300	1.1267
0.95	12	70	(0.25,4)	(8,180)	0.4165	2.7981
0.99	15	70	(0.25,3)	(10,175)	0.7027	11.1714

Table 4.7: Linear model combined with nonparametric smoothing for $k = 3$

4.5 Concluding Remarks

Four different methods of estimating the E quantiles were considered in this chapter. They are sample quantile method, nonparametric smoothing method, linear regression method and the method based on the combination of nonparametric smoothing and linear regression. Based on the previous discussion, the linear model combined with nonparametric regression (the combined method) is the best way so far to model sample quantiles of the E distribution. It can generate the comprehensive quantile table of the E distribution, simply, accurately and systematically. Sometimes, the sample quantile method can be used to generate the E quantiles as well. It is a simple and easy way when we just want to compute the critical values of one particular E distribution.

Bibliography

[1] Barndorff-Nielsen, O.E. & Cox, D. R. (1989), *Asymptotic techniques for use in statistics.* Chapman and Hall, London.

[2] Bowman, A.W. (1997), *Applied smoothing techniques for data analysis: The kernel approach with S-Plus illustrations.* Clarendon Press, Oxford.

[3] Chen, S. & Hall, P. (1993), Smoothed empirical likelihood ratios for quantiles. *Ann. Stat.* **21**, 1166- 1181.

[4] DiCiccio, T., Hall, P. & Romano, J. (1991), Empirical likelihood is bartlett-correctable. *Ann. Statist.* **19**, 1053-1061.

[5] Fujikoshi, Y. (1987), *Error bounds for asymptotic expansions of scale mixtures of distributions.* Hiroshima mathematical Journal. **17**, 309-324.

[6] Hall, P. & La Scala, B. (1990), Methodology and algorithms of empirical likelihood. *Int. Statist. Review.* **58**, 109-127.

[7] Hall, P. & Owen, A.B. (1993), *Empirical likelihood confidence bands in density estimation.* Jour. Comp. Graph. Stat. **2**, 273-289.

[8] Kiefer, J. & Wolfowiz, J. (1956), Consistency of the Maximum Likelihood Estimator in the Presence of Infinitely Many Incidental Parameters. *Ann. Statist.* **27**, 887-906.

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

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

[11] Owen, A.B. (1991), Empirical likelihood for linear models. *Ann. Statist.* **19**, 1725-1747.

[12] Owen, A.B. (1992), Empirical likelihood and generalized projection pursuit. *Dept. of Statistics Tech.* Rep. **393**, Standford University, Stanford CA.

[13] Owen, A.B. (1995), Empirical likelihood confidence bands for a distribution function. *J. Amer. Statist. Assoc.* **90**, 516-521.

[14] Owen, A.B. (2001), *Empirical likelihood.* Chapman and Hall, London.

[15] Qin, J. (1994), Empirical Likelihood Ratio confidence intervals for the difference of two sample means. *Ann. Inst. Meth. Statist.* **47**, 117 -126.

[16] Qin, J. & Lwaless, J.F. (1994), Empirical Likelihood and General Estimating Equations. *Ann, Statist.* **22**, 300-325.

[17] Serfling, R. J. (1980), *Approximation theorems of mathematical statistics.* Wiley, New York.

[18] Tsao, M. (2004a), A new method of calibration for the empirical log-likelihood ratio. *Statist. Prob. Lett.* **68**, 305-314.

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

[20] Tsao, M. (2004c), Computing the E distributions. *Technical Report*, Department of Mathematics and Statistics, University of Victoria, British Columbia, Canada.

[21] Wood, A.T.A., Do, K.A. & Broom, B.M. (1994), Sequential Linearization of Empirical Likelihood Constraints with Application to U-Statistics. Center for Mathematics and its Applications, *Research Report.* **No. SRR** 33-94.

[22] Wendel, J.G. (1962), A problem in geometric probability. *Math. Scandinavica* **11**, 109-111.

Appendix A

Table 1 contains commonly used critical values for E distributions with $k = 1$ and 3 and $n \leq 80$ by the combined method of nonparametric smoothing and linear models. In Table 1, each column has a value marked with a *, indicating values below this value were computed by the linear model and values above were smoothed by nonparametric smoothing.

Table 1: Estimated critical values of $E_{k,n}$

	Dimension $k = 1$					Dimension $k = 3$				
$n\backslash\alpha$	0.50	0.20	0.10	0.05	0.01	0.50	0.20	0.10	0.05	0.01
2	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	NA	NA	NA	NA	NA
3	1.1528	$+\infty$	$+\infty$	$+\infty$	$+\infty$	NA	NA	NA	NA	NA
4	0.8686	5.3783	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$
5	0.7317	3.5922	8.7508	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$
6	0.6439	2.8721	6.3968	13.3352	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$	$+\infty$
7	0.6071	2.5218	4.9862	9.3301	$+\infty$	7.2824	$+\infty$	$+\infty$	$+\infty$	$+\infty$
8	0.5802	2.3512	4.3241	7.3920	40.6302	5.6759	$+\infty$	$+\infty$	$+\infty$	$+\infty$
9	0.5592	2.2355	4.0371	6.4892	19.6858	4.6732	18.9767	$+\infty$	$+\infty$	$+\infty$
10	0.5419	2.1421	3.8429	6.0700	15.3131	4.1719	13.3265	38.4067	$+\infty$	$+\infty$
11	*0.5297	2.0683	3.6782	5.7522	13.1470	3.8992	10.4834	21.2451	$+\infty$	$+\infty$
12	0.5224	2.0099	3.5379	5.4676	12.2216	3.6888	9.2273	16.7099	38.6503	$+\infty$
13	0.5170	1.9635	3.4192	5.2265	11.3883	3.5082	8.5112	13.6151	21.9027	$+\infty$
14	0.5125	1.9276	3.3229	5.0355	10.6919	3.3579	7.9584	12.3782	19.1957	$+\infty$
15	0.5087	1.9019	3.2510	4.8941	10.1487	3.2359	7.4896	11.2988	16.7097	53.4751
16	0.5053	1.8844	3.2022	4.7923	9.7383	3.1381	7.1011	10.2314	14.7302	33.9416
17	0.5023	1.8714	3.1695	4.7123	9.4151	3.0598	6.7874	9.8741	13.6272	29.4630
18	0.4997	1.8592	3.1425	4.6360	9.1305	2.9970	6.5412	9.3782	12.7739	24.1717
19	0.4973	1.8460	3.1123	4.5547	8.8577	2.9459	6.3511	9.1222	12.6736	25.6625
20	0.4952	1.8319	3.0768	4.4732	8.6014	2.9031	6.2010	8.8876	12.0230	21.7649
21	0.4932	1.8181	3.0406	4.4044	8.3828	2.8668	6.0754	8.6983	11.6997	21.5067
22	0.4915	1.8071	3.0108	4.3576	8.2166	2.8359	5.9636	8.5216	11.4003	20.1598
23	0.4899	1.8003	2.9904	4.3310	8.1005	2.8086	5.8615	8.3580	11.1257	18.5798
24	0.4884	1.7965	2.9768	4.3145	8.0219	2.7816	5.7685	8.2077	10.8759	17.4605
25	0.4871	1.7915	2.9651	4.2972	7.9632	2.7521	5.6847	8.0706	10.6499	17.0404
26	0.4858	1.7824	2.9523	4.2744	7.9033	2.7203	5.6109	7.9463	10.4459	16.7336
27	0.4847	1.7697	2.9387	4.2474	7.8264	2.6889	5.5468	7.8340	10.2615	16.4223
28	0.4836	1.7571	2.9257	4.2178	7.7318	2.6606	5.4901	7.7329	10.0945	16.1159
29	0.4826	1.7474	2.9125	4.1854	7.6311	2.6369	5.4378	7.6418	9.9427	15.8216
30	0.4817	*1.7410	*2.8979	*4.1505	7.5372	*2.6179	5.3885	7.5598	9.8044	15.5449
31	0.4808	1.7367	2.8826	4.1176	7.4582	2.6032	5.3434	7.4860	9.6785	15.2895
32	0.4800	1.7332	2.8692	4.0914	7.3997	2.5921	5.3038	7.4192	9.5639	15.0573
33	0.4793	1.7302	2.8591	4.0735	7.3626	2.5834	5.2700	7.3585	9.4601	14.8476
34	0.4786	1.7275	2.8522	4.0621	7.3403	2.5763	5.2418	7.3031	9.3663	14.6564
35	0.4779	1.7250	2.8472	4.0542	7.3191	2.5701	5.2191	7.2521	9.2818	14.4778
36	0.4772	1.7227	2.8430	4.0479	7.2864	2.5643	5.1995	7.2049	9.2059	14.3059
37	0.4766	1.7205	2.8392	4.0422	7.2386	2.5589	5.1796	7.1611	9.1378	14.1381
38	0.4761	1.7184	2.8356	4.0369	7.1819	2.5538	5.1574	7.1207	9.0768	13.9764
39	0.4755	1.7164	2.8323	4.0318	7.1274	2.5490	5.1351	7.0836	9.0219	13.8270
40	0.4750	1.7146	2.8291	4.0270	*7.0833	2.5444	5.1161	7.0501	8.9724	13.6963
χ^2_k	0.4549	1.6424	2.7055	3.8415	6.6349	2.3660	4.6416	6.2514	7.8147	11.345

Table 1 Cont'ed: Estimated critical values of $E_{k,n}$

	Dimension $k = 1$					Dimension $k = 3$				
$n\backslash\alpha$	0.50	0.20	0.10	0.05	0.01	0.50	0.20	0.10	0.05	0.01
41	0.4745	1.7128	2.8261	4.0225	7.0524	2.5400	5.1021	7.0198	8.9274	13.5876
42	0.4741	1.7111	2.8232	4.0182	7.0324	2.5358	5.0922	6.9925	8.8858	13.4990
43	0.4736	1.7095	2.8204	4.0140	7.0190	2.5319	5.0847	6.9674	8.8468	13.4235
44	0.4732	1.7080	2.8178	4.0101	7.0089	2.5281	5.0784	6.9436	8.8095	13.3515
45	0.4728	1.7065	2.8153	4.0063	7.0002	2.5245	5.0718	6.9204	8.7734	13.2745
46	0.4724	1.7051	2.8129	4.0027	6.9921	2.5210	5.0623	6.8969	8.7383	13.1875
47	0.4720	1.7038	2.8106	3.9993	6.9845	2.5177	5.0475	6.8730	8.7042	13.0916
48	0.4717	1.7025	2.8084	3.9960	6.9772	2.5145	5.0276	6.8486	8.6712	12.9934
49	0.4713	1.7013	2.8063	3.9928	6.9702	2.5115	5.0069	6.8242	8.6397	12.9023
50	0.4710	1.7001	2.8043	3.9898	6.9634	2.5086	4.9910	6.8004	8.6095	12.8265
51	0.4707	1.6989	2.8024	3.9869	6.9570	2.5058	4.9812	6.7776	8.5807	12.7692
52	0.4704	1.6979	2.8005	3.9841	6.9508	2.5031	4.9731	6.7563	8.5527	12.7281
53	0.4701	1.6968	2.7987	3.9814	6.9448	2.5005	4.9611	6.7368	8.5256	12.6963
54	0.4698	1.6958	2.7970	3.9788	6.9390	2.4980	4.9452	6.7194	8.4991	12.6663
55	0.4695	1.6948	2.7953	3.9763	6.9335	2.4956	4.9306	6.7040	8.4738	12.6332
56	0.4693	1.6939	2.7937	3.9739	6.9281	2.4932	4.9217	6.6910	8.4503	12.5958
57	0.4690	1.6930	2.7921	3.9715	6.9230	2.4910	4.9176	6.6804	8.4294	12.5568
58	0.4688	1.6921	2.7906	3.9693	6.9180	2.4888	4.9131	6.6722	8.4119	12.5205
59	0.4685	1.6913	2.7892	3.9671	6.9132	2.4868	4.9049	6.6663	8.3981	12.4902
60	0.4683	1.6904	2.7878	3.9650	6.9086	2.4847	*4.8935	6.6622	8.3878	12.4660
61	0.4681	1.6897	2.7864	3.9630	6.9041	2.4828	4.8820	6.6592	8.3800	12.4458
62	0.4679	1.6889	2.7851	3.9610	6.8997	2.4809	4.8727	6.6564	8.3735	12.4258
63	0.4677	1.6881	2.7839	3.9591	6.8955	2.4791	4.8661	6.6529	8.3668	12.4031
64	0.4675	1.6874	2.7826	3.9573	6.8914	2.4773	4.8613	6.6478	8.3585	12.3761
65	0.4673	1.6867	2.7815	3.9555	6.8875	2.4756	4.8576	6.6404	8.3475	12.3441
66	0.4671	1.6861	2.7803	3.9538	6.8836	2.4739	4.8542	6.6306	8.3335	12.3067
67	0.4669	1.6854	2.7792	3.9521	6.8799	2.4723	4.8510	6.6186	8.3168	12.2640
68	0.4667	1.6848	2.7781	3.9504	6.8763	2.4707	4.8479	6.6048	8.2981	12.2177
69	0.4666	1.6842	2.7770	3.9489	6.8728	2.4692	4.8449	*6.5902	*8.2784	*12.1708
70	0.4664	1.6836	2.7760	3.9473	6.8694	2.4677	4.8420	6.5757	8.2591	12.1275
71	0.4662	1.6830	2.7750	3.9458	6.8661	2.4663	4.8392	6.5619	8.2410	12.0912
72	0.4661	1.6824	2.7741	3.9444	6.8629	2.4649	4.8364	6.5495	8.2249	12.0632
73	0.4659	1.6819	2.7731	3.9430	6.8597	2.4636	4.8337	6.5387	8.2110	12.0427
74	0.4658	1.6813	2.7722	3.9416	6.8567	2.4622	4.8311	6.5297	8.1994	12.0279
75	0.4656	1.6808	2.7713	3.9403	6.8537	2.4609	4.8286	6.5223	8.1899	12.0165
76	0.4655	1.6803	2.7705	3.9390	6.8508	2.4597	4.8261	6.5161	8.1820	12.0069
77	0.4654	1.6798	2.7696	3.9377	6.8480	2.4585	4.8237	6.5110	8.1753	11.9982
78	0.4652	1.6793	2.7688	3.9365	6.8453	2.4573	4.8214	6.5067	8.1696	11.9898
79	0.4651	1.6789	2.7680	3.9353	6.8426	2.4561	4.8191	6.5029	8.1644	11.9817
80	0.4650	1.6784	2.7672	3.9341	6.8400	2.4550	4.8169	6.4994	8.1597	11.9737
χ^2_k	0.4549	1.6424	2.7055	3.8415	6.6349	2.3660	4.6416	6.2514	7.8147	11.345