GOODNESS-OF-FIT TEST AND BILINEAR MODEL

A Thesis
Presented to
The Academic Faculty

by

Huijun Feng

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Mathematics

Georgia Institute of Technology
May 2013

GOODNESS-OF-FIT TEST AND BILINEAR MODEL

Approved by:

Professor Liang Peng,
Committee Chair
School of Mathematics
Georgia Institute of Technology

Professor Liang Peng, Advisor
School of Mathematics
Georgia Institute of Technology

Professor Vladimir Koltchinskii
School of Mathematics
Georgia Institute of Technology

Professor Yichuan Zhao
Department of Mathematics and Statistics
Georgia State University

Professor Shijie Deng
School of Industrial and Systems Engineering
Georgia Institute of Technology

Professor Xiaoming Huo
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: 4 December 2012

To Linwei, my husband

Dianyong Feng and Yan Xue, my parents

ACKNOWLEDGEMENTS

Atlanta is a lovely city with a great climate, beautiful views, enthusiastic people, a long history, Martin Luther King Jr's birthplace, and the setting of my favorite novel "*Gone with the Wind*". Georgia Insititute of Technology, a top research university with numerous brilliant researchers, is located in this gorgeous city. I have enjoyed living and studying here. During the past four years, I made a lot of friends, learned much, and grew up. What's more, my husband and I met, fell in love, and married while at Georgia Tech. These four years of graduate study are fated to be the most memorable time in my life.

During my study here, the most person to whom I owe the most thanks is my advisor, Professor Liang Peng, a great statistician, an excellent researcher, and a respected advisor. Professor. Peng gave me a flexible environment in which to tackle interesting projects. Without his patience, guidance, kindness and support, I could not have completed this thesis and found my career goal. To me, Professor. Peng is more than just an academic advisor; he has been as supportive as a father.

Now, please allow me to thank Professors. Shijie Deng, Xiaoming Huo, Vladimir Koltchinskii, and Yichuan Zhao for being my oral exam and dissertation committee members. I would also like to thank Professor Luca Dieci, the former graduate coordinator, who offered me the precious opportunity to join the warm family in the School of Mathematics. It was great to work with Ms. Christy Dalton, Karen Hinds, Jan Lewis, Sharon MacDowell, Annette Rohrs, and Genola Turner, our School of Mathematics lovely, helpful and friendly staff members. Thank you all!

School of Math cares a lot for graduate students and international students. I would like to express my sincere gratitude to Ms. Cathy Jacobson for her help with my English. It's also my pleasure to thank the IT group led by Professor Lew Lefton and Justin Filoseta for their expertise and help.

Georgia Tech offers a wonderful platform for students to pursue their professional careers through many career fairs and the famous CO-OP program. Thanks go to Mr. Ken Little for his help in the Graduate CO-OP program, and thanks so much to those people who work behind the scene. Moreover, thank to Mazliah Analytics for giving me my first internship, and thanks to the president, Mr. Rich Saleh and to the senior consultant, Mr. Hongyi Qu, for teaching me about software and business thinking!

My life in Atlanta is a treasure, which would not have happened without my friends. In particular, Yun Gong and Ruodu Wang who share the same advisor with me, helped me a lot in study and college life. I also would like to thank Professors. Chunling Cao and Fukang Zhu, two former visiting scholars from Jilin University, China for their sharing their experience and research. Thanks is also due Professor. Fukang Zhu for his research collaboration.

I now want to thank my dearest friends: Jingfang Liu, Tianjun Ye, Lili Hu, Qingqing Liu, Xiaolin Wang, Tianyu Ye, Dong Xia, Wuchen Li, Chenliang Zhang, Ruidong Wang, Yunlong He, Jun Lu, Chun-hung Liu, Ke Yin, Emma Cohen, Sarah Fletcher, Christopher Pryby, Huy Huynh, James Conway, Alan Diaz, Fabio Difonzo, and the many others I may have failed to mention for their friendship and support. I am also very thankful that many other friends sent their care and love to me from places far away from Atlanta.

Most importantly, I want to thank my beloved husband, Linwei Xin for his selfless love, my parents for showing the world to me and bringing me up, and my grandparents for their care. Without their support throughout my life, I could not have achieved this goal. Becoming a Ph.D is a new start for me. Tomorrow is a new day! With your love and hope, I will live better and work harder. Thank you!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

The Empirical Likelihood method was introduced by A. B. Owen to test hypotheses in the early 1990s. It's a nonparametric method and uses the data directly to do statistical tests and to compute confidence intervals/regions. Because of its distribution free property and generality, it has been studied extensively and employed widely in statistical topics. There are many classical test statistics such as the Cramer-von Mises test statistic, the Anderson-Darling test statistic, and the Watson test statistic, to name a few. However, none is universally most powerful. This thesis is dedicated to extending the Empirical Likelihood Method to several interesting statistical topics in hypothesis tests. First of all, we focus on testing the fit of distributions. Based on the Cramer-von Mises test statistic, we propose a novel Jackknife Empirical Likelihood test via estimating equations in testing the goodness-of-fit. The proposed new test allows one to add more relevant constraints so as to improve the power. Also, this idea can be generalized to other classical test statistics. Second, when aiming at testing the error distributions generated from a statistical model (e.g., the regression model), we introduce the Jackknife Empirical Likelihood idea to the regression model, and further compute the confidence regions with the merits of distribution free limiting chi-square property. Third, the empirical likelihood method based on some weighted score equations are proposed for constructing confidence intervals for the coefficient in the simple bilinear model. The effectiveness of all presented methods (rooted in the Empirical Likelihood Method) are demonstrated by some extensive simulation studies.

CHAPTER I

INTRODUCTION

Goodness-of-fit studies how well a distribution or a family of distributions fits the data. It often shows up in the format of hypothesis test. The null hypothesis H_0 is that a random variable X follows a distribution $F(x)$ or a family of distributions $F(x;\theta)$. The alternative hypothesis H_a states that H_0 is false. Goodness-of-fit techniques measure the discrepancy between the observed sample data and the objective distribution. Many researchers have proposed kinds of test statistics to perform this goal, for example, Pearson's chi-square tests, Yate's chi-square tests, Kolmogorov-Smirnov Test, Cramer-Von Miser Test, and so on. These tests are generally divided into three categories: Chi-Square tests, Tests based on Empirical Distribution Function (EDF), and Tests based on Regression and Correlation.

Empirical Likelihood, introduced by Art B. Owen in 1988, is a non-parametric technique based on the EDF and likelihood ratio. It avoids picking up parametric distribution families for the data, and inherits the advantage of EDF. Following likelihood ratio test, Wilks theorem holds and gives Empirical Likelihood Test distribution free property. This can be used to test not only the hypothesis about the parameters, but also the goodness-of-fit. The popularity of Empirical Likelihood is mainly due to its efficiency and flexibility. Although Empirical Likelihood Method (ELM) has been extensively studied, there are still numerous questions to be solved. Note that ELM is not omnipotent. In some cases, it fails. This dissertation addresses three extensions of Empirical Likelihood Method (ELM).

Chapter II proposes a new method to test the goodness-of-fit problem. This method is based on Cramer-Von Mises Test and Empirical Likelihood Test. In order to increase the test power, we add the estimating equations in testing. Since ELM fails for nonlinear functional statistics, we introduce Jackknife's idea to overcome this problem.

Chapter III is dedicated to test whether the error term of a regression model follows some kind of distributions. More specifically, we propose a Jackknife Empirical Likelihood Method based on the regression model, EDF of the regression errors, and ELM to do the test.

Chapter IV is the application of Profile Empirical Likelihood Method to a simple bilinear time series. With employment of Conditional Least Square Estimator, weighted score equations and Profile ELM's idea, we construct a confidence region for the parameters without assuming the normality of the errors and estimating the asymptotic variance.

This Chapter is served as an introduction to the concepts and tools we are going to use later.

1.1 Chi-Squared Tests

Pearson in 1900 invented his Chi-Squared Test for fitting the data to a multinomial setting. Later on many researchers have developed Chi-Squared Tests for solving other questions. More details can be found in [7] and [40].

1.1.1 Pearson Chi-Squared Test

To test the simple null hypothesis H_0: a random sample $X_1, \cdots, X_n$ has distribution function F(x), Pearson first partitioned the data into M cells. Second, he computed $N_i = \sum_j 1_{\{X_j \in ith \text{ cell}\}}$. Third, compute

$$p_i = \int_{ith \text{ cell}} dF(x).$$

Finally, the Test statistic is defined as $X^2 = \sum_{i=1}^{M} \frac{(N_i - np_i)^2}{np_i}$. It is proved that X^2 goes to a chi-square distribution with degrees of freedom (df) $M-1$, as n increases. When we specify a significant level α, if p-value is smaller than α, we could reject the null hypothesis.

However, in many cases, the null hypothesis is a composite one. Instead of testing the random sample follows a specific distribution (discrete or continuous), we may only know that the random sample probably comes from a distribution family $\{F(\cdot|\theta) : \theta \in \Omega\}$. In this case, Pearson proposed to estimate θ by $\tilde{\theta}_n$, a function of $X_1, \cdots, X_n$, and then tested whether the data fits $F(\cdot|\tilde{\theta}_n)$. Thus,

$$p_i(\tilde{\theta}_n) = \int_{ith \text{ cell}} dF(x|\tilde{\theta}_n),$$

$$X^2(\tilde{\theta}_n) = \sum_{i=1}^{M} \frac{(N_i - np_i(\tilde{\theta}_n))^2}{np_i(\tilde{\theta}_n)}.$$

Unfortunately the limit of $X^2(\tilde{\theta}_n)$ is not $\chi^2(M-1)$ any more.

Later, Fisher used the minimum chi-squared estimator $\bar{\theta}_n$, which is the solution of

$$\sum_{i=1}^{M} \{\frac{N_i}{p_i(\theta)}\}^2 \frac{\partial p_i(\theta)}{\partial \theta_k} = 0, \quad k = 1, \cdots, p$$

Then Pearson-Fisher statistic $X^2(\bar{\theta}_n)$ approximates $\chi^2(M-p-1)$ distribution under the null hypothesis.

In 1949, Neyman proposed the modified chi-squared statistic

$$X_M^2 = \sum_{i=1}^{M} \frac{[N_i - np_i(\theta)]^2}{N_i},$$

and θ can be estimated by solving

$$\sum_{i=1}^{M} \frac{p_i(\theta)}{N_i} \frac{\partial p_i(\theta)}{\partial \theta_k} = 0, \quad k = 1, \cdots, p.$$

Neyman method's advantage is that the equation above is more likely solvable in a closed form than Pearson's statistic and Fisher's one.

1.1.2 General Chi-Squared Test

Rao and Robson (1974) put up a more powerful test statistic, which is taken as a standard chi-squared test of fit. They use random cells in their procedure. Denote $E_i(X_1, \cdots, X_n)$ as the i-th random cell and N_i is still the frequency in E_i. Under null hypothesis $F(\cdot|\theta)$,

$$p_i = \int_{E_i} dF(x|\theta), i = 1, \cdots, M.$$

Let $V_n(\theta) = (\frac{[N_i - np_i(\theta)]}{(np_i(\theta))^{1/2}})_1^M$, a M-dimensional vector. Define $Q_n = Q_n(X_1, \cdots, X_n)$ a M-dim symmetric nonnegative definite matrix. Then, the general test statistic is defined as

$$V_n(\theta)' Q_n V_n(\theta),$$

where θ should be replaced by its estimator. Rao and Robson used MLE $\hat{\theta}_n$ and proved that the Rao-Robson statistic

$$R_n = V_n(\hat{\theta}_n)' Q_n(\hat{\theta}_n) V_n(\hat{\theta}_n)$$

has the $\chi^2(M-1)$ limiting distribution under the null hypothesis. Moreover, the limiting matrix of $Q_n(\hat{\theta}_n)$ is

$$Q(\theta) = I_M + B(\theta)[J(\theta) - B(\theta)'B(\theta)]^{-1}B(\theta)',$$

where $J(\theta)$ is the p-dim Fisher Information matrix for $F(\cdot|\theta)$, and $B(\theta)$ is a $M \times p$ matrix with the (i,j)th entry $p_i(\theta)^{-1/2}\frac{\partial p_i(\theta)}{\partial \theta_j}$. Actually, R_n equals to the Pearson chi-squared test plus an extra term:

$$R_n = X^2(\hat{\theta}_n) + (V_n'B)(J - B'B)^{-1}(V_n'B)'.$$

There are many test statistics based on Chi-Squared Tests, which are more flexible in many situations than Pearson's original one. The merit of Chi-Squared Tests is distribution free limiting chi-square null distribution, and so the critical values can be obtained easily from the chi-square distribution's table. However, generally it's not as flexible and powerful as the test statistics based on Empirical Distribution Function (EDF).

1.2 Tests based on EDF

EDF tests don't have any assumption on random samples. They only use the information derived from the sample data. Suppose $X_1, X_2, \cdots, X_n$ is a random sample from some distribution (in most cases, we assume it's continuous in this section), the definition of EDF is

$$F_n(x) = \frac{\text{the number of observations} \leq x}{n}, -\infty < x < \infty,$$

so, $F_n(x)$ is a step function. Sort the random sample and denote the order statistics as $X_{(1)} < X_{(2)} < \cdots < X_{(n)}$, then we have

$$F_n(x) = \begin{cases} 0 & x < X_{(1)} \\ \frac{1}{n} & X_{(i)} \leq x < X_{(i+1)} \quad i = 1, \cdots, n-1 \\ 1 & X_{(n)} \leq x. \end{cases}$$

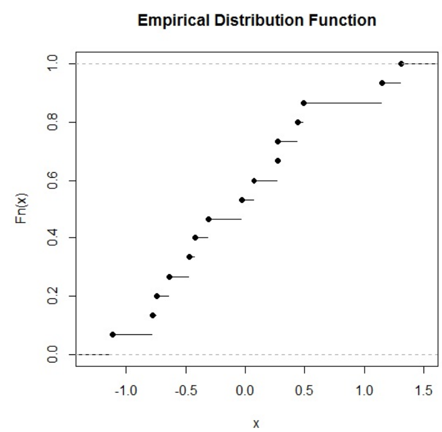

Figure 1: EDF for 15 observation from Normal Distribution

1.2.1 EDF Statistics

Kolmogorov-Smirnov Test

To test the null hypothesis that data is from a distribution $F(x)$, Kolmogorov measured the largest difference between EDF $F_n(x)$ and $F(x)$ defined as

$$D_n = \sup_x |F_n(x) - F(x)|.$$

Kolmogorov showed that under null hypothesis $\sqrt{n}D_n \xrightarrow{d} \sup_t |B(F(t))|$, where $B(t)$ is a Brownian bridge. He also provided a table for the Kolmogorov distribution

$$K = \sup_t |B(t)|.$$

When $F(x)$ is continuous, then reject H_0 with significant level α if $\sqrt{n}D_n > K_\alpha$.

Cramer-von Mises Test

To test whether $F(x)$ is the cdf of the data, Cramer and Von Mises gave an alternative test to K-S test. Instead of looking at the maximum difference between the empirical distribution function and the test function, they used the expectation of the squared of the difference as

$$W^2 = \int_{-\infty}^{\infty} (F_n(x) - F(x))^2 dF(x).$$

If $X_1, X_2, \cdots, X_n$ is the random sample and $X_{[1]} \leq X_{[2]} \leq \cdots \leq X_{[n]}$ are the order statistics, then the CM test statistic is

$$T = nW^2 = \frac{1}{12n} + \sum_{i=1}^{n} [\frac{2i-1}{2n} - F(X_{[i]})]^2.$$

Cramer and Von Mises also gave a table. If under a specific significant level, T's value is larger than the critical value, then the null hypothesis is rejected.

Watson Test

An improved test of Cramer-von Mises test is the Watson test defined as

$$U^2 = T - n(\bar{F} - \frac{1}{2})^2,$$

where

$$\bar{F} = \frac{1}{n} \sum F(X_i).$$

Anderson-Darling Test

Suppose $F(x)$ is the distribution function we want to test and $F_n(x)$ is the EDF. Anderson-Darling test is defined as

$$A^2 = n \int_{-\infty}^{\infty} \frac{(F_n(x) - F(x))^2}{[F(x)(1 - F(x))]} dF(x).$$

If we denote $w(x) = [F(x)(1 - F(x))]^{-1}$, then Anderson-Darling test can be written as $A^2 = n\int_{-\infty}^{\infty}(F_n(x) - F(x))^2 w(x) dF(x)$.

We can see that Cramer-von Mises Test is the above statistics with $w(x) = 1$. Thus, we can view Anderson-Darling test is a weighted CM test, which gives more weight to the tails of the underlying distribution.

1.2.2 How to compute

After we have the sample data and the hypothesis, how will we fulfill the test by these statistics? The computation of the statistics is done by the Probability Integral Thansformation: $Z = F(X)$. If $F(x)$ is the true distribution, then Z is a uniform random variable on $[0, 1]$. Suppose we have the random sample $X_1, X_2, \ldots, X_n$, then we can compute $Z_i = F(X_i), i = 1, 2, \ldots, n$. Denote the cdf of uniform distribution by F^*, and let F_n^* be the EDF of $Z_i's$, that is, $F_n^*(z) = \frac{1}{n}\sum_{i=1}^n 1_{Z_i \le z}$. Thus,

$$F_n(X_i) - F(X_i) = F_n^*(Z_i) - F^*(Z_i) = F_n^*(Z_i) - Z_i.$$

We sort $Z_1, Z_2, \ldots, Z_n$ by ascending order $Z_{(1)} \le Z_{(2)} \le \cdots \le Z_{(n)}$. With employment of Z_i's, the K-S statistic is

$$D_n = \max_{1 \le i \le n} |F_n(X_i) - F(X_i)| = \max\{D^+, D^-\}$$

where $D^+ = \max_{1 \le i \le n}\{F_n(X_i) - F(X_i)\} = \max_i\{F_n^*(Z_i) - Z_i\} = \max_i\{\frac{i}{n} - Z_{(i)}\}$, and $D^- = \max_{1 \le i \le n}\{F(X_i) - F_n(X_i)\} = \max_i\{Z_i - F_n^*(Z_i)\} = \max_i\{Z_{(i)} - \frac{i-1}{n}\}$.

Kuiper in 1960 proposed to use $V = D^+ + D^-$ to test the observations on a circle.

By using $Z'_{(i)}s$, CM test statistic becomes

$$T = \frac{1}{12n} + \sum_{i=1}^{n}[\frac{2i-1}{2n} - Z_{(i)}]^2$$

and

$$U^2 = T - n(\bar{Z} - \frac{1}{2})^2$$

where $\bar{Z} = \frac{1}{n}\sum Z_i$.

The Anderson-Darling test statistic becomes

$$A^2 = -n - \frac{1}{n}\sum_i (2i-1)[\log Z_{(i)} + \log(1 - Z_{(n-i+1)})]$$

Hence, we can easily compute these test statistics by their numeric expressions.

1.3 Tests based on Regression Models

Suppose the covariates are $X_1, X_2, \ldots, X_n$ which are p-dim vectors, and the corresponding responses are $Y_1, Y_2, \ldots, Y_n$. We want to fit a regression model

$$Y = \alpha + \beta T + \epsilon = \alpha + \beta g(X) + \epsilon.$$

If we order the responses $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$, and rewrite our model to

$$Y_{(i)} = \alpha + \beta T_i + \epsilon_i$$

Then T_i has another expression besides a function of indicators. If Y satisfies a continuous distribution F_0, $F_0(x) = F(u), u = \frac{x-\alpha}{\beta}$, and its inverse function exists, then $T_i = F^{-1}(\frac{i}{n+1})$.

There are three approaches to test the goodness-of-fit of this regression model:

(a) Tests based on $R(Y, T)$ which are the correlation coefficients between the response and T.

(b) Tests based on residues $\epsilon_i = Y_{(i)} - \hat{Y}_{(i)}$, where $\hat{Y}_{(i)} = \alpha + \beta T_i$. However, we need to estimate α and β in the regression, so $\hat{Y}_{(i)}$ should be $\hat{\alpha} + \hat{\beta} T_i$ where $\hat{\alpha}$ and $\hat{\beta}$

are two estimators of α and β by some method, for example, minimizing least squared error, etc.

(c) We get the estimator $\hat{\beta}$ of β, and compare $\hat{\beta}^2$ with some other estimate method. This is also an idea for the measuring the fit.

Here we look at the test method in category (b), which is the most common idea in goodness-of-fit test for regression models (More test methods can be found in [7]).

1.3.1 F-test based on Regression Residuals

Denote $m_i = E[g(X_{(i)})]$, where $X_{(i)}$ is the indicators corresponding to $Y_{(i)}$. Then $E[Y_{(i)}] = \alpha + \beta m_i$. Let $v_{ij} = E[(g(X_{(i)}) - m_i)^T(g(X_{(j)}) - m_j)]$ be the covariance of $g(X_{(i)})$ and $g(X_{(j)})$, $Y = (Y_{(1)}, Y_{(2)}, \ldots, Y_{(n)})^T$, $m = (m_1, m_2, \ldots, m_n)^T$, and $V = (v_{ij})_{i,j}$, then the generalized least squares estimates of α and β are

$$\hat{\alpha} = -m^T GX, \quad \hat{\beta} = 1^T GX,$$

where $G = \frac{V^{-1}(1m^T - m1^T)V^{-1}}{(1^TV^{-1}1)(m^TV^{-1}m)-(1^TV^{-1}m)^2}$.

Based on the ANOVA table, the test for fit is

$$Z_1(Y, X) = SS_{residule}/SS_{total}$$

where $SS_{residual} = \sum_i (Y_i - \hat{Y}_i)^2$ and $SS_{total} = \sum_i (Y_i - \bar{Y})^2$. The test statistic satisfies F distribution under the null hypothesis and some regularity conditions.

The generalized least squares employment gives the test

$$Z_2(Y, X) = SS_{residual}/SS_{total}$$

where $SS_{residual} = (Y - \hat{Y})^T V^{-1} (Y - \hat{Y})$.

1.3.2 Likelihood Ratio Test based on Regression models

In many cases, we compare two regression models to see which one has a better fit. In this case, we can use likelihood ratio test:

$$D = -2 \log \frac{\text{likelihood for null model}}{\text{likelihood for alternative model}}.$$

When the null model is a special case of the alternative model, the test statistic is approximately a chi-squared distribution. That is, the asymptotic distribution of the test statistic is distribution free.

1.4 Empirical Likelihood Method

Empirical Likelihood Method is introduced by A.B. Owen (1988, 1990), which can be used to construct confidence intervals and to test hypotheses. When it applies to testing hypothesis, it's a combination of the Chi-Squared Test and EDF test. Moreover, it can also be used to test the hypothesis based on Regression Models. Due to its flexibility and effectiveness, it has been quickly extended and studied.

Empirical likelihood (ELM) test has asymptotic chi-square distribution, so the test statistic is distribution free, which avoids the bootstrap step to get the critical values. ELM is also a method based on EDF, which doesn't require extra information but the data. Our research focuses on the application and improvement of ELM.

1.4.1 Basic concepts about ELM

Definition 1.1. *Let $X_1, \ldots, X_n \in R^d$ be a random sample from a common distribution function, the nonparametric likelihood function is defined as, for any distribution* *function F,*

$$L(F|X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P_F(X_i),$$

where $P_F(X_i)$ *is the probability of getting the observation* X_i *under the cdf F.*

Since $L(F_n|X_1, X_2, \ldots, X_n) = n^{-n}$, if $F \neq F_n$, then

$$L(F|X_1, X_2, \ldots, X_n) < L(F_n|X_1, X_2, \ldots, X_n). \tag{1}$$

Write $X = (X_1, X_2, \cdots, X_n)^T$. Now, consider X is from $F \in \{F_\theta : \theta \in \Theta\}, \Theta \subset R^k$. So X is from a parametric distribution family, and Θ is a vector space. Suppose the pdf f_θ of X exists, and we want to test

$$H_0 : \theta \in \Theta_0 \quad v.s. \quad H_a : \theta \in \Theta_1 \tag{2}$$

where $\Theta_0 \cup \Theta_1 = \Theta$ and $\Theta_0 \cap \Theta_1 = \emptyset$.

Definition 1.2. *Let* $l(\theta) = f_\theta(X)$ *be the likelihood function of random sample X. For testing (2), a likelihood ratio (LR) test is a test that rejects* H_0 *if and only if* $\lambda(X) < c$*, where* $c \in [0, 1]$ *and* $\lambda(X)$ *is the likelihood ratio defined by*

$$\lambda(X) = \frac{\sup_{\theta \in \Theta_0} l(\theta)}{\sup_{\theta \in \Theta} l(\theta)}.$$

Thus, if $\hat{\theta}$ is an MLE of θ and $\hat{\theta}_0$ is an MLE of θ subject to $\theta \in \Theta_0$, then

$$\lambda(X) = l(\hat{\theta}_0)/l(\hat{\theta}).$$

The Wilks' theorem (Wilks 1938) states that under some regularity conditions and the null hypothesis,

$$-2 \log \lambda(X) \xrightarrow{d} \chi_q^2$$

where $q = dim(\Theta) - dim(\Theta_0)$. When the null hypothesis H_0 is simple, i.e., $H_0 : \theta = \theta_0$, we have $q = dim(\Theta)$. Using this good property, we can make rejection region $\lambda < e^{-\chi^2_{q,\alpha}/2}$ when n is sufficiently large, where $\chi^2_{q,\alpha}$ is the $(1-\alpha)$th quantile of the chi-square distribution χ^2_q. Hence we reject H_0 if the test statistic falls in the rejection region. Besides, we can also use the asymptotic property of -2 log-likelihood ratio to construct confidence regions for the parameters.

As we can see, the Likelihood Ratio Test is useful in parametric models. The corresponding nonparametric method is called Empirical Likelihood Method (ELM), which is based on EDF.

1.4.2 Empirical Likelihood Ratio Test

(1) tells us that F_n maximizes the nonparametric likelihood function $L(F|X)$ over $p_i > 0$ and $\sum_i p_i = 1$, where $p_i = P_F(X_i)$. (Owen 1988, 2001) pointed out that we could do some modification on the nonparametric likelihood, and the modification of the likelihood is called empirical likelihood. By maximizing an empirical likelihood, we can get an estimator of the distribution F, and this estimator is called maximum empirical likelihood estimator (MELE).

Now let's look at the definition of Empirical Likelihood Ratio Test from [40]. Consider that we are interested in a functional of cdf F, and we want to test

$$H_0 : T(F) = \theta_0 \quad v.s. \quad H_1 : T(F) \neq \theta_0$$

Let $l(F), F \in \mathcal{F}$, where $\mathcal{F}$ is a class of cdf's on R^d, be a given empirical likelihood, $\hat{F}$ be an MELE of F, and $\hat{F}_{H_0}$ be an MELE of F under H_0, i.e., $\hat{F}_{H_0}$ is an MELE of F subject to $T(F) = \theta_0$. Then the empirical likelihood ratio is defined as

$$\lambda_n(X) = l(\hat{F}_{H_0})/l(\hat{F}).$$

A test with rejection region $\lambda_n(X) < c$ is called an empirical likelihood ratio test. Now let's look at a particular case: $X = (X_1, X_2, \cdots, X_n)^T$, where $X_i \in R^d$ is a random sample from cdf $F(\cdot)$. In this case, the empirical likelihood is defined as

$$l(F) = \prod_{i=1}^{n} p_i \text{ subject to } p_i \geq 0, \sum_{i=1}^{n} p_i = 1, \tag{3}$$

where $p_i = P_F(X_i), i = 1, 2, \ldots, n$. We want to test

$$H_0 : F \in \mathcal{F}_0 \quad v.s. \quad H_a : F \in \mathcal{F}, F \notin \mathcal{F}_0$$

where $\mathcal{F}_0 \subset \mathcal{F}$ and $\mathcal{F}$ is the set of all distribution functions on R^d. Then our empirical likelihood ratio function is

$$\begin{aligned}
\lambda_n(X) &= \frac{\sup\{l(F|X) : F \in \mathcal{F}_0\}}{\sup\{l(F|X) : F \in \mathcal{F}\}} \\
&= \frac{\sup\{l(F|X) : F \in \mathcal{F}_0\}}{l(F_n|X)} \\
&= \sup_{F \in \mathcal{F}_0} \{\prod_{i=1}^{n}(np_i) : p_i = P_F(X_i) \geq 0, \sum_{i=1}^{n} p_i = 1\}
\end{aligned}$$

Under some regularity conditions, Owen (1988, 1990) proved that Wilks' Theorem still holds for some situations, for example, $T(F)$ is a linear functional of F. Particularly, when $T(F) = E(X_i) = \theta$, the empirical likelihood ratio is

$$R(\theta) = \sup\{\prod_{i=1}^{n}(np_i) : \ p_1 \geq 0, \cdots, p_n \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i X_i = \theta\}, \tag{4}$$

and Owen (1988) showed that:

Theorem 1. *Let $X_1, \ldots, X_n$ be independent random vectors in R^p with common distribution F_0 having mean μ_0 and finite covariance matrix V_0 of rank $q > 0$. Then $l(\mu_0)$ converges in distribution to a χ^2_q random variable as $n \to \infty$, where $l(\mu_0) = -2 \log R(\mu_0)$.*

Qin and Lawless (1994) introduced estimating equations to ELM, and they proved that Wilks' theorem still holds under some regularity conditions. More interestingly, the ELM with estimating equations has better power than the original ELM test in some cases. For example, if we have more information about the distribution and parameter $E[g(X_1; \theta)] = 0$, where $E[g(\cdot)]$ is a d-dimensional linear functional of the underlying distribution, then we could add the estimating equations $E[g(\cdot)]$ in our empirical likelihood ratio and then maximizing it. Our EL function for the parameter θ becomes

$$L(\theta) = \sup\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i g(X_i, \theta) = 0, p_i \geq 0, \sum_{i=1}^{n} p_i = 1\}. \tag{5}$$

If the null hypothesis is $H_0 : \theta = \theta_0$ and $\hat{\theta}$ maximizes (5), then Qin and Lawless (1994) showed that under some regularity conditions, $-2\log(L(\theta_0)/L(\hat{\theta})) \xrightarrow{d} \chi^2_r$, where $r = d \vee q$ and $\theta \in R^q$.

1.4.3 Profile Empirical Likelihood

$X_1, X_2, \ldots, X_n$ are a random sample from the d-dim distribution $F_0(\cdot|\theta)$ and $\theta = (\vartheta, \varphi)$, where ϑ is a r-dim vector and φ is a $(p-r)$-dim vector. Moreover, we know one more information $E[g(X_1, \theta)] = 0$, and we are going to test

$$H_0 : \vartheta = \vartheta_0 \quad v.s. \quad H_a : \vartheta \neq \vartheta_0.$$

Hence, we are only interested in the first r-dim of the whole parameter.

In this case, the EL is

$$L(\theta) = \sup\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i g(X_i, \theta) = 0, p_i \geq 0, \sum_{i=1}^{n} p_i = 1\}.$$

We could use *emplik* in R to solve the above optimization problem, and let $\hat{\theta}$ be the solution. And let $\hat{\varphi}$ be a maximum solution of $L(\varphi) = L(\vartheta_0, \varphi)$. The profile empirical likelihood ratio is

$$\lambda_n(X) = \prod_{i=1}^{n} \frac{1 + [\xi_n(\hat{\theta})]^T g(X_i; \hat{\theta})}{1 + [\varsigma_n(\vartheta_0, \hat{\varphi})]^T g(X_i; \vartheta_0, \hat{\varphi})}$$

where $\xi_n(\hat{\theta})$ satisfies

$$\sum_{i=1}^{n} \frac{g(X_i; \hat{\theta})}{1 + [\xi_n(\hat{\theta})]^T g(X_i; \hat{\theta})} = 0,$$

and $\varsigma_n(\vartheta_0, \hat{\varphi})$ satisfies

$$\sum_{i=1}^{n} \frac{g(X_i; \vartheta_0, \hat{\varphi})}{1 + [\varsigma_n(\vartheta_0, \hat{\varphi})]^T g(X_i; \vartheta_0, \hat{\varphi})} = 0.$$

Qin and Lawless (1994) showed that Wilks' Theorem holds with the profile empirical likelihood ratio $\lambda_n(X)$ above under some regularity conditions. Therefore, we could make the rejection decision when $\lambda_n(X) < c$, where c is the corresponding chi-square distribution's α significant level critical value.

All above optimization problems can be solved by the Lagrange Multiplier Method. More details about Empirical Likelihood Ratio test and Profile Empirical Likelihood Ratio Test can be found in Owen (1988, 1990, 2001), Chen and Qin (1993), Qin (1993), and Qin and Lawless (1994).

CHAPTER II

JACKKNIFE EMPIRICAL LIKELIHOOD TESTS FOR DISTRIBUTION FUNCTIONS

This Chapter is based on the published paper:

H. Feng and L. Peng (2012). Jackknife empirical likelihood tests for distribution functions. JSPI 142, 1571–1585.

2.1 Introduction

Suppose $X_1, \cdots, X_n$ are independent and identically distributed random variables with distribution function F. Let F_0 denote a specified distribution function. Testing the hypothesis $H_0 : F \equiv F_0$ has been studied extensively in the literature. Some well-known testing procedures are based on the distance between F_0 and the empirical distribution $F_n(x) = n^{-1} \sum_{i=1}^n I(X_i \le x)$ such as $\sup_x |F_n(x) - F_0(x)|$ (Kolmogorov test statistic), $n \int \{F_n(x) - F_0(x)\}^2 \, dF_0(x)$ (Cramer-von Mises test statistic), $n \int \{F_n(x) - F_0(x)\}^2 F_0^{-1}(x)\{1 - F_0(x)\}^{-1} \, dF_0(x)$ (Anderson-Darling test statistic), $n \int \{F_n(x) - F_0(x) - \int (F_n(x) - F_0(x)) \, dF_0(x)\}^2 \, dF_0(x)$ (Watson test statistic); see Chapter 4 of D'Agostino and Stephens (1986) for details. A more general testing procedure via phi-divergence was proposed by Jager and Wellner (2007) which still measures some type of distance between $F_n(x)$ and $F_0(x)$ and includes Cramer-von Mises type of statistics and the test statistic proposed by Berk and Jones (1979) as special cases. Recently, Einmahl and McKeague (2003) proposed a local empirical likelihood method, which results in the same test statistic of Berk and Jones (1979) when the method is applied to testing $H_0 : F \equiv F_0$.

Here we focus on the Cramer-von Mises test. By writing

$$\int \{F(x)-F_0(x)\}^2 dF_0(x) = \int \{F(x)-F_0(x)\}F(x)dF_0(x) - \int \{F(x)-F_0(x)\}F_0(x)dF_0(x)$$

it is easy to verify that $F \equiv F_0$ is equivalent to

$$\int \{F(x) - F_0(x)\}F(x)\, dF_0(x) = 0 \quad \text{and} \quad \int \{F(x) - F_0(x)\}F_0(x)\, dF_0(x) = 0,$$

which is equivalent to

$$\int F^2(x)\, dF_0(x) = 1/3 \quad \text{and} \quad \int F(x)F_0(x)\, dF_0(x) = 1/3. \tag{6}$$

Therefore, one can test $H_0 : F(x) \equiv F_0(x)$ via estimating equations (6). Unfortunately, under H_0, the joint asymptotic limit of the nonparametric estimators of $\int F^2(x)\, dF_0(x) - 1/3$ and $\int F(x)F_0(x)\, dF_0(x) - 1/3$ is degenerate, i.e., the asymptotic covariance matrix of

$$\sqrt{n}\left(\int F_n^2(x)\, dF_0(x) - 1/3, \int F_n(x)F_0(x)\, dF_0(x) - 1/3\right)$$

has rank one rather than two. In order to overcome this difficulty, we propose to consider either

$$\int F^2(x)\, dF_0(x) = \frac{1}{3} \quad \text{and} \quad \int F(x)F_0(x)\, dF_0(x) + \int \{1 - F_0^2(x)\}^{1/2}\, dF(x) = \frac{1}{3} + \frac{\pi}{4} \tag{7}$$

or

$$\int F^2(x)\, dF_0(x) = \frac{1}{3} \quad \text{and} \quad \int F(x)F_0(x)\, dF_0(x) - \int \{1 - F_0^2(x)\}^{1/2}\, dF(x) = \frac{1}{3} - \frac{\pi}{4}. \tag{8}$$

Note that $\int\{1 - F_0^2(x)\}^{1/2}\, dF(x) = E\{1 - F_0^2(X_i)\}^{1/2} = \pi/4$ was employed by Stephens (1966) to test $H_0 : F \equiv F_0$. Although $H_0 : F \equiv F_0$ is neither equivalent to (7) nor equivalent to (8), rejecting H_0 does imply that at least one of them is not true.

In this chapter, we first propose novel empirical likelihood tests via (7) or (8) to test a simple null hypothesis. Extension to testing composite null hypothesis is given too. An important feature of the proposed new methods is to allow more relevant constraints to be included freely. Since Owen (1988, 1990) introduced the empirical likelihood method for constructing confidence intervals for a mean, it has been extended and applied to many different fields; see Owen (2001) for details. Applying the empirical likelihood method to estimating equations was studied by Qin and Lawless (1994).

We organize this chapter as follows. Section 2 presents the new methodologies and theoretical results. A simulation comparison is given in Section 3. All proofs are put in Section 4.

2.2 Methodologies

Throughout we assume that $X_1 = (X_{11}, \cdots, X_{1d})^T, \cdots, X_n = (X_{n1}, \cdots, X_{nd})^T$ are independent and identically distributed random vectors with distribution function F. We also simply write $\int \cdots \int F^2(x)\, dF_0(x)$ as $\int F^2(x)\, dF_0(x)$ for $x = (x_1, \cdots, x_d)^T$.

2.2.1 Simple null hypothesis.

In this subsection, we are interested in testing $H_0 : F \equiv F_0$ against $H_a : F \not\equiv F_0$, where F_0 is a given distribution function. Let $F_i's$ and $F_{0i}'s$ denote the marginal distributions of F and F_0, respectively. Based on the Cramer-von Mises type of distance $\int\{F(x)-F_0(x)\}^2\,dF_0(x)$, we propose to construct tests via

$$\begin{cases} \int F^2(x)\,dF_0(x) = \int F_0^2(x)\,dF_0(x) \\ \int F(x)F_0(x)\,dF_0(x) + \sum_{l=1}^d \int\{1-F_{0l}^2(x_l)\}^{1/2}\,dF_l(x_l) = \int F_0^2(x)\,dF_0(x) + \frac{d\pi}{4} \end{cases} \tag{9}$$

or

$$\begin{cases} \int F^2(x)\,dF_0(x) = \int F_0^2(x)\,dF_0(x) \\ \int F(x)F_0(x)\,dF_0(x) - \sum_{l=1}^d \int\{1-F_{0l}^2(x_l)\}^{1/2}\,dF_l(x_l) = \int F_0^2(x)\,dF_0(x) - \frac{d\pi}{4}, \end{cases} \tag{10}$$

which are extensions of (7) and (8) to multivariate distribution functions.

Since equation $\int F^2(x)\,dF_0(x) = \int F_0^2(x)\,dF_0(x)$ is nonlinear, we can not directly apply the empirical likelihood method to either equations (9) or equations (10). Recently, Jing, Yuan and Zhou (2009) proposed a so-called jackknife empirical likelihood method to deal with nonlinear functionals. Here we propose to employ this jackknife empirical likelihood method to either equations (9) or (10) as follows.

For $i=1,\cdots,n$ and $l=1,\cdots,d$, define

$$\begin{cases} F_{n,i}(x) = \frac{1}{n-1}\sum_{j=1,j\neq i}^n I(X_j \le x) \\ F_{n,i}^{(l)}(x_l) = \frac{1}{n-1}\sum_{j=1,j\neq i}^n I(X_{j,l} \le x_l) \\ F_n^{(l)}(x_l) = \frac{1}{n}\sum_{j=1}^n I(X_{j,l} \le x_l). \end{cases}$$

As in Jing, Yuan and Zhou (2009), a jackknife sample based on (9) is defined as $Y_i = (Y_{i,1}, Y_{i,2})^T$ for $i=1,\cdots,n$, where

$$Y_{i,1} = n\int F_n^2(x)\,dF_0(x) - (n-1)\int F_{n,i}^2(x)\,dF_0(x) - \int F_0^2(x)\,dF_0(x)$$

and

$$\begin{aligned} Y_{i,2} &= n\int F_n(x)F_0(x)\,dF_0(x) - (n-1)\int F_{n,i}(x)F_0(x)\,dF_0(x) + \sum_{l=1}^d\{1-F_{0l}^2(X_{i,l})\}^{1/2} \\ &\quad - \int F_0^2(x)\,dF_0(x) - \frac{d\pi}{4}. \end{aligned}$$

Based on this jackknife sample, one can define the jackknife empirical likelihood function as

$$L_s = \sup\{\prod_{i=1}^{n}(np_i) : p_1 \geq 0, \cdots, p_n \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i Y_i = 0\}.$$

By Lagrange multiplier technique, we obtain $p_i = n^{-1}\{1 + \lambda^T Y_i\}^{-1}$ and $-2\log L_s = 2\sum_{i=1}^{n}\log\{1 + \lambda^T Y_i\}$, where λ satisfies $\sum_{i=1}^{n} \frac{Y_i}{1+\lambda^T Y_i} = 0$.

Similarly, we can define the jackknife sample based on (10) as $\bar{Y}_i = (\bar{Y}_{i,1}, \bar{Y}_{i,2})^T$ for $i = 1, \cdots, n$, where

$$\bar{Y}_{i,1} = n\int F_n^2(x)\,dF_0(x) - (n-1)\int F_{n,i}^2(x)\,dF_0(x) - \int F_0^2(x)\,dF_0(x)$$

and

$$\begin{aligned}\bar{Y}_{i,2} &= n\int F_n(x)F_0(x)\,dF_0(x) - (n-1)\int F_{n,i}(x)F_0(x)\,dF_0(x) - \sum_{l=1}^{d}\{1 - F_{0l}^2(X_{i,l})\}^{\frac{1}{2}} \\ &\quad - \int F_0^2(x)\,dF_0(x) + \tfrac{d\pi}{4}.\end{aligned}$$

Based on this jackknife sample, the jackknife empirical likelihood function is defined as

$$\bar{L}_s = \sup\{\prod_{i=1}^{n}(np_i) : p_1 \geq 0, \cdots, p_n \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \bar{Y}_i = 0\}.$$

The following theorem shows that Wilks theorem holds for the above proposed jackknife empirical likelihood methods.

Theorem 2. *Under $H_0 : F \equiv F_0$, both $l_s = -2\log L_s$ and $\bar{l}_s = -2\log \bar{L}_s$ converge in distribution to a chi-square distribution with degrees of freedom two.*

Remark 1. A very appealing feature of the above proposed empirical likelihood tests is that it allows one to add more constraints freely especially when the sample size is large enough. Ideally, adding constraints characterizing the departure from F_0 will improve the test power. On the other hand, having more constraints may reduce the accuracy of the test size and also increases the computational burden.

2.2.2 Composite null hypothesis.

Here we are interested in testing $H_0 : F \in \mathcal{F} = \{F(;\theta) : \theta \in \Omega \subset R^q\}$ against $H_a : F \notin \mathcal{F}$. In this case, the Cramer-von Mises test with θ replaced by some estimators has a complicated asymptotic limit, which depends on the underlying distribution function when $d > 1$. Hence, obtaining critical points requires additional effort such as bootstrap method. Similar to the idea in Section 2.1, one may apply the profile jackknife empirical likelihood method to equation (9) and some estimating equations for θ such as score equations, and then expect Wilks theorem holds. However, when the dimension of θ is large, i.e., q is large, the above profile jackknife empirical likelihood method is computationally intensive. Here we propose to employ the idea in Li, Peng and Qi (2011) as follows.

Let $f(x;\theta)$ denote the density function of X_i, put $G(x;\theta) = \frac{\partial}{\partial\theta} \log f(x;\theta) := (g_1(x;\theta), \cdots, g_q(x;\theta))^T$ and let $\hat{\theta}_n$ and $\hat{\theta}_{n,i}$ denote the consistent solutions to

$$\frac{1}{n}\sum_{j=1}^{n} G(X_j;\theta) = 0 \quad \text{and} \quad \frac{1}{n-1}\sum_{j=1, j\neq i}^{n} G(X_j;\theta) = 0,$$

respectively. A jackknife sample based on (9) is defined as $Y_i^* = (Y_{i,1}^*, Y_{i,2}^*)^T$ for $i = 1, \cdots, n$, where

$$\begin{aligned} Y_{i,1}^* &= n\int F_n^2(x)\,dF(x;\hat{\theta}_n) - (n-1)\int F_{n,i}^2(x)\,dF(x;\hat{\theta}_n) \\ &\quad -\{n\int F^2(x;\hat{\theta}_n)\,dF(x;\hat{\theta}_n) - (n-1)\int F^2(x;\hat{\theta}_{n,i})\,dF(x;\hat{\theta}_n)\} \end{aligned}$$

and

$$\begin{aligned} Y_{i,2}^* &= n\int F_n(x)F(x;\hat{\theta}_n)\,dF(x;\hat{\theta}_n) - (n-1)\int F_{n,i}(x)F(x;\hat{\theta}_{n,i})\,dF(x;\hat{\theta}_n) \\ &\quad + \textstyle\sum_{l=1}^{d}\sum_{j=1}^{n}\{1 - F_l^2(X_{j,l};\hat{\theta}_n)\}^{1/2} - \sum_{l=1}^{d}\sum_{j=1, j\neq i}^{n}\{1 - F_l^2(X_{j,l};\hat{\theta}_{n,i})\}^{1/2} \\ &\quad -\{n\int F^2(x;\hat{\theta}_n)\,dF(x;\hat{\theta}_n) - (n-1)\int F^2(x;\hat{\theta}_{n,i})\,dF(x;\hat{\theta}_n)\} - \tfrac{d\pi}{4}. \end{aligned}$$

Based on this jackknife sample, one can define the jackknife empirical likelihood function as

$$L_c = \sup\{\prod_{i=1}^{n}(np_i) : p_1 \geq 0, \cdots, p_n \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i Y_i^* = 0\}.$$

By Lagrange multiplier technique, we obtain $p_i = n^{-1}\{1+\lambda^T Y_i^*\}^{-1}$ and $-2\log L_c = 2\sum_{i=1}^{n}\log\{1+\lambda^T Y_i^*\}$, where λ satisfies $\sum_{i=1}^{n}\frac{Y_i^*}{1+\lambda^T Y_i^*} = 0$.

Similarly, we can define the jackknife sample based on (10) as $\bar{Y}_i^* = (\bar{Y}_{i,1}^*, \bar{Y}_{i,2}^*)^T$ for $i = 1, \cdots, n$, where

$$\begin{aligned}\bar{Y}_{i,1}^* &= n\int F_n^2(x)\,dF(x;\hat{\theta}_n) - (n-1)\int F_{n,i}^2(x)\,dF(x;\hat{\theta}_n)\\ &\quad -\{n\int F^2(x;\hat{\theta}_n)\,dF(x;\hat{\theta}_n) - (n-1)\int F^2(x;\hat{\theta}_{n,i})\,dF(x;\hat{\theta}_n)\}\end{aligned}$$

and

$$\begin{aligned}\bar{Y}_{i,2}^* &= n\int F_n(x)F(x;\hat{\theta}_n)\,dF(x;\hat{\theta}_n) - (n-1)\int F_{n,i}(x)F(x;\hat{\theta}_{n,i})\,dF(x;\hat{\theta}_n)\\ &\quad -\sum_{l=1}^{d}\sum_{j=1}^{n}\{1-F_l^2(X_{j,l};\hat{\theta}_n)\}^{1/2} + \sum_{l=1}^{d}\sum_{j=1,j\neq i}^{n}\{1-F_l^2(X_{j,l};\hat{\theta}_{n,i})\}^{1/2}\\ &\quad -\{n\int F^2(x;\hat{\theta}_n)\,dF(x;\hat{\theta}_n) - (n-1)\int F^2(x;\hat{\theta}_{n,i})\,dF(x;\hat{\theta}_n)\} + \tfrac{d\pi}{4}.\end{aligned}$$

Based on this jackknife sample, the jackknife empirical likelihood function is defined as

$$\bar{L}_c = \sup\{\prod_{i=1}^{n}(np_i) : p_1 \geq 0, \cdots, p_n \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \bar{Y}_i^* = 0\}.$$

Before proving that Wilks theorem holds for the above proposed jackknife empirical likelihood methods, we list some regularity conditions.

- A1) There is a neighborhood of θ_0, say Ω_0, such that $G(x;\theta)$ are continuous functions of $\theta \in \Omega_0$ for all x, y, and $||G(x;\theta)||^6$, $||\frac{\partial}{\partial\theta}G(x;\theta)||^6$, $||\frac{\partial^2}{\partial\theta^T\partial\theta}g_l(x;\theta)||^3$ and $|\frac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_s}g_l(x;\theta)|$ for $l, i, j, s = 1, \cdots, q$ are bounded by $K(x)$ uniformly in $\theta \in \Omega_0$, where $EK(X_1) < \infty$ and θ_0 denotes the true value of θ;

- A2) $\Sigma_1 = E\{\frac{\partial}{\partial\theta}G(X_1;\theta_0)\}$ is invertible;

- A3) Σ^* defined in Lemma 4 is positive definite.

Theorem 3. *Under conditions A1)–A3) and $H_0 : F \in \mathcal{F}$, both $l_c = -2\log L_c$ and $\bar{l}_c = -2\log \bar{L}_c$ converge in distribution to a chi-square distribution with degrees of freedom two.*

Remark 2. Theorem 3 still holds when G are replaced by q independent estimating equations for θ such that $EG(X_1;\theta) = 0$. When $\hat{\theta}_n$ does not have an explicit formula, one could use the approximate jackknife empirical likelihood method in Peng (2011) to reduce the computational burden of the proposed jackknife empirical likelihood methods.

2.3 Simulation study

We investigate the finite sample behavior of the proposed methods and compare them with the Cramer-von Mises test (T_{CM}), the test statistic T_n in Page 268 of Einmahl and McKeague (2009) (EM), and the test statistic $T_n(s)$ in Page 2020 of Jager and Wellner (2007) with $s = 1/2$ (JW) in terms of both size and power. For computing the test statistics EM and JW we employ the parametric bootstrap method with $1,000$ replications to obtain critical points. When the null hypothesis is composite, we estimate the parameters by maximum likelihood estimation.

2.3.1 Simple null hypothesis with $d = 1$.

We draw $10,000$ random samples of sizes $n = 100$ and $1,000$ from the following mixture distributions

$$\text{Model 1} : (1 - \frac{\delta}{\sqrt{n}})N(\mu,\sigma^2) + \frac{\delta}{\sqrt{n}}t(\nu)$$

and

$$\text{Model 2} : (1 - \frac{\delta}{\sqrt{n}}) LogNormal(\mu, \sigma^2) + \frac{\delta}{\sqrt{n}} exponential(\nu).$$

The null hypothesis is $\delta = 0$. The significance levels are chosen at 0.1 and 0.05. We calculate the empirical powers of the proposed tests $l_s, \bar{l}_s$, and the existing tests T_{CM}, EM and JW mentioned above. We only report the cases in Tables 1–4 for level 0.05 since similar results hold for level 0.1. When $n = 1000$, Tables 2 and 4 show that all tests have an accurate size, and the Cramer-von Mises test is less powerful than the other tests for most cases. The tests in Einmahl and McKeague (2003) and Jager and Wellner (2007) are much more powerful than the proposed jackknife empirical likelihood tests for some cases. But for some other cases, they are much less powerful. Note that the tests in Einmahl and McKeague (2003) and Jager and Wellner (2007) are type of the Anderson-Darling test. Therefore it would be of interest to develop corresponding jackknife empirical likelihood tests based on the Anderson-Darling test instead of the Cramer-von Mises test. Although it is hard to draw any conclusion from the comparison between the tests in Einmahl and McKeague (2003) and Jager and Wellner (2007) and the proposed jackknife empirical likelihood tests, the tests in Einmahl and McKeague (2003) and Jager and Wellner (2007) are more computationally intensive since bootstrap methods are required to obtain critical points. Tables 1 and 3 show that the proposed jackknife empirical likelihood tests have a large size when $n = 100$.

2.3.2 Composite null hypothesis with $d = 1$.

We draw $10,000$ random samples of sizes $n = 100$ and $1,000$ from Models 1 and 2. The null hypothesis is H_0 : normal distribution for Model 1 and H_0 : Log-normal distribution for Model 2. Here $q = 2$. The empirical powers of the proposed tests $l_c, \bar{l}_c$, the Cramer-von Mises test and the tests in Einmahl and McKeague (2003) and Jager and Wellner (2007) are reported in Tables 5–8 for level 0.05. When $n = 1000$, Tables 6 and 8 show that all tests have an accurate size and the proposed jackknife empirical likelihood tests are slightly more powerful than the other tests. The proposed jackknife empirical likelihood tests have a large size when $n = 100$, but have much less computation than the other tests, which require using bootstrap methods to obtain critical points. Similar conclusions are true for level 0.1.

2.3.3 Simple null hypothesis with $d = 2$.

We draw $5,000$ random samples of size $n = 1,000$ from the following mixture distribution

$$\text{Model 3} : (1 - \frac{\delta}{\sqrt{n}})N(\begin{pmatrix} 0 \\ \mu \end{pmatrix}, \begin{pmatrix} 1 & 0.5\sigma \\ 0.5\sigma & \sigma^2 \end{pmatrix}) + \frac{\delta}{\sqrt{n}} t(\nu).$$

The null hypothesis is $\delta = 0$. The significance level is chosen at 0.05 and 0.1. For calculating the Cramer-von Mises test, we employ a bootstrap method with $1,000$ replications to obtain the critical points. Note that the tests in Einmahl and McKeague (2003) and Jager and Wellner (2007) are developed for $d = 1$. Although these tests may be extended to the case of $d > 1$, computing them would not be easy. Hence we do not compare with them. The empirical powers of the proposed tests $l_s, \bar{l}_s$, and the Cramer-von Mises test with level 0.05 are reported in Table 9, which shows that the proposed empirical likelihood tests are more powerful. Also they are less computationally intensive than the Cramer-von Mises test which requires using a bootstrap method to obtain critical points. We do not report results for level 0.1, which are similar to those for level 0.05.

2.3.4 Composite null hypothesis with $d = 2$.

We draw $5,000$ random samples of size $n = 1000$ from Model 3. The null hypothesis is H_0 : bivariate normal distribution. Here $q = 5$. The significance level is chosen at levels 0.05 and 0.1. The empirical powers of the proposed tests $l_c, \bar{l}_c$, and the Cramer-von Mise test with level 0.05 are reported in Table 10, which shows that the proposed jackknife empirical likelihood tests are slightly more powerful than the Cramer-von Mises test for most cases. Similar conclusions are true for level 0.1.

In summary, when $n = 1000$, the proposed empirical likelihood tests are much more powerful than the Cramer-von Mises test for most considered cases and are much less computationally intensive than the other considered tests. It remains interesting to develop similar jackknife empirical likelihood tests based on the Anderson-Darling test instead of the Cramer-von Mises test so as to better compare with tests in Einmahl and McKeague (2003) and Jager and Wellner (2007).

2.4 Proofs

It is known that there exists a Wiener process W such that

$$\sup_{x\in(-\infty,\infty)^d} |\sqrt{n}\{F_n(x) - F_0(x)\} - W(x)| = o_p(1) \tag{11}$$

as $n \to \infty$, where $EW(x) = 0$ and $E\{W(x)W(y)\} = F_0(x \wedge y) - F_0(x)F_0(y)$. We use $W_l(x_l)$ to denote $W(x)$ with $x = (x_1, \cdots, x_d)^T$ and $x_k = \infty$ for $k \neq l$.

Before proving Theorem 2, we show the following two lemmas.

Lemma 1. Under $H_0 : F \equiv F_0$, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Y_i \xrightarrow{d} N(0, \Sigma),$$

as $n \to \infty$, where $\Sigma = (\sigma_{ij})_{1 \le i,j \le 2}$, $\sigma_{11} = 4A_1$, $\sigma_{22} = A_1 + 2A_2 + 2A_3 + \frac{2d}{3} - (\frac{d\pi}{4})^2$, $\sigma_{12} = \sigma_{21} = 2A_1 + 2A_3$,

$$A_1 = \int \{F_0(x \wedge y) - F_0(x)F_0(y)\}F_0(x)F_0(y)\, dF_0(x)dF_0(y),$$

$$A_2 = \sum\sum_{1 \le l < k \le d} \int (1 - F_{0l}^2(x_l))^{1/2}(1 - F_{0k}(y_k))^{1/2}\, dF_{0lk}(x_l, y_k),$$

and

$$A_3 = \int \{I(x < y) - F_0(y)\} \sum_{l=1}^{d} (1 - F_{0l}^2(x_l))^{1/2} F_0(y)\, dF_0(x)dF_0(y)$$

for $x = (x_1, \cdots, x_d)^T$ and $y = (y_1, \cdots, y_d)^T$.

Proof. Since

$$\begin{cases} F_{n,i}(x) = \frac{n}{n-1}F_n(x) - \frac{1}{n-1}I(X_i \le x), \\ F_{n,i}^2(x) = \frac{n^2}{(n-1)^2}F_n^2(x) - 2\frac{n}{(n-1)^2}F_n(x)I(X_i \le x) + \frac{1}{(n-1)^2}I(X_i \le x), \end{cases} \tag{12}$$

we have

$$\begin{aligned} Y_{i,1} &= n\int F_n^2(x)\, dF_0(x) - (n-1)\int F_{n,i}^2(x)\, dF_0(x) - \int F_0^2(x)\, dF_0(x) \\ &= -\frac{n}{n-1}\int F_n^2(x)\, dF_0(x) + 2\frac{n}{n-1}\int F_n(x)I(X_i \le x)\, dF_0(x) \\ &\quad -\frac{1}{n-1}\int I(X_i \le x)\, dF_0(x) - \int F_0^2(x)\, dF_0(x) \end{aligned} \tag{13}$$

and

$$Y_{i,2} = \int I(X_i \le x)F_0(x)\, dF_0(x) + \sum_{l=1}^{d}\{1 - F_{0l}^2(X_{i,l})\}^{1/2} - \int F_0^2(x)\, dF_0(x) - \frac{d\pi}{4}. \tag{14}$$

Therefore, it follows from (11) that

$$\begin{aligned}
& \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n Y_{i,1} \\
= \;& -\tfrac{n\sqrt{n}}{n-1} \textstyle\int F_n^2(x)\, dF_0(x) + 2\tfrac{n\sqrt{n}}{n-1} \int F_n^2(x)\, dF_0(x) \\
& -\tfrac{\sqrt{n}}{n-1} \textstyle\int F_n(x)\, dF_0(x) - \sqrt{n} \int F_0^2(x)\, dF_0(x) \\
= \;& \tfrac{n}{n-1} \textstyle\int \sqrt{n}\{F_n(x) - F_0(x)\} F_n(x)\, dF_0(x) + \tfrac{n}{n-1} \int \sqrt{n}\{F_n(x) - F_0(x)\} F_0(x)\, dF_0(x) \\
& -\tfrac{1}{n-1} \textstyle\int \sqrt{n}\{F_n(x) - F_0(x)\}\, dF_0(x) + \tfrac{\sqrt{n}}{n-1} \int F_0^2(x)\, dF_0(x) - \tfrac{\sqrt{n}}{n-1} \int F_0(x)\, dF_0(x) \\
= \;& 2 \textstyle\int W(x) F_0(x)\, dF_0(x) + o_p(1)
\end{aligned} \tag{15}$$

and

$$\begin{aligned}
& \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n Y_{i,2} \\
= \;& \sqrt{n} \textstyle\int F_n(x) F_0(x)\, dF_0(x) + \sum_{l=1}^d \sqrt{n} \int \{1 - F_{0l}^2(x_l)\}^{1/2}\, dF_n^{(l)}(x_l) \\
& -\sqrt{n} \textstyle\int F_0^2(x)\, dF_0(x) - \tfrac{d\pi\sqrt{n}}{4} \\
= \;& \sqrt{n} \textstyle\int F_n(x) F_0(x)\, dF_0(x) + \sum_{l=1}^d \sqrt{n} \int F_n^{(l)}(x_l) \{1 - F_{0l}^2(x_l)\}^{-1/2} F_{0l}(x_l)\, dF_{0l}(x_l) \\
& -\sqrt{n} \textstyle\int F_0^2(x)\, dF_0(x) - \tfrac{d\pi\sqrt{n}}{4} \\
= \;& \textstyle\int \sqrt{n}\{F_n(x) - F_0(x)\} F_0(x)\, dF_0(x) \\
& + \textstyle\sum_{l=1}^d \int \sqrt{n}\{F_n^{(l)}(x_l) - F_{0l}(x_l)\}\{1 - F_{0l}^2(x_l)\}^{-1/2} F_{0l}(x_l)\, dF_{0l}(x_l) \\
= \;& \textstyle\int W(x) F_0(x)\, dF_0(x) + \sum_{l=1}^d \int W_l(x_l)\{1 - F_{0l}^2(x_l)\}^{-1/2} F_{0l}(x_l)\, dF_{0l}(x_l) + o_p(1).
\end{aligned} \tag{16}$$

It is straightforward to check that

$$E\{\int W(x) F_0(x)\, dF_0(x)\}^2 = \int E\{W(x)W(y)\} F_0(x) F_0(y)\, dF_0(x) dF_0(y) = A_1, \tag{17}$$

$$
\begin{aligned}
& E\{\textstyle\sum_{l=1}^{d} \int W_l(x_l)(1-F_{0l}^2(x_l))^{-1/2} F_{0l}(x_l)\, dF_{0l}(x_l)\}^2 \\
= \; & \textstyle\sum_{l=1}^{d} \int \{F_{0l}(x_l \wedge y_l) - F_{0l}(x_l)F_{0l}(y_l)\}(1-F_{0l}^2(x_l))^{-1/2} F_{0l}(x_l) \times \\
& (1-F_{0l}^2(y_l))^{-1/2} F_{0l}(y_l)\, dF_{0l}(x_l) dF_{0l}(y_l) \\
& +2 \textstyle\sum\sum_{1\le l<k\le d} \int \{F_{0lk}(x_l, y_l) - F_{0l}(x_l)F_{0l}(y_k)\}(1-F_{0l}^2(x_l))^{-1/2} F_{0l}(x_l) \times \\
& (1-F_{0k}^2(y_k))^{-1/2} F_{0k}(y_k)\, dF_{0l}(x_l) dF_{0k}(y_k) \\
= \; & \textstyle\sum_{l=1}^{d} \int (x_l \wedge y_l - x_l y_l)(1-x_l^2)^{-1/2} x_l (1-y_l^2)^{-1/2} y_l\, dx_l dy_l \\
& +2 \textstyle\sum\sum_{1\le l<k\le d} \int F_{0lk}(x_l, y_k)(1-F_{0l}^2(x_l))^{-\frac{1}{2}} F_{0l}(x_l) \times \\
& (1-F_{0k}^2(y_k))^{-\frac{1}{2}} F_{0k}(y_k)\, dF_{0l}(x_l) dF_{0k}(y_k) \\
& -2 \textstyle\sum\sum_{1\le l<k\le d} \int x_l y_k (1-x_l^2)^{-\frac{1}{2}} x_l (1-y_k^2)^{-\frac{1}{2}} y_k\, dx_l dy_k \\
= \; & \tfrac{2d}{3} - d(\tfrac{\pi}{4})^2 + 2 \textstyle\sum\sum_{1\le l<k\le d} \int (1-F_{0l}^2(x_l))^{\frac{1}{2}} (1-F_{0k}^2(y_k))^{\frac{1}{2}}\, dF_{0lk}(x_l, y_k) \\
& -(d^2-d)(\tfrac{\pi}{4})^2 \\
= \; & \tfrac{2d}{3} - (\tfrac{d\pi}{4})^2 + 2A_2,
\end{aligned}
\tag{18}
$$

and

$$
\begin{aligned}
& E\{\int W(x)F_0(x)\,dF_0(x)\}\{\sum_{l=1}^d \int W_l(x_l)(1-F_{0l}^2(x_l))^{-1/2}F_{0l}(x_l)\,dF_{0l}(x_l)\} \\
= \; & \sum_{l=1}^d \int \{F_0((y_1,\cdots,y_{l-1},x_l\wedge y_l,y_{l+1},\cdots,y_d)^T) - F_{0l}(x_l)F_0(y)\}\times \\
& F_0(y)(1-F_{0l}^2(x_l))^{-1/2}F_{0l}(x_l)\,dF_{0l}(x_l)dF_0(y) \\
= \; & -\sum_{l=1}^d \int F_0((y_1,\cdots,y_{l-1},x_l\wedge y_l,y_{l+1},\cdots,y_d)^T)F_0(y)\,d(1-F_{0l}^2(x_l))^{1/2}dF_0(y) \\
& +\sum_{l=1}^d \int F_{0l}(x_l)F_0^2(y)\,d(1-F_{0l}^2(x_l))^{1/2}dF_0(y) \\
= \; & \sum_{l=1}^d \int I(x_l\le y_l)F_0(y)(1-F_{0l}^2(x_l))^{1/2}\times \\
& \{\tfrac{\partial}{\partial x_l}F_0((y_1,\cdots,y_{l-1},x_l,y_{l+1},\cdots,y_d)^T)\}\,dx_l dF_0(y) \\
& -\sum_{l=1}^d \int F_0^2(y)(1-F_{0l}^2(x_l))^{1/2}\,dF_{0l}(x_l)dF_0(y) \\
= \; & \sum_{l=1}^d \int I(x\le y)F_0(y)(1-F_{0l}^2(x_l))^{1/2}\,dF_0(x)dF_0(y) \\
& -\sum_{l=1}^d \int F_0^2(y)(1-F_{0l}^2(x_l))^{1/2}\,dF_0(x)dF_0(y) \\
= \; & A_3,
\end{aligned}
\tag{19}
$$

where F_{0lk} denotes the distribution function of $(X_{i,l}, X_{i,k})$ under H_0. Hence, the lemma follows from (15)–(19).

Lemma 2. Under $H_0 : F \equiv F_0$, we have

$$
\frac{1}{n}\sum_{i=1}^n Y_iY_i^T \xrightarrow{p} \Sigma
$$

as $n \to \infty$.

Proof. It follows from (13) and (14) that

$$
\begin{aligned}
& \tfrac{1}{n}\textstyle\sum_{i=1}^{n} Y_{i,1}^2 \\
= \ & \{\tfrac{n}{n-1}\textstyle\int F_n^2(x)\, dF_0(x)\}^2 \\
& + \tfrac{4n^2}{(n-1)^2}\textstyle\int F_n(x)F_n(y)\tfrac{1}{n}\sum_{i=1}^{n} I(X_i \le x \wedge y)\, dF_0(x)dF_0(y) \\
& + \{\textstyle\int F_0^2(x)\, dF_0(x)\}^2 \\
& + 2\{-\tfrac{n}{n-1}\textstyle\int F_n^2(x)\, dF_0(x)\}\{2\tfrac{n}{n-1}\int F_n(x)\tfrac{1}{n}\sum_{i=1}^{n} I(X_i \le x)\, dF_0(x)\} \\
& + \tfrac{2n}{n-1}\{\textstyle\int F_n^2(x)\, dF_0(x)\}\{\int F_0^2(x)\, dF_0(x)\} \\
& - \tfrac{4n}{n-1}\{\textstyle\int F_n(x)\tfrac{1}{n}\sum_{i=1}^{n} I(X_i \le x)\, dF_0(x)\}\{\int F_0^2(x)\, dF_0(x)\} + o_p(1) \\
= \ & 4A_1 + o_p(1),
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} Y_{i,2}^2 \\
= \quad &\int \frac{1}{n}\sum_{i=1}^{n} I(X_i \le x \wedge y) F_0(x) F_0(y)\, dF_0(x) dF_0(y) \\
&+\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{d}\{1 - F_{0l}^2(X_{i,l})\} \\
&+\frac{1}{n}\sum_{i=1}^{n} 2\sum\sum_{1\le l<k\le d}\{1 - F_{0l}^2(X_{i,l})\}^{1/2}\{1 - F_{0k}^2(X_{i,k})\}^{1/2} \\
&+\{\int F_0^2(x)\, dF_0(x) + \tfrac{d\pi}{4}\}^2 \\
&+2\int I(y \le x) F_0(x) \sum_{l=1}^{d}\{1 - F_{0l}^2(y_l)\}^{1/2}\, dF_0(x) dF_n(y) \\
&-2\{\int F_n(x) F_0(x)\, dF_0(x)\}\{\int F_0^2(x)\, dF_0(x) + \tfrac{d\pi}{4}\} \\
&-2\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{d}\{1 - F_{0l}^2(X_{i,l})\}^{1/2}\{\int F_0^2(x)\, dF_0(x) + \tfrac{d\pi}{4}\} + o_p(1) \\
= \quad &\int F_0(x \wedge y) F_0(x) F_0(y)\, dF_0(x) dF_0(y) + \tfrac{2d}{3} \\
&+2\sum\sum_{1\le l<k\le d}\int\{1 - F_{0l}^2(x_l)\}^{1/2}\{1 - F_{0k}^2(y_k)\}^{1/2}\, dF_{0lk}(x_l, y_k) \\
&+\{\int F_0^2(x)\, dF_0(x) + \tfrac{d\pi}{4}\}^2 \\
&+2\int I(y \le x) F_0(x) \sum_{l=1}^{d}\{1 - F_{0l}(y_l)\}^{1/2}\, dF_0(x) dF_0(y) \\
&-2\{\int F_0^2(x)\, dF_0(x)\}\{\int F_0^2(x)\, dF_0(x) + \tfrac{d\pi}{4}\} \\
&-2\tfrac{d\pi}{4}\{\int F_0^2(x)\, dF_0(x) + \tfrac{d\pi}{4}\} + o_p(1) \\
= \quad &A_1 + \tfrac{2d}{3} + 2A_2 + 2A_3 - (\tfrac{d\pi}{4})^2,
\end{aligned}
$$

and

$$\begin{aligned}
& \tfrac{1}{n}\textstyle\sum_{i=1}^n Y_{i,1}Y_{i,2} \\
= \; & \{-\tfrac{n}{n-1}\textstyle\int F_n^2(x)\,dF_0(x)\}\{\int F_n(x)F_0(x)\,dF_0(x)\} \\
& +2\tfrac{n}{n-1}\textstyle\int F_n(x)F_n(x\wedge y)F_0(y)\,dF_0(x)dF_0(y) \\
& +\{-\textstyle\int F_0^2(x)\,dF_0(x)\}\{\int F_n(x)F_0(x)\,dF_0(x)\} \\
& +\{-\tfrac{n}{n-1}\textstyle\int F_n^2(x)\,dF_0(x)\}\{\frac{1}{n}\sum_{i=1}^n\sum_{l=1}^d(1-F_{0l}^2(X_{i,l}))^{1/2}\} \\
& +2\tfrac{n}{n-1}\textstyle\int F_n(x)\frac{1}{n}\sum_{i=1}^n I(X_i\le x)\sum_{l=1}^d(1-F_{0l}^2(X_{i,l}))^{1/2}\,dF_0(x) \\
& +\{-\textstyle\int F_0^2(x)\,dF_0(x)\}\frac{1}{n}\sum_{i=1}^n\sum_{l=1}^d(1-F_{0l}^2(X_{i,l}))^{1/2}+o_p(1) \\
= \; & -\{\textstyle\int F_0^2(x)\,dF_0(x)\}^2+2\int F_0(x\wedge y)F_0(x)F_0(y)\,dF_0(x)dF_0(y) \\
& -\{\textstyle\int F_0^2(x)\,dF_0(x)\}^2-\frac{d\pi}{4}\int F_0^2(x)\,dF_0(x) \\
& +2\textstyle\int I(y\le x)F_0(x)\sum_{l=1}^d(1-F_{0l}^2(y_l))^{\frac{1}{2}}\,dF_0(x)dF_0(y)-\frac{d\pi}{4}\int F_0^2(x)\,dF_0(x)+o_p(1) \\
= \; & 2A_1+2A_3+o_p(1),
\end{aligned}$$

which imply the lemma.

Proof of Theorem 2. Note that $\sup_{1\le i\le n}||Y_i||$ is bounded. Hence the convergence of l_c follows from Lemmas 1, 2 and the standard arguments in the empirical likelihood method (see Owen (1990)). Similarly, we can show the convergence of $\bar{l}_c$.

Before proving Theorem 3, we need the following lemmas.

Lemma 3. Under conditions of Theorem 3, we have

$$\hat{\theta}_n-\theta_0+\Sigma_1^{-1}\frac{1}{n}\sum_{j=1}^n G(X_j;\theta_0)=O_p(n^{-1}), \tag{20}$$

$$\max_{1\le i\le n}||\hat{\theta}_{n,i}-\theta_0+\Sigma_1^{-1}\frac{1}{n-1}\sum_{j=1,j\ne i}^n G(X_j;\theta_0)||=O_p(n^{-1}) \tag{21}$$

and

$$\max_{1\le i\le n}||\hat{\theta}_n-\theta_0+\Sigma_{n1}^{-1}D_n(i)||=o_p(n^{-3/2}), \tag{22}$$

where $\Sigma_{n1} = \frac{1}{n}\sum_{j=1}^{n}\frac{\partial}{\partial\theta}G(X_j;\theta_0)$ and

$$D_n(i) = n^{-1}G(X_i;\theta_0) - n^{-1}\{\frac{\partial}{\partial\theta}G(X_i;\theta_0)\}\Sigma_1^{-1}\frac{1}{n-1}\sum_{j=1}^{n}G(X_j;\theta_0).$$

Proof. Since condition A1) implies that

$$\max_{1\le i\le n}\sup_{\theta\in\Omega_0}\frac{1}{n-1}||G(X_i;\theta)|| = o(n^{-1/2}) \tag{23}$$

almost surely, we have

$$\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}G(X_j;\theta) \to EG(X_1;\theta)$$

almost surely and uniformly for $1\le i\le n$ and $\theta\in\Omega_0$, which imply that

$$\hat{\theta}_n - \theta_0 = o_p(1) \quad \text{and} \quad \max_{1\le i\le n}||\hat{\theta}_{n,i}-\theta_0|| = o_p(1)$$

by Theorem B in Section 7.2.1 of Serfling (2002). Hence, by Taylor expansion, we have

$$\begin{aligned} 0 &= \tfrac{1}{n-1}\textstyle\sum_{j=1,j\neq i}^{n}G(X_j;\hat{\theta}_{n,i}) \\ &= \tfrac{1}{n-1}\textstyle\sum_{j=1,j\neq i}^{n}G(X_j;\theta_0) \\ &\quad + \tfrac{1}{n-1}\textstyle\sum_{j=1,j\neq i}^{n}\{\frac{\partial}{\partial\theta}G(X_j;\gamma_i\theta_0+(1-\gamma_i)\hat{\theta}_{n,i})\}\{\hat{\theta}_{n,i}-\theta_0\}, \end{aligned} \tag{24}$$

where $\gamma_i\in[0,1]$ for $i=1,\cdots,n$. Using (23), we can show that

$$\max_{1\le i\le n}||\frac{1}{n-1}\sum_{j=1,j\neq i}^{n}\frac{\partial}{\partial\theta}G(X_j;\gamma_i\theta_0+(1-\gamma_i)\hat{\theta}_{n,i}) - \Sigma_1|| = o_p(1).$$

Hence (21) follows from (24). Similarly we can show (20).

It is easy to check that (23), (20) and (21) imply that

$$\max_{1\le i\le n}||\hat{\theta}_n-\hat{\theta}_{n,i}|| = o_p(n^{-1/2}) \quad \text{and} \quad \max_{1\le i\le n}||\hat{\theta}_{n,i}-\theta_0|| = O_p(n^{-1/2}), \tag{25}$$

which imply that

$$\begin{aligned}
0 &= \tfrac{1}{n}\{\textstyle\sum_{j=1}^n G(X_j;\hat\theta_n) - \sum_{j=1,j\neq i}^n G(X_j;\hat\theta_{n,i})\}\\
&= \tfrac{1}{n}\textstyle\sum_{j=1}^n\{G(X_j;\hat\theta_n) - G(X_j;\hat\theta_{n,i})\} + n^{-1}G(X_i;\hat\theta_{n,i})\\
&= \tfrac{1}{n}\textstyle\sum_{j=1}^n\{\frac{\partial}{\partial\theta}G(X_j;\hat\theta_n)\}\{\hat\theta_n - \hat\theta_{n,i}\} + n^{-1}G(X_i;\hat\theta_{n,i}) + o_p(n^{-1})\\
&= \tfrac{1}{n}\textstyle\sum_{j=1}^n\{\frac{\partial}{\partial\theta}G(X_j;\theta_0)\}\{\hat\theta_n - \hat\theta_{n,i}\} + n^{-1}G(X_i;\theta_0) + o_p(n^{-1})\\
&= \Sigma_1\{\hat\theta_n - \hat\theta_{n,i}\} + n^{-1}G(X_i;\theta_0) + o_p(n^{-1})
\end{aligned}$$

uniformly in $1 \le i \le n$, i.e.,

$$\max_{1\le i\le n} ||\hat\theta_n - \hat\theta_{n,i} + \Sigma_1^{-1}n^{-1}G(X_i;\theta_0)|| = o_p(n^{-1}). \tag{26}$$

Using (25), (26) and Taylor expansion again, we have for $l = 1, \cdots, q$

$$\begin{aligned}
0 &= \tfrac{1}{n}\{\textstyle\sum_{j=1}^n g_l(X_j;\hat\theta_n) - \sum_{j=1,j\neq i}^n g_l(X_j;\hat\theta_{n,i})\}\\
&= \tfrac{1}{n}\textstyle\sum_{j=1}^n\{g_l(X_j;\hat\theta_n) - g_l(X_j;\hat\theta_{n,i})\} + n^{-1}g_l(X_i;\hat\theta_{n,i})\\
&= \tfrac{1}{n}\textstyle\sum_{j=1}^n\{\frac{\partial}{\partial\theta}g_l(X_j;\hat\theta_n)\}\{\hat\theta_n - \hat\theta_{n,i}\}\\
&\quad +\tfrac{1}{2n}\textstyle\sum_{j=1}^n\{\hat\theta_n - \hat\theta_{n,i}\}^T\{\frac{\partial^2}{\partial\theta^T\partial\theta}g_l(X_j;\hat\theta_n)\}\{\hat\theta_n - \hat\theta_{n,i}\}\\
&\quad +n^{-1}g_l(X_i;\theta_0) + n^{-1}\{\tfrac{\partial}{\partial\theta}g_l(X_i;\theta_0)\}\{\hat\theta_{n,i} - \theta_0\} + o_p(n^{-3/2})\\
&= \tfrac{1}{n}\textstyle\sum_{j=1}^n\{\frac{\partial}{\partial\theta}g_l(X_j;\hat\theta_n)\}\{\hat\theta_n - \hat\theta_{n,i}\}\\
&\quad +\tfrac{1}{2}\{n^{-1}G^T(X_i;\theta_0)\Sigma_1^{-1}\}\{\tfrac{1}{n}\textstyle\sum_{j=1}^n \frac{\partial^2}{\partial\theta^T\partial\theta}g_l(X_j;\hat\theta_n)\}\{\Sigma_1^{-1}n^{-1}G(X_i;\theta_0)\}\\
&\quad +n^{-1}g_l(X_i;\theta_0) - n^{-1}\{\tfrac{\partial}{\partial\theta}g_l(X_i;\theta_0)\}\Sigma_1^{-1}\tfrac{1}{n-1}\textstyle\sum_{j=1,j\neq i}^n G(X_j;\theta_0) + o_p(n^{-3/2})\\
&= \{\tfrac{1}{n}\textstyle\sum_{j=1}^n \frac{\partial}{\partial\theta}g_l(X_j;\hat\theta_n)\}^T\{\hat\theta_n - \hat\theta_{n,i}\}\\
&\quad +n^{-1}g_l(X_i;\theta_0) - n^{-1}\{\tfrac{\partial}{\partial\theta}g_l(X_i;\theta_0)\}\Sigma_1^{-1}\tfrac{1}{n-1}\textstyle\sum_{j=1}^n G(X_j;\theta_0) + o_p(n^{-3/2})
\end{aligned}$$

uniformly in $1 \le i \le n$, which imply (22).

Lemma 4. Under conditions of Theorem 3, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n Y_i^* \xrightarrow{d} N(0,\Sigma^*)$$

as $n \to \infty$, where $\Sigma^* = (\sigma^*_{ij})_{1\le i,j\le 2}$, $\sigma'_{ij}s$ are defined in Lemma 1,

$$\begin{cases} \sigma^*_{11} = \sigma_{11} + 4A_6\Sigma_1^{-1}A_6 - 8A_6\Sigma_1^{-1}\int A_5(x)F(x;\theta_0)\,dF(x;\theta_0) \\ \sigma^*_{22} = \sigma_{22} + A_6\Sigma_1^{-1}A_6 - 2A_6\Sigma_1^{-1}\int A_5(x)F(x;\theta_0)\,dF(x;\theta_0) - 2A_6\Sigma_1^{-1}A_4 \\ \sigma^*_{12} = \sigma^*_{21} = \sigma_{12} - 4A_6\Sigma_1^{-1}\int A_5(x)F(x;\theta_0)\,dF(x;\theta_0) - 2A_6\Sigma_1^{-1}A_4 + 2A_6\Sigma_1^{-1}A_6 \\ A_4 = E\{\sum_{l=1}^d (1-F_l^2(X_{1,l};\theta_0))^{1/2}G(X_1;\theta_0)\} \\ A_5(x) = E\{G(X_1;\theta_0)I(X_1\le x)\} \\ A_6 = \int F(x;\theta_0)\frac{\partial}{\partial\theta}F(x;\theta_0)\,dF(x;\theta_0). \end{cases}$$

Proof. It is straightforward to check that

$$\sum_{i=1}^n D_n(i) = \{-n^{-1}\Sigma_1 + \Sigma_1 - \frac{1}{n}\sum_{i=1}^n \frac{\partial}{\partial\theta}G(X_i;\theta_0)\}\Sigma_1^{-1}\frac{1}{n-1}\sum_{j=1}^n G(X_j;\theta_0) = o_p(n^{-1/2})$$

and $\sum_{i=1}^n D_n^T(i)\Delta D_n(i) = o_p(n^{-1/2})$ for any $q\times q$ matrix Δ. Put $H(x;\theta) = \sum_{l=1}^d\{1 - F_l^2(x_l;\theta)\}^{1/2}$ for $x = (x_1,\cdots,x_d)^T$. Hence, by Lemma 3 and Taylor expansion, we have

$$\begin{aligned} &\sum_{i=1}^n\sum_{j=1}^n H(X_j;\hat\theta_n) - \sum_{i=1}^n\sum_{j=1,j\ne i}^n H(X_j;\hat\theta_{n,i}) \\ =&\sum_{i=1}^n\sum_{j=1}^n\{H(X_j;\hat\theta_n) - H(X_j;\hat\theta_{n,i})\} + \sum_{i=1}^n H(X_i;\hat\theta_{n,i}) \\ =&\sum_{i=1}^n\sum_{j=1}^n\{\tfrac{\partial}{\partial\theta}H(X_j;\hat\theta_n)\}\{\hat\theta_n - \hat\theta_{n,i}\} \\ &+\tfrac12\sum_{i=1}^n\sum_{j=1}^n\{\hat\theta_n-\hat\theta_{n,i}\}^T\{\tfrac{\partial^2}{\partial\theta^T\partial\theta}H(X_j;\hat\theta_n)\}\{\hat\theta_n-\hat\theta_{n,i}\} \\ &+\sum_{i=1}^n H(X_i;\theta_0) + \sum_{i=1}^n\{\tfrac{\partial}{\partial\theta}H(X_i;\theta_0)\}\{\hat\theta_{n,i}-\theta_0\} + o_p(n^{1/2}) \\ =&-\sum_{i=1}^n\sum_{j=1}^n\{\tfrac{\partial}{\partial\theta}H(X_j;\hat\theta_n)\}\Sigma_{n1}^{-1}D_n(i) \\ &+\tfrac12\sum_{i=1}^n\sum_{j=1}^n D_n^T(i)\Sigma_{n1}^{-1}\{\tfrac{\partial^2}{\partial\theta^T\partial\theta}H(X_j;\hat\theta_n)\}\Sigma_{n1}^{-1}D_n(i) \\ &+\sum_{i=1}^n H(X_i;\theta_0) - \sum_{i=1}^n\{\tfrac{\partial}{\partial\theta}H(X_i;\theta_0)\}\Sigma_1^{-1}\tfrac{1}{n-1}\sum_{l=1}^n G(X_l;\theta_0) \\ &+\sum_{i=1}^n\{\tfrac{\partial}{\partial\theta}H(X_i;\theta_0)\}\Sigma_1^{-1}\tfrac{1}{n-1}G(X_i;\theta_0) + o_p(n^{1/2}) \\ =&\sum_{i=1}^n H(X_i;\theta_0) - E\{\tfrac{\partial}{\partial\theta}H(X_1;\theta_0)\}\Sigma_1^{-1}\sum_{i=1}^n G(X_i;\theta_0) + o_p(n^{1/2}). \end{aligned} \tag{27}$$

Similarly we can show that

$$\begin{aligned}
&\textstyle\sum_{i=1}^{n}\{n\int F^2(x;\hat\theta_n)\,dF(x;\hat\theta_n)-(n-1)\int F^2(x;\hat\theta_{n,i})\,dF(x;\hat\theta_n)\}\\
=\;&\textstyle\int F^2(x;\theta_0)\,dF(x;\theta_0)+\{2\int F(x;\theta_0)\frac{\partial}{\partial\theta}F(x;\theta_0)\,dF(x;\theta_0)\}\Sigma_1^{-1}\sum_{i=1}^{n}G(X_i;\theta_0)\\
&+o_p(n^{1/2}).
\end{aligned}\tag{28}$$

Similar to the proofs of (15) and (16), it follows from (27) and (28) that

$$\begin{aligned}
&\textstyle\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Y_{i,1}^*\\
=\;&\textstyle 2\int\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(I(X_i\le x)-F(x;\theta_0))F(x;\theta_0)\,dF(x;\theta_0)\\
&\textstyle -2\{\int F(x;\theta_0)\frac{\partial}{\partial\theta}F(x;\theta_0)\,dF(x;\theta_0)\}\Sigma_1^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}G(X_i;\theta_0)+o_p(1)
\end{aligned}\tag{29}$$

and

$$\begin{aligned}
&\textstyle\frac{1}{\sqrt{n}}\sum_{i=1}^{n}Y_{i,2}^*\\
=\;&\textstyle\int\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(I(X_i\le x)-F(x;\theta_0))F(x;\theta_0)\,dF(x;\theta_0)\\
&\textstyle +\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{d}\{(1-F_l^2(X_{i,l};\theta_0))^{1/2}-\frac{\pi}{4}\}\\
&\textstyle -\{\int F(x;\theta_0)\frac{\partial}{\partial\theta}F(x;\theta_0)\,dF(x;\theta_0)\}\Sigma_1^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}G(X_i;\theta_0)+o_p(1).
\end{aligned}\tag{30}$$

It is easy to verify that

$$E\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(I(X_i\le x)-F(x;\theta_0))\}\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}G(X_i;\theta_0)\}=A_5(x)\tag{31}$$

and

$$E\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{l=1}^{d}(1-F_l^2(X_{i,l};\theta_0))^{1/2}-\frac{d\pi}{4}\}\{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}G(X_i;\theta_0)\}=A_4.\tag{32}$$

Hence the lemma follows from (17)–(19), (29)–(32).

Lemma 5. Under conditions of Theorem 2, we have

$$\frac{1}{n}\sum_{i=1}^{n}Y_i^*Y_i^{*T}\xrightarrow{p}\Sigma^*$$

as $n\to\infty$.

Proof. Similar to the proof of Lemma 2.

Proof of Theorem 3. It can be shown by Lemmas 4-5 and the standard arguments in the empirical likelihood method.

2.5 Tables

Table 1: Empirical powers are reported for the proposed tests in Section 2.2.1, the Cramer-von Mises test (T_{CM}), the test proposed by Einmahl and McKeague (2003) (EM) and the test proposed by Jager and Wellner (2007)(JW) with nominal level 0.05 for Model 1, $n = 100$ and the simple null hypothesis $H_0 : F \sim N(\mu, \sigma^2)$.

$(\delta, \mu, \sigma, \nu)$	T_{CM}	EM	JW	l_s	$\bar{l}_s$
$(0, -1, 0.3, 1)$	0.0542	0.0542	0.0543	0.0954	0.0717
$(1, -1, 0.3, 1)$	0.1809	0.3545	0.2606	0.5514	0.7263
$(1.5, -1, 0.3, 1)$	0.3721	0.6882	0.5556	0.8695	0.9386
$(2, -1, 0.3, 1)$	0.6094	0.9043	0.8126	0.9755	0.9920
$(1, 0, 0.3, 1)$	0.0797	0.1624	0.1103	0.1850	0.3453
$(1.5, 0, 0.3, 1)$	0.1349	0.3716	0.2486	0.4156	0.6121
$(2, 0, 0.3, 1)$	0.2374	0.6418	0.4751	0.6740	0.8180
$(1, 1, 0.3, 1)$	0.1816	0.3543	0.2604	0.0864	0.2026
$(1.5, 1, 0.3, 1)$	0.3652	0.6812	0.5494	0.2281	0.4161
$(2, 1, 0.3, 1)$	0.6131	0.9107	0.8188	0.4480	0.6580
$(1, -1, 3, 1)$	0.0666	0.0778	0.0798	0.1110	0.0946
$(1.5, -1, 3, 1)$	0.0909	0.1049	0.1001	0.1302	0.1191
$(2, -1, 3, 1)$	0.1334	0.1427	0.1404	0.1590	0.1435
$(1, 0, 3, 1)$	0.0460	0.0533	0.0559	0.0800	0.0675
$(1.5, 0, 3, 1)$	0.0510	0.0607	0.0623	0.0728	0.0680
$(2, 0, 3, 1)$	0.0535	0.0682	0.0707	0.0703	0.0684
$(1, 1, 3, 1)$	0.0654	0.0675	0.0673	0.0691	0.0637
$(1.5, 1, 3, 1)$	0.0900	0.0902	0.0883	0.0736	0.0823
$(2, 1, 3, 1)$	0.1309	0.1303	0.1261	0.0890	0.1081

Table 2: Empirical powers are reported for the proposed tests in Section 2.2.1, the Cramer-von Mises test (T_{CM}), the test proposed by Einmahl and McKeague (2003)(EM) and the test proposed by Jager and Wellner (2007)(JW) with nominal level 0.05 for Model 1, $n = 1000$ and the simple null hypothesis $H_0 : F \sim N(\mu, \sigma^2)$.

$(\delta, \mu, \sigma, \nu)$	T_{CM}	EM	JW	l_s	$\bar{l}_s$
$(0, -1, 0.3, 1)$	0.0492	0.0494	0.0497	0.0598	0.0515
$(1, -1, 0.3, 1)$	0.1613	0.4681	0.3869	0.8837	0.9067
$(1.5, -1, 0.3, 1)$	0.3502	0.8794	0.8052	0.9950	0.9961
$(2, -1, 0.3, 1)$	0.6131	0.9915	0.9773	1	1
$(1, 0, 0.3, 1)$	0.0787	0.2055	0.1699	0.4163	0.4769
$(1.5, 0, 0.3, 1)$	0.1199	0.5158	0.4209	0.7621	0.8090
$(2, 0, 0.3, 1)$	0.2098	0.8435	0.7538	0.9408	0.9555
$(1, 1, 0.3, 1)$	0.1639	0.4753	0.3993	0.2190	0.2787
$(1.5, 1, 0.3, 1)$	0.3557	0.8772	0.8078	0.4882	0.5544
$(2, 1, 0.3, 1)$	0.6333	0.9931	0.9791	0.7711	0.8073
$(1, -1, 3, 1)$	0.0727	0.0726	0.0718	0.0808	0.0642
$(1.5, -1, 3, 1)$	0.0936	0.0916	0.0892	0.0900	0.0741
$(2, -1, 3, 1)$	0.1302	0.1245	0.1229	0.1119	0.0919
$(1, 0, 3, 1)$	0.0517	0.0540	0.0555	0.0534	0.0502
$(1.5, 0, 3, 1)$	0.0532	0.0574	0.0576	0.0510	0.0507
$(2, 0, 3, 1)$	0.0617	0.0667	0.0667	0.0566	0.0539
$(1, 1, 3, 1)$	0.0678	0.0678	0.0671	0.0543	0.0623
$(1.5, 1, 3, 1)$	0.0945	0.0895	0.0899	0.0690	0.0856
$(2, 1, 3, 1)$	0.1347	0.1289	0.1299	0.0967	0.1230

Table 3: Empirical powers are reported for the proposed tests in Section 2.2.1, the Cramer-von Mises test (T_{CM}), the test proposed by Einmahl and McKeague (2003) (EM) and the test proposed by Jager and Wellner (2007) (JW) with nominal level 0.05 for Model 2, $n = 100$ and the simple null hypothesis $H_0 : F \sim LogNormal(\mu, \sigma^2)$.

$(\delta, \mu, \sigma, \nu)$	T_{CM}	EM	JW	l_s	$\bar{l}_s$
$(0, -1, 0.3, 0.5)$	0.0515	0.0501	0.0509	0.0902	0.0710
$(1, -1, 0.3, 0.5)$	0.2445	0.4465	0.3475	0.6746	0.8092
$(1.5, -1, 0.3, 0.5)$	0.4905	0.7824	0.6683	0.9278	0.9707
$(2, -1, 0.3, 0.5)$	0.7470	0.9452	0.8842	0.9894	0.9971
$(1, 0, 0.3, 0.5)$	0.1027	0.1895	0.1366	0.2652	0.4453
$(1.5, 0, 0.3, 0.5)$	0.1645	0.3861	0.2739	0.5267	0.7030
$(2, 0, 0.3, 0.5)$	0.2849	0.6453	0.5008	0.7739	0.8835
$(1, 1, 0.3, 0.5)$	0.1447	0.2532	0.1903	0.0533	0.0929
$(1.5, 1, 0.3, 0.5)$	0.3025	0.5345	0.4192	0.0824	0.1738
$(2, 1, 0.3, 0.5)$	0.4821	0.7663	0.6563	0.1319	0.2864
$(1, -1, 3, 0.5)$	0.0911	0.0916	0.0963	0.1728	0.1821
$(1.5, -1, 3, 0.5)$	0.1434	0.1435	0.1542	0.2240	0.2876
$(2, -1, 3, 0.5)$	0.2240	0.2268	0.2429	0.2897	0.4080
$(1, 0, 3, 0.5)$	0.0551	0.0693	0.0757	0.1219	0.1154
$(1.5, 0, 3, 0.5)$	0.0612	0.0895	0.1038	0.1353	0.1502
$(2, 0, 3, 0.5)$	0.0856	0.1322	0.1561	0.1514	0.1914
$(1, 1, 3, 0.5)$	0.0693	0.0785	0.0834	0.0963	0.0823
$(1.5, 1, 3, 0.5)$	0.1062	0.1219	0.1310	0.1356	0.1132
$(2, 1, 3, 0.5)$	0.1636	0.1865	0.2038	0.1705	0.1484

Table 4: Empirical powers are reported for the proposed tests in Section 2.2.1, the Cramer-von Mises test (T_{CM}), the test proposed by Einmahl and McKeague (2003) (EM) and the test proposed by Jager and Wellner (2007) (JW) with nominal level 0.05 for Model 2, $n = 1000$ and the simple null hypothesis $H_0 : F \sim LogNormal(\mu, \sigma^2)$.

$(\delta, \mu, \sigma, \nu)$	T_{CM}	EM	JW	l_s	$\bar{l}_s$
$(0, -1, 0.3, 0.5)$	0.0489	0.0485	0.0490	0.0578	0.0492
$(1, -1, 0.3, 0.5)$	0.2138	0.5911	0.5040	0.9492	0.9612
$(1.5, -1, 0.3, 0.5)$	0.4929	0.9485	0.8996	0.9990	0.9991
$(2, -1, 0.3, 0.5)$	0.7824	0.9986	0.9940	1	1
$(1, 0, 0.3, 0.5)$	0.0835	0.2061	0.1717	0.5216	0.5884
$(1.5, 0, 0.3, 0.5)$	0.1514	0.5365	0.4539	0.8663	0.8923
$(2, 0, 0.3, 0.5)$	0.2669	0.8465	0.7691	0.9823	0.9870
$(1, 1, 0.3, 0.5)$	0.1441	0.3245	0.2806	0.0869	0.1216
$(1.5, 1, 0.3, 0.5)$	0.2808	0.6905	0.6104	0.1720	0.2281
$(2, 1, 0.3, 0.5)$	0.4949	0.9322	0.8835	0.3094	0.3747
$(1, -1, 3, 0.5)$	0.0851	0.0854	0.0858	0.1649	0.1217
$(1.5, -1, 3, 0.5)$	0.1473	0.1441	0.1470	0.2724	0.2044
$(2, -1, 3, 0.5)$	0.2394	0.2346	0.2355	0.4128	0.3433
$(1, 0, 3, 0.5)$	0.0553	0.0625	0.0655	0.0992	0.0775
$(1.5, 0, 3, 0.5)$	0.0718	0.0860	0.0926	0.1360	0.1022
$(2, 0, 3, 0.5)$	0.1088	0.1379	0.1418	0.1881	0.1546
$(1, 1, 3, 0.5)$	0.0761	0.0782	0.0799	0.0768	0.0665
$(1.5, 1, 3, 0.5)$	0.1154	0.1201	0.1221	0.1035	0.0962
$(2, 1, 3, 0.5)$	0.1751	0.1873	0.1933	0.1527	0.1427

Table 5: Empirical powers are reported for the proposed tests in Section 2.2.2, the Cramer-von Mises test (T_{CM}), the test proposed by Einmahl and McKeague (2003) (EM) and the test proposed by Jager and Wellner (2007) (JW) with nominal level 0.05 for Model 1, $n = 100$ and the composite null hypothesis that F has a normal distribution.

$(\delta, \mu, \sigma, \nu)$	T_{CM}	EM	JW	l_c	$\bar{l}_c$
$(0, -1, 0.3, 1)$	0.0518	0.0516	0.0513	0.0967	0.1181
$(0.1, -1, 0.3, 1)$	0.3595	0.3824	0.3845	0.4340	0.4199
$(0.3, -1, 0.3, 1)$	0.7367	0.7627	0.7633	0.7985	0.7794
$(1, -1, 0.3, 1)$	0.9922	0.9942	0.9944	0.9941	0.9933
$(0.1, 0, 0.3, 1)$	0.3063	0.3205	0.3212	0.3731	0.3589
$(0.3, 0, 0.3, 1)$	0.6509	0.6750	0.6766	0.7157	0.6889
$(1, 0, 0.3, 1)$	0.9736	0.9789	0.9781	0.9818	0.9783
$(0.1, 1, 0.3, 1)$	0.3614	0.3814	0.3827	0.4310	0.3993
$(0.3, 1, 0.3, 1)$	0.7312	0.7596	0.7594	0.7914	0.7543
$(1, 1, 0.3, 1)$	0.9909	0.9930	0.9931	0.9553	0.9912
$(0.1, -1, 3, 1)$	0.0932	0.0980	0.0998	0.1419	0.1490
$(0.3, -1, 3, 1)$	0.1670	0.1278	0.1701	0.2165	0.2075
$(1, -1, 3, 1)$	0.3835	0.3983	0.3926	0.4582	0.4106
$(0.1, 0, 3, 1)$	0.0854	0.0852	0.0855	0.1333	0.1413
$(0.3, 0, 3, 1)$	0.1497	0.1573	0.1567	0.2083	0.2002
$(1, 0, 3, 1)$	0.3773	0.3933	0.3916	0.4375	0.4017
$(0.1, 1, 3, 1)$	0.0849	0.0867	0.0875	0.1318	0.1408
$(0.3, 1, 3, 1)$	0.1477	0.1567	0.1572	0.2031	0.2004
$(1, 1, 3, 1)$	0.3776	0.3953	0.3925	0.4156	0.3941

Table 6: Empirical powers are reported for the proposed tests in Section 2.2.2, the Cramer-von Mises test (T_{CM}), the test proposed by Einmahl and McKeague (2003) (EM) and the test proposed by Jager and Wellner (2007) (JW) with nominal level 0.05 for Model 1, $n = 1000$ and the composite null hypothesis that F has a normal distribution.

$(\delta, \mu, \sigma, \nu)$	T_{CM}	EM	JW	l_c	$\bar{l}_c$
$(0, -1, 0.3, 1)$	0.0527	0.0516	0.0522	0.0595	0.0596
$(0.1, -1, 0.3, 1)$	0.5954	0.6371	0.6378	0.6549	0.6348
$(0.3, -1, 0.3, 1)$	0.9595	0.9701	0.9700	0.9779	0.9751
$(1, -1, 0.3, 1)$	1	1	1	1	1
$(0.1, 0, 0.3, 1)$	0.5296	0.5468	0.5657	0.5625	0.5416
$(0.3, 0, 0.3, 1)$	0.9187	0.9370	0.9379	0.9438	0.9364
$(1, 0, 0.3, 1)$	1	1	1	1	1
$(0.1, 1, 0.3, 1)$	0.5911	0.6303	0.6324	0.6214	0.5923
$(0.3, 1, 0.3, 1)$	0.9577	0.9688	0.9695	0.9708	0.9642
$(1, 1, 0.3, 1)$	1	1	1	1	1
$(0.1, -1, 3, 1)$	0.1128	0.1181	0.1188	0.1258	0.1183
$(0.3, -1, 3, 1)$	0.2395	0.2570	0.2577	0.2693	0.2523
$(1, -1, 3, 1)$	0.6001	0.6300	0.6300	0.6603	0.6375
$(0.1, 0, 3, 1)$	0.1124	0.1175	0.1182	0.1245	0.1178
$(0.3, 0, 3, 1)$	0.2412	0.2567	0.2577	0.2628	0.2488
$(1, 0, 3, 1)$	0.6088	0.6335	0.6338	0.6471	0.6246
$(0.1, 1, 3, 1)$	0.1128	0.1175	0.1186	0.1231	0.1162
$(0.3, 1, 3, 1)$	0.2394	0.2581	0.2571	0.2587	0.2430
$(1, 1, 3, 1)$	0.5989	0.6334	0.6343	0.6254	0.6039

Table 7: Empirical powers are reported for the proposed tests in Section 2.2.2, the Cramer-von Mises test (T_{CM}), the test proposed by Einmahl and McKeague (2003) (EM) and the test proposed by Jager and Wellner (2007) (JW) with nominal level 0.05 for Model 2, $n = 100$ and the composite null hypothesis that F has a log-normal distribution.

$(\delta, \mu, \sigma, \nu)$	T_{CM}	EM	JW	l_c	$\bar{l}_c$
$(0, -1, 0.3, 1)$	0.0503	0.0516	0.0523	0.0988	0.1136
$(0.1, -1, 0.3, 1)$	0.3281	0.3675	0.3704	0.4348	0.4109
$(0.3, -1, 0.3, 1)$	0.7086	0.7485	0.7511	0.7960	0.7719
$(1, -1, 0.3, 1)$	0.9906	0.9935	0.9930	0.9960	0.9946
$(0.1, 0, 0.3, 1)$	0.1753	0.2039	0.2055	0.2695	0.2457
$(0.3, 0, 0.3, 1)$	0.4165	0.4663	0.4661	0.5553	0.4975
$(1, 0, 0.3, 1)$	0.8763	0.9054	0.9020	0.9411	0.9232
$(0.1, 1, 0.3, 1)$	0.2235	0.2419	0.2448	0.3036	0.2734
$(0.3, 1, 0.3, 1)$	0.5141	0.5512	0.5532	0.6077	0.5376
$(1, 1, 0.3, 1)$	0.9250	0.9397	0.9392	0.8616	0.9342
$(0.1, -1, 3, 1)$	0.0492	0.0496	0.0494	0.1047	0.1173
$(0.3, -1, 3, 1)$	0.0584	0.0574	0.0574	0.1130	0.1107
$(1, -1, 3, 1)$	0.1081	0.1054	0.0999	0.1758	0.1339
$(0.1, 0, 3, 1)$	0.0490	0.0490	0.0497	0.0990	0.1137
$(0.3, 0, 3, 1)$	0.0573	0.0557	0.0535	0.0994	0.0996
$(1, 0, 3, 1)$	0.0911	0.0889	0.0834	0.1193	0.0912
$(0.1, 1, 3, 1)$	0.0492	0.0488	0.0496	0.0969	0.1153
$(0.3, 1, 3, 1)$	0.0543	0.0572	0.0545	0.0887	0.1011
$(1, 1, 3, 1)$	0.0991	0.0929	0.0860	0.0910	0.0917

Table 8: Empirical powers are reported for the proposed tests in Section 2.2.2, the Cramer-von Mises test (T_{CM}), the test proposed by Einmahl and McKeague (2003) (EM) and the test proposed by Jager and Wellner (2007) (JW) with nominal level 0.05 for Model 2, $n = 1000$ and the composite null hypothesis that F has a log-normal distribution.

$(\delta, \mu, \sigma, \nu)$	T_{CM}	EM	JW	l_c	$\bar{l}_c$
$(0, -1, 0.3, 1)$	0.0544	0.0556	0.0563	0.0607	0.0616
$(0.1, -1, 0.3, 1)$	0.4272	0.5128	0.5173	0.5670	0.5370
$(0.3, -1, 0.3, 1)$	0.9094	0.9455	0.9473	0.9645	0.9598
$(1, -1, 0.3, 1)$	1	1	1	1	1
$(0.1, 0, 0.3, 1)$	0.1954	0.2367	0.2378	0.2862	0.2563
$(0.3, 0, 0.3, 1)$	0.5767	0.6641	0.6645	0.7512	0.7243
$(1, 0, 0.3, 1)$	0.9933	0.9970	0.9975	0.9991	0.9990
$(0.1, 1, 0.3, 1)$	0.3102	0.3566	0.3594	0.3382	0.3030
$(0.3, 1, 0.3, 1)$	0.7387	0.7951	0.7961	0.7837	0.7539
$(1, 1, 0.3, 1)$	0.9987	0.9995	0.9995	0.9993	0.9989
$(0.1, -1, 3, 1)$	0.0517	0.0520	0.0511	0.0605	0.0556
$(0.3, -1, 3, 1)$	0.0530	0.0535	0.0535	0.0643	0.0580
$(1, -1, 3, 1)$	0.1074	0.1042	0.1021	0.1148	0.0958
$(0.1, 0, 3, 1)$	0.0525	0.0511	0.0504	0.0599	0.0546
$(0.3, 0, 3, 1)$	0.0512	0.0500	0.0507	0.0593	0.0552
$(1, 0, 3, 1)$	0.0887	0.0840	0.0829	0.0899	0.0730
$(0.1, 1, 3, 1)$	0.0527	0.0507	0.0508	0.0579	0.0541
$(0.3, 1, 3, 1)$	0.0535	0.0541	0.0545	0.0585	0.0586
$(1, 1, 3, 1)$	0.0922	0.0860	0.0853	0.0772	0.0739

Table 9: Empirical powers are reported for the proposed tests in Section 2.2.1 and the Cramer-von Mises test (T_{CM}) with nominal level 0.05 for Model 3 and the simple null hypothesis.

$(\delta, \mu, \sigma, \nu)$	T_{CM}	l_s	$\bar{l}_s$
$(0, -1, 0.3, 1)$	0.0556	0.0570	0.0542
$(1, -1, 0.3, 1)$	0.0952	0.3518	0.3818
$(1.5, -1, 0.3, 1)$	0.1886	0.6672	0.6922
$(2, -1, 0.3, 1)$	0.3412	0.8788	0.8936
$(1, 0, 0.3, 1)$	0.0732	0.2848	0.3130
$(1.5, 0, 0.3, 1)$	0.1082	0.5740	0.6030
$(2, 0, 0.3, 1)$	0.1804	0.8108	0.8278
$(1, 1, 0.3, 1)$	0.1028	0.1604	0.1804
$(1.5, 1, 0.3, 1)$	0.1678	0.3308	0.3590
$(2, 1, 0.3, 1)$	0.3058	0.5436	0.5678
$(1, -1, 3, 1)$	0.0506	0.0746	0.0852
$(1.5, -1, 3, 1)$	0.0590	0.1360	0.1524
$(2, -1, 3, 1)$	0.0694	0.2068	0.2232
$(1, 0, 3, 1)$	0.0564	0.0690	0.0806
$(1.5, 0, 3, 1)$	0.0600	0.1186	0.1378
$(2, 0, 3, 1)$	0.0786	0.1886	0.2104
$(1, 1, 3, 1)$	0.0656	0.0634	0.0708
$(1.5, 1, 3, 1)$	0.0806	0.1018	0.1184
$(2, 1, 3, 1)$	0.1216	0.1574	0.1810

Table 10: Empirical powers are reported for the proposed tests in Section 2.2.2 and the Cramer-von Mises test (T_{CM}) with nominal level 0.05 for Model 3 and the composite null hypothesis.

$(\delta, \mu, \sigma, \nu)$	T_{CM}	l_c	$\bar{l}_c$
$(0, -1, 0.3, 1)$	0.0570	0.0540	0.0552
$(0.1, -1, 0.3, 1)$	0.5536	0.5854	0.5734
$(1, -1, 0.3, 1)$	1	1	1
$(0.1, 0, 0.3, 1)$	0.5066	0.5050	0.4934
$(1, 0, 0.3, 1)$	0.9996	0.9996	0.9996
$(0.1, 1, 0.3, 1)$	0.5518	0.5248	0.5106
$(1, 1, 0.3, 1)$	1	1	0.9998
$(0.1, -1, 3, 1)$	0.2126	0.2142	0.2086
$(1, -1, 3, 1)$	0.9176	0.9242	0.9204
$(0.1, 0, 3, 1)$	0.2132	0.2160	0.2102
$(1, 0, 3, 1)$	0.9140	0.9306	0.9270
$(0.1, 1, 3, 1)$	0.2144	0.2172	0.2118
$(1, 1, 3, 1)$	0.9124	0.9330	0.9288

CHAPTER III

JACKKNIFE EMPIRICAL LIKELIHOOD TEST FOR REGRESSION MODELS

This chapter is based on the published paper:

H. Feng and L. Peng (2012). Jackknife Empirical Likelihood Tests For Error Distributions in Regression Models. Journal of Multivariate Analysis 112, 63–75.

3.1 Introduction

Let Y and X denote a univariate response and a d-variate covariate, respectively. For modeling the relationship between Y and X, a widely employed tool is the regression model $Y = m(X;\alpha) + \epsilon$, where m is a known function depending on a q-dimensional unknown parameter α and ϵ is a random error with mean zero. Suppose $\{(X_i^T, Y_i)^T\}_{i=1}^n$ is a random sample from this regression model, i.e.,

$$Y_i = m(X_i;\alpha) + \epsilon_i, \quad i = 1, \cdots, n, \tag{33}$$

where $\epsilon_1, \cdots, \epsilon_n$ are independent and identically distributed random variables with zero mean, $X_1, \cdots, X_n$ are independent and identically distributed random variables and independent of $\epsilon_i' s$. A standard way to estimate the unknown parameter α is the least squares estimate

$$\hat{\alpha} = \arg\min_{\alpha} \sum_{i=1}^{n} \{Y_i - m(X_i;\alpha)\}^2,$$

which says that $\hat{\alpha}$ is a solution of the following score equations

$$\sum_{i=1}^{n}\{Y_i - m(X_i;\alpha)\}\frac{\partial}{\partial\alpha}m(X_i;\alpha) = 0. \tag{34}$$

In some applications such as predicting conditional Value-at-Risk in risk management, it is useful to fit a parametric distribution family to the random error ϵ_i so as to improve the accuracy of inference. This results in a corresponding parametric distribution family for the conditional distribution of Y_i given X_i. See Chapters 8, 9 and 10 in Davison (2008) for more details on regression models. Here we are interested in testing whether the distribution of ϵ_i follows from a particular parametric family, i.e., test $H_0 : F_\epsilon \in \mathcal{F} = \{F(\cdot;\beta) : \beta \in \Omega \subset R^s\}$, where F_ϵ denotes the distribution of ϵ_i, and $F(\cdot;\beta)$ denotes a distribution function depending on the parameter β. Obviously one can simply employ some classical goodness-of-fit tests to the estimated errors $\hat{\epsilon}_i = Y_i - m(X_i;\hat{\alpha}), i = 1, \cdots, n$. More specifically, one can estimate β first by using the maximum likelihood estimate $\hat{\beta}$ based on $\hat{\epsilon}_1, \cdots, \hat{\epsilon}_n$, and then consider either the Kolmogorov-Smirnov test

$$T_1 = \sup_z \sqrt{n}|\frac{1}{n}\sum_{i=1}^{n} I(\hat{\epsilon}_i \le z) - F(z;\hat{\beta})|$$

or the Cramér-von-Mises test

$$\begin{aligned} T_2 &= n\int_{-\infty}^{\infty}\{\tfrac{1}{n}\textstyle\sum_{i=1}^{n} I(\hat{\epsilon}_i \le z) - F(z;\hat{\beta})\}^2\, dF(z;\hat{\beta}) \\ &= \tfrac{1}{12n} + \textstyle\sum_{i=1}^{n}\{\tfrac{2i-1}{2n} - F(\hat{\epsilon}_{n,i};\hat{\beta})\}^2, \end{aligned}$$

where $\hat{\epsilon}_{n,1} \le \cdots \le \hat{\epsilon}_{n,n}$ denote the order statistics of $\hat{\epsilon}_1, \cdots, \hat{\epsilon}_n$ (see D'Agostino and Stephens (1986)). Due to the plug-in estimators $\hat{\alpha}$ in $\hat{\epsilon}_i's$ and $\hat{\beta}$, the limiting distributions of T_1 and T_2 become quite complicated, which depend on the underlying distribution and thus are no longer distribution free. Hence some ad hoc procedure such as bootstrap method is needed in order to calculate the critical values.

Recently, Khmaladze and Koul (2004) proposed a new goodness-of-fit test via martingale transforms for testing the error distribution. It turns out that the new test statistic is asymptotically distribution free in testing a simple null hypothesis or a composite null hypothesis with a scale distribution family, and hence critical values can be tabulated. Some numeric analyses are given in Khmaladze and Koul (2004) and Koul and Sakhanenko (2005) for the Kolmogorov-Smirnov type of test. However, when the Cramér-von-Mises type of test is concerned, the calculation of the proposed test in Khmaladze and Koul (2004) becomes quite complicated, which requires to evaluate some integrals numerically.

In this chapter, we propose a novel jackknife empirical likelihood test for testing the error distribution in the regression model (33). It turns out that the asymptotic distribution of the new test has a chi-square limit, and the calculation of the test statistic is quite straightforward and involves no numeric integration. As a powerful tool in interval estimation and hypothesis test, empirical likelihood method has been applied to many different settings. We refer to Owen (2001) for an overview. Some advantages of empirical likelihood method include that the shape of confidence interval/region is determined by the sample automatically. When the empirical likelihood method is applied to nonlinear functionals directly, Wilks theorem fails in general, i.e., the limit is no longer a chi-square distribution. To overcome this difficulty, Jing, Yuan and Zhou (2009) proposed to apply the empirical likelihood method to some jackknife sample constructed from the targeted nonlinear functionals. This is called jackknife empirical likelihood method. A smoothed jackknife empirical likelihood method is applied to copulas, tail copulas and ROC curves; see Gong, Peng and Qi (2010), Peng and Qi (2010), Peng, Qi and Van Keilegom (2011).

We organize this chapter as follows. Methodology and main asymptotic results are given in Section 2. Section 3 presents a simulation study. Proofs are put in Section 4.

3.2 Methodology

To motivate our new method, we assume $\epsilon_i's$ are observable and β is known for the time being. That is, we want to test $H_0 : F_\epsilon(x) \equiv F(x;\beta)$. This is equivalent to test $H_0 : \int_{-\infty}^{\infty} \{F_\epsilon(x) - F(x;\beta)\}^2 dF(x;\beta) = 0$, which results in the Cramér-von-Mises test when $F_\epsilon(x)$ is replaced by the empirical distribution function based on $\epsilon_1, \cdots, \epsilon_n$.

By noting that $H_0 : \int_{-\infty}^{\infty} \{F_\epsilon(x) - F(x;\beta)\}^2 \, dF(x;\beta) = 0$ is equivalent to

$$H_0 : E \int_{-\infty}^{\infty} \{I(\epsilon_1 \vee \epsilon_2 \leq x) - 2I(\epsilon_1 \leq x)F(x;\beta) + F^2(x;\beta)\} \, dF(x;\beta) = 0,$$

i.e.,

$$H_0 : E\{F^2(\epsilon_1;\beta) - F(\epsilon_1 \vee \epsilon_2;\beta) + 1/3\} = 0,$$

one can directly apply the empirical likelihood method to the above estimating equation based on sample $\{(\epsilon_i, \epsilon_{i+k})^T\}_{i=1}^k$, where $k = [n/2]$. More specifically, by defining the empirical likelihood function as

$$\begin{aligned} L(\beta) \; &= \sup\{\textstyle\prod_{i=1}^k (kp_i) : p_1 \geq 0, \cdots, p_k \geq 0, \sum_{i=1}^k p_i = 1, \\ &\textstyle\sum_{i=1}^k p_i (\frac{F^2(\epsilon_i;\beta) + F^2(\epsilon_{i+k};\beta)}{2} - F(\epsilon_i \vee \epsilon_{i+k};\beta) + 1/3) = 0\}, \end{aligned}$$

it follows from Owen (1988) that $-2\log L(\beta)$ converges in distribution to a chi-square limit with one degree of freedom under H_0. Hence, one can use the empirical likelihood ratio test statistic $-2\log L(\beta)$ to test $H_0 : F_\epsilon(x) \equiv F(x;\beta)$. Unfortunately, this test is not powerful since

$$E\int_{-\infty}^{\infty}\{I(\epsilon_1\vee\epsilon_2\le x)-2I(\epsilon_1\le x)F(x;\beta)+F^2(x;\beta)\}\,dF(x;\beta)=O(\delta^2)$$

rather than $O(\delta)$ when $\sup_x|F_\epsilon(x)-F(x;\beta)|=O(\delta)$. To overcome this difficulty, we propose to apply the empirical likelihood method to the following two equations:

$$\begin{cases} E\{F^2(\epsilon_1;\beta)-F(\epsilon_1\vee\epsilon_2;\beta)+1/3\}=0 \\ EF(\epsilon_1;\beta)-2EF^3(\epsilon_1;\beta)=0. \end{cases} \tag{35}$$

Note that Li and Peng (2011) proposed to employ different estimating equations when $\epsilon_i's$ are observable and β is either known or unknown. We remark that the second equation in (35) can be replaced by some other linear estimating equations. Hence this new method is quite flexible and easy in taking more relevant constraints into account.

Now we are ready to extend the above idea to test the error distribution in the regression model (33). We consider the cases of simple null hypothesis and composite null hypothesis separately. Throughout we assume that α_0 and β_0 denote the true values of α and β respectively.

3.2.1 Simple null hypothesis

In this subsection, we are interested in testing $H_0: F_\epsilon(x)\equiv F(x;\beta_0)$ under model (33).

Put $k=[\frac{n}{2}]$ and define $\epsilon_i(\alpha)=Y_i-m(X_i;\alpha)$, $\tilde\epsilon_i(\alpha)=Y_{k+i}-m(X_{k+i};\alpha)$, $\epsilon_i^*(\alpha)=\max(\epsilon_i(\alpha),\tilde\epsilon_i(\alpha))$ and $h_i(\alpha)=\frac{\partial}{\partial\alpha}\{\epsilon_i^2(\alpha)+\tilde\epsilon_i^2(\alpha)\}$ for $i=1,\cdots,k$. Therefore the least squares estimator $\hat\alpha$ of α is defined as a solution to the equation $\sum_{i=1}^k h_i(\alpha)=0$.

Unfortunately we can not directly apply the empirical likelihood method to equations (35) based on the sample $\{(\epsilon_i(\hat\alpha),\tilde\epsilon_i(\hat\alpha))\}_{i=1}^k$ since this fails to catch the variance of $\hat{\alpha}$. Generally speaking, Wilks theorem does not hold when the empirical likelihood method is applied to nonlinear functionals directly. Motivated by the recent jackknife empirical likelihood method in Jing, Yuan and Zhou (2009), we propose to apply the empirical likelihood method to some jackknife pseudo sample. In order to formulate the jackknife sample, it follows from the idea in Jing, Yuan and Zhou (2009) that $\hat{\alpha}_j$ is the solution of the leave-jth item equation $\sum_{i=1,i\neq j}^{k} h_i(\alpha) = 0$ for each $j = 1, \cdots, k$. When m is a nonlinear function, the above equation does not admit an explicit solution in general. Therefore, the above way of formulating jackknife sample is computationally intensive. Here we propose to app ly the approximate jackknife empir ical likelihood method in Peng (2011) as follows.

Note that

$$\begin{aligned} 0 &= \sum_{j=1,j\neq i}^{k} h_j(\alpha) \\ &= \sum_{j=1,j\neq i}^{k} h_j(\alpha) - \sum_{j=1}^{k} h_j(\hat{\alpha}) \\ &= \sum_{j=1}^{k}\{h_j(\alpha) - h_j(\hat{\alpha})\} - h_i(\alpha) \\ &\approx \sum_{j=1}^{k}\{\tfrac{\partial}{\partial\alpha^T} h_j(\hat{\alpha})\}\{\alpha - \hat{\alpha}\} - h_i(\hat{\alpha}). \end{aligned} \tag{36}$$

Instead of solving $0 = \sum_{j=1,j\neq i}^{k} h_j(\alpha)$, we propose to approximate the solution by

$$\hat{\alpha}_i = \hat{\alpha} + \{\frac{1}{k}\sum_{j=1}^{k} \frac{\partial}{\partial\alpha^T} h_j(\hat{\alpha})\}^{-1} \frac{1}{k} h_i(\hat{\alpha}).$$

Using equation (35), $\hat{\alpha}$ and $\hat{\alpha}_i's$, we define the approximate jackknife sample as

$$\begin{aligned} G_1(i) &= \sum_{j=1}^{k}\{\tfrac{F^2(\epsilon_j(\hat{\alpha});\beta_0)+F^2(\tilde{\epsilon_j}(\hat{\alpha});\beta_0)}{2} - F(\epsilon_j^*(\hat{\alpha});\beta_0) + 1/3\} \\ &\quad - \sum_{j=1,j\neq i}^{k}\{\tfrac{F^2(\epsilon_j(\hat{\alpha}_i);\beta_0)+F^2(\tilde{\epsilon_j}(\hat{\alpha}_i);\beta_0)}{2} - F(\epsilon_j^*(\hat{\alpha}_i);\beta) + 1/3\} \end{aligned}$$

and

$$\begin{aligned} G_2(i) &= \sum_{j=1}^k \{F(\epsilon_j(\hat{\alpha});\beta_0) + F(\tilde{\epsilon}_j(\hat{\alpha});\beta_0)\} \\ &\quad - \sum_{j=1,j\neq i}^k \{F(\epsilon_j(\hat{\alpha}_i);\beta_0) + F(\tilde{\epsilon}_j(\hat{\alpha}_i);\beta_0)\} \\ &\quad -2\sum_{j=1}^k \{F^3(\epsilon_j(\hat{\alpha});\beta_0) + F^3(\tilde{\epsilon}_j(\hat{\alpha});\beta_0)\} \\ &\quad +2\sum_{j=1,j\neq i}^k \{F^3(\epsilon_j(\hat{\alpha}_i);\beta_0) + F^3(\tilde{\epsilon}_j(\hat{\alpha}_i);\beta_0)\} \end{aligned}$$

for $i = 1, \cdots, k$. Based on the above approximate jackknife sample, we define the jackknife empirical likelihood function as

$$L_n^J = \sup\{\prod_{i=1}^k (kp_i) : p_1 \geq 0, \cdots, p_k \geq 0, \sum_{i=1}^k p_i = 1, \sum_{i=1}^k p_i G(i) = 0\}$$

where $G(i) = (G_1(i), G_2(i))^T$. By the Lagrange multiplier technique, we have

$$l_n^J := -2\log L_n^J = 2\sum_{i=1}^k \log\{1 + \lambda^T G(i)\},$$

where λ satisfies

$$\sum_{i=1}^k \frac{G(i)}{1 + \lambda^T G(i)} = 0. \tag{37}$$

Before proving that Wilks theorem holds for the above jackknife empirical likelihood test, we list some regularity conditions:

- A1) there are a neighborhood of α_0, say Ω_0 and a function $K(x)$ such that $EK(X_1) < \infty$, where $\{(X_i^T, Y_i)^T\}_{i=1}^n$ are iid random samples from the regression model (1), and

  $$\sup_{\alpha\in\Omega_0}\{|\frac{\partial}{\partial\alpha_i}m(x;\alpha)| + |\frac{\partial^2}{\partial\alpha_i\partial\alpha_j}m(x;\alpha)| + |\frac{\partial^3}{\partial\alpha_i\partial\alpha_j\partial\alpha_l}m(x;\alpha)|\} \leq K(x)$$

  for $1 \leq i, j, l \leq q$;

- A2) $E\frac{\partial}{\partial\alpha^T}h_1(\alpha_0)$ is invertible;

- A3) $\sup_{y\in\Omega_1}|F''(y;\beta_0)| < \infty$, where Ω_1 denotes the support of ϵ_1.

Theorem 4. *Suppose model (33) holds with* $E\epsilon_i = 0$ *and* $E\epsilon_i^{2+\delta_0} < \infty$ *for some* $\delta_0 > 0$*. Further assume conditions A1)-A3) hold. Then, under* $H_0 : F_\epsilon(x) \equiv F(x; \beta_0)$*, we have* $l_n^J \xrightarrow{d} \chi^2(2)$ *as* $n \to \infty$*.*

Using Theorem 4, a jackknife empirical likelihood test for testing $H_0 : F_\epsilon(x) \equiv F(x; \beta_0)$ against $H_a : F_\epsilon(x) \not\equiv F(x; \beta_0)$ can be constructed, which rejects H_0 when $l_n^J \geq \chi^2_{2,1-\gamma}$, where γ is the significance level and $\chi^2_{2,1-\gamma}$ denotes the $(1-\gamma)$-th quantile of a chi-square distribution with two degrees of freedom.

3.2.2 Composite null hypothesis

In this subsection, we are interested in testing $H_0 : F_\epsilon \in \mathcal{F} = \{F(\cdot; \beta) : \beta \in \Omega \subset \mathcal{R}^s\}$ against $H_a : F_\epsilon \notin \mathcal{F}$ under model (33).

Define

$$\bar{h}_i(\alpha, \beta) = \frac{\partial}{\partial \beta} \log f(\epsilon_i(\alpha); \beta) + \frac{\partial}{\partial \beta} \log f(\tilde{\epsilon}_i(\alpha); \beta)$$

for $i = 1, \cdots, k$, where $f(x; \beta) = \frac{\partial}{\partial x} F(x; \beta)$. Next we estimate β by solving the score equation $\sum_{i=1}^k \bar{h}_i(\hat{\alpha}, \beta) = 0$, and denote the solution by $\hat{\beta}$. Although one may prefer to estimate α and β simultaneously by solving the equations

$$\sum_{i=1}^k \frac{\partial}{\partial \alpha} \{\log f(\epsilon_i(\alpha); \beta) + \log f(\tilde{\epsilon}_i(\alpha); \beta)\} = 0 \quad \text{and} \quad \sum_{i=1}^k \bar{h}_i(\alpha, \beta) = 0,$$

we propose to estimate them separately, which has less computation in general. In order to formulate the jackknife sample, one needs to solve $\sum_{i=1, i \neq j}^k \bar{h}_i(\hat{\alpha}_j, \beta) = 0$ for each $j = 1, \cdots, k$. Like (36), we have

$$\begin{aligned} 0 &= \sum_{i=1, i\neq j}^{k} \bar{h}_i(\hat{\alpha}_j, \beta) \\ &= \sum_{i=1, i\neq j}^{k} \bar{h}_i(\hat{\alpha}_j, \beta) - \sum_{i=1}^{k} \bar{h}_i(\hat{\alpha}, \hat{\beta}) \\ &= \sum_{i=1, i\neq j}^{k} \bar{h}_i(\hat{\alpha}_j, \beta) - \sum_{i=1}^{k} \bar{h}_i(\hat{\alpha}_j, \hat{\beta}) + \sum_{i=1}^{k} \bar{h}_i(\hat{\alpha}_j, \hat{\beta}) - \sum_{i=1}^{k} \bar{h}_i(\hat{\alpha}, \hat{\beta}) \\ &= \sum_{i=1}^{k} (\bar{h}_i(\hat{\alpha}_j, \beta) - \bar{h}_i(\hat{\alpha}_j, \hat{\beta})) + \sum_{i=1}^{k} (\bar{h}_i(\hat{\alpha}_j, \hat{\beta}) - \bar{h}_i(\hat{\alpha}, \hat{\beta})) - \bar{h}_j(\hat{\alpha}_j, \beta) \\ &\approx \sum_{i=1}^{k} \{\tfrac{\partial}{\partial \beta} \bar{h}_i(\hat{\alpha}_j, \hat{\beta})\}(\beta - \hat{\beta}) + \sum_{i=1}^{k} \{\tfrac{\partial}{\partial \alpha} \bar{h}_i(\hat{\alpha}, \hat{\beta})\}(\hat{\alpha}_j - \hat{\alpha}) - \bar{h}_j(\hat{\alpha}, \hat{\beta}). \end{aligned}$$

Thus, instead of solving $\sum_{i=1, i\neq j}^{k} \bar{h}_i(\hat{\alpha}_j, \beta) = 0$, we propose to approximate the solution by

$$\begin{aligned} \hat{\beta}_j &= \hat{\beta} + \{\tfrac{1}{k} \sum_{i=1}^{k} \tfrac{\partial}{\partial \beta^T} \bar{h}_i(\hat{\alpha}, \hat{\beta})\}^{-1} \tfrac{1}{k} \bar{h}_j(\hat{\alpha}, \hat{\beta}) \\ &\quad - \{\tfrac{1}{k} \sum_{i=1}^{k} \tfrac{\partial}{\partial \beta^T} \bar{h}_i(\hat{\alpha}, \hat{\beta})\}^{-1} \{\tfrac{1}{k} \sum_{i=1}^{k} \tfrac{\partial}{\partial \alpha^T} \bar{h}_i(\hat{\alpha}, \hat{\beta})\}(\hat{\alpha}_j - \hat{\alpha}) \end{aligned}$$

for $j = 1, \cdots, k$.

Based on $\hat{\alpha}, \hat{\alpha}_i, \hat{\beta}, \hat{\beta}_i$ and (35), we formulate the approximate jackknife sample as

$$\begin{aligned} \bar{G}_1(i) &= \sum_{j=1}^{k} \{\tfrac{F^2(\epsilon_j(\hat{\alpha}); \hat{\beta}) + F^2(\tilde{\epsilon}_j(\hat{\alpha}); \hat{\beta})}{2} - F(\epsilon_j^*(\hat{\alpha}); \hat{\beta}) + 1/3\} \\ &\quad - \sum_{j=1, j\neq i}^{k} \{\tfrac{F^2(\epsilon_j(\hat{\alpha}_i); \hat{\beta}_i) + F^2(\tilde{\epsilon}_j(\hat{\alpha}_i); \hat{\beta}_i)}{2} - F(\epsilon_j^*(\hat{\alpha}_i); \hat{\beta}_i) + 1/3\} \end{aligned}$$

and

$$\begin{aligned} \bar{G}_2(i) &= \sum_{j=1}^{k} \{F(\epsilon_j(\hat{\alpha}); \hat{\beta}) + F(\tilde{\epsilon}_j(\hat{\alpha}); \hat{\beta})\} \\ &\quad - \sum_{j=1, j\neq i}^{k} \{F(\epsilon_j(\hat{\alpha}_i); \hat{\beta}_i) + F(\tilde{\epsilon}_j(\hat{\alpha}_i); \hat{\beta}_i)\} \\ &\quad - 2\sum_{j=1}^{k} \{F^3(\epsilon_j(\hat{\alpha}); \hat{\beta}) + F^3(\tilde{\epsilon}_j(\hat{\alpha}); \hat{\beta})\} \\ &\quad + 2\sum_{j=1, j\neq i}^{k} \{F^3(\epsilon_j(\hat{\alpha}_i); \hat{\beta}_i) + F^3(\tilde{\epsilon}_j(\hat{\alpha}_i); \hat{\beta}_i)\} \end{aligned}$$

for $i = 1, \cdots, k$. Based on the above approximate jackknife sample, the jackknife empirical likelihood function is defined as

$$\bar{L}_n^J = \sup\{\prod_{i=1}^{k} (kp_i) : p_1 \geq 0, \cdots, p_k \geq 0, \sum_{i=1}^{k} p_i = 1, \sum_{i=1}^{k} p_i \bar{G}(i) = 0\} \tag{38}$$

where $\bar{G}(i) = (\bar{G}_1(i), \bar{G}_2(i))^T$. By the Lagrange multiplier technique, we have

$$\bar{l}_n^J := -2\log \bar{L}_n^J = 2\sum_{i=1}^k \log\{1 + \bar{\lambda}^T \bar{G}(i)\},$$

where $\bar{\lambda}$ satisfies

$$\sum_{i=1}^k \frac{\bar{G}(i)}{1 + \bar{\lambda}^T \bar{G}(i)} = 0. \tag{39}$$

Before showing that Wilks theorem holds for the above jackknife empirical likelihood method, we list some regularity conditions:

- A4) there are a neighborhood of β_0, say Ω_2, and a function $\bar{K}(\cdot)$ such that $E\bar{K}(\epsilon_1(\alpha_0), \tilde{\epsilon}_1(\alpha_0), X_1, X_{k+1}) < \infty$ and

  $$\sup_{\alpha\in\Omega_0,\beta\in\Omega_2}\{|\tfrac{\partial}{\partial\theta_i}\bar{h}_1(\alpha,\beta)| + |\tfrac{\partial^2}{\partial\theta_i\partial\theta_j}\bar{h}_1(\alpha,\beta)|$$
  $$+|\tfrac{\partial^3}{\partial\theta_i\partial\theta_j\partial\theta_l}\bar{h}_1(\alpha,\beta)|\} \le \bar{K}(\epsilon_1(\alpha_0), \tilde{\epsilon}_1(\alpha_0), X_1, X_{k+1}),$$

  where $\theta = (\alpha^T, \beta^T)^T$ and $1 \le i, j, l \le q + s$;

- A5) $E\frac{\partial}{\partial\beta^T}\bar{h}_1(\alpha_0, \beta_0)$ is invertible;

- A6) $\sup_{y\in\Omega_3}\sup_{\beta\in\Omega_2}|\frac{\partial^2}{\partial\bar{\theta}^2}F(y;\beta)| < \infty$, where $\bar{\theta} = (y, \beta^T)^T$ and Ω_3 denotes the support of ϵ_i which is independent of β.

Theorem 5. *Suppose model (33) hold with $E\epsilon_i = 0$ and $E\epsilon_i^{2+\delta_0} < \infty$ for some $\delta_0 > 0$. Further assume A1)–A2) and A4)–A6) hold. Then $\bar{l}_n^J \xrightarrow{d} \chi^2(2)$ as $n \to \infty$.*

As before, Theorem 5 can be employed to test $H_0 : F_\epsilon \in \mathcal{F}$ against $H_a : F_\epsilon \notin \mathcal{F}$.

Remark 1. Theorems 4 and 5 still hold when estimators for α and β are replaced by solving some other estimating equations.

3.3 Simulation Study

In this section, we investigate the finite sample behavior of the proposed jackknife empirical likelihood test and compare it with the Cramér-von-Mises test. Since the test in Khmaladze and Koul (2004) is hard to implement for the type of Cramér-von Mises test and only applicable to testing a simple null hypothesis or a composite null hypothesis with a scale distribution family, we do not compare our new test with it.

Consider the model $Y_i = \exp(\alpha X_i) + \epsilon_i$ in Section 7 of Khmaladze and Koul (2004) with $\alpha = 0.25$ and $X_i \sim Uniform(2, 4)$. We draw $10,000$ random samples of size $n = 200$ and 500 from the above model with either

$$F_\epsilon(x) = (1 - \frac{\delta}{\sqrt{n}})N(0,1) + \frac{\delta}{\sqrt{n}}t(\nu) \tag{40}$$

or

$$F_\epsilon(x) = (1 - \frac{\delta}{\sqrt{n}})t(\nu) + \frac{\delta}{\sqrt{n}}N(0,1) \tag{41}$$

for $\delta = 0, 0.5, 1, 2, 3$.

The aim is to test either $H_0 : \epsilon_i \sim N(0,1)$ or $H_0 : \epsilon_i \sim t(3)$ or $H_0 : \epsilon_i \sim t(8)$ or $H_0 : F_\epsilon \in \mathcal{F}^n = \{N(0,\sigma^2) : \sigma > 0\}$ or $H_0 : F_\epsilon \in \mathcal{F}^t = \{t(\nu) : \nu > 2\}$. In case of composite null hypothesis, β equals either σ or ν and $\hat{\beta}$ is the corresponding moment estimator based on the estimated errors $\hat{\epsilon}_i's$. For computing the power of the Cramér-von-Mises test, a parametric bootstrap method with repetition $1,000$ is employed to obtain the critical values. More specifically, we generate $1,000$ random samples with size n from F_ϵ in case of simple null hypothesis or $F_\epsilon(;\hat{\beta})$ in case of composite hypothesis. Denote them by $\{\epsilon_i^{*(j)}\}_{i=1}^n$ for $j = 1, \cdots, 1000$. For each $j = 1, \cdots, 1000$, we further generate a bootstrap sample

$$Y_i^{*(j)} = m(X_i; \hat{\alpha}) + \epsilon_i^{*(j)} \quad \text{for} \quad i = 1, \cdots, n.$$

Based on $\{(X_i, Y_i^{*(j)})^T\}_{i=1}^n$ for each $j = 1, \cdots, 1000$, we compute the corresponding least squares estimator for α, the moment estimator for β in case of composite null hypothesis and estimated errors, say $\hat{\alpha}^{*(j)}, \hat{\beta}^{*(j)}, \{\hat{\epsilon}_i^{*(j)}\}_{i=1}^n$. Using these bootstrap quantities, we obtained 1000 bootstrapped Cramér-von Mises test statistics, which give the critical values. Note that Koul and Sakhanenko (2005) employed the naive bootstrap method, i.e., resampling from the estimated errors nonparametrically, for obtaining critical values for the Kolmogorov-Smirnov test. Since we are testing a parametric distribution family for ϵ_i, it prefers to employing the parametric bootstrap method.

The empirical sizes and powers of the proposed jackknife empirical likelihood method and the Cramér-von-Mises test are reported in Tables 11–14. From these tables, we observe that (i) results for $\delta = 0$ show that the size of the proposed jackknife empirical likelihood method is close to the nominal level and its accuracy is improved when the sample size becomes large; (ii) results for $\delta = 0.5, 1, 2, 3$ show that the proposed jackknife empirical likelihood method is more powerful than the Cramér-von-Mises test for most cases, especially for simple null hypothesis; (iii) both tests almost have no power for testing $H_0 : F_\epsilon \in \mathcal{F}^n$ when δ is not large. Second we consider the case of small sample size by drawing $10,000$ random samples with size $n = 50$ and 100 from the above model. It turns out that the size of the proposed jackknife empirical likelihood method is larger than the nominal level for $n = 50$. Hence we propose the following bootstrap calibration method. More details on calibration for empirical likelihood methods can be found in Owen (2001).

Draw $1,000$ resamples from $\{(\epsilon_i(\hat{\alpha}), \tilde{\epsilon}_i(\hat{\alpha}))\}_{i=1}^k$ with size $k = [n/2]$, say $\{(\epsilon_i^{*(b)}(\hat{\alpha}), \tilde{\epsilon}_i^{*(b)}(\hat{\alpha}))\}_{i=1}^k$ for $b = 1, \cdots, 1000$. For each resample $\{(\epsilon_i^{*(b)}(\hat{\alpha}),$

$\tilde{\epsilon}_i^{*(b)}(\hat{\alpha}))\}_{i=1}^k$, we use the model (33) to generate a resample

$$Y_i^{*(b)} = m(X_i; \hat{\alpha}) + \epsilon_i^{*(b)}, \quad Y_{k+i}^{*(b)} = m(X_{k+i}; \hat{\alpha}) + \tilde{\epsilon}_i^{*(b)}$$

for $i = 1, \cdots, k$. Next based on $\{(X_i, Y_i^{*(b)})\}_{i=1}^{2k}$, we re-estimate the parameters and calculate the jackknife empirical likelihood function, which results in $1,000$ jackknife empirical likelihood functions. Therefore, the bootstrap calibrated jackknife empirical likelihood test is computed by obtaining critical values from the computed $1,000$ jackknife empirical likelihood functions instead of the chi-square distribution with two degrees of freedom.

In Tables 15 and 16 we report the empirical sizes and powers of the proposed jackknife empirical likelihood method, its bootstrap calibrated version and the Cramér-von-Mises test. From these two tables we observe that i) the size of the jackknife empirical likelihood test is larger than the nominal level for $n = 50$, but gets more accurate when $n = 100$; ii) the size of the bootstrap calibrated jackknife empirical likelihood test is comparable with that of the Cramér-von-Mises test; iii) for testing t distributions, the bootstrap calibrated jackknife empirical likelihood test is more powerful than the Cramér-von-Mises test for simple null hypothesis, but less powerful for composite null hypothesis; iv) both the bootstrap calibrated jackknife empirical likelihood test and the Cramér-von-Mises test perform similar for testing normal distributions; v) for sample size $n = 100$, the jackknife empirical likelihood test has a reasonably accurate size and is most powerful.

3.4 Conclusions

We propose some jackknife empirical likelihood methods to test whether the error distribution in a regression model belongs to a particular parametric family. Unlike classical goodness-of-fit tests, the new tests always have a chi-square limit and so no ad hoc techniques such as bootstrap method are needed to obtain critical values. Also the calculation of the proposed tests is quite straightforward and involves no numeric integration unlike the method in Khmaladze and Koul (2004). When the sample size is small ($n = 50$), the sizes of the jackknife empirical likelihood tests are larger than the nominal level and a bootstrap calibration is proposed to improve the size. A simulation study confirms that the sizes of the new methods are reasonably accurate for sample size larger than 100 and powerful too.

3.5 Proofs

Lemma 1. Under conditions of Theorem 4, we have as $n \to \infty$

$$\begin{aligned}
& \tfrac{1}{\sqrt{k}} \textstyle\sum_{i=1}^{k} G_1(i) \\
= \quad & \tfrac{1}{\sqrt{k}} \textstyle\sum_{i=1}^{k} \{ \frac{F^2(\epsilon_i(\alpha_0);\beta_0)+F^2(\tilde{\epsilon}_i(\alpha_0);\beta_0)}{2} - F(\epsilon_i(\alpha_0) \vee \tilde{\epsilon}_i(\alpha_0);\beta_0) + \tfrac{1}{3} \} + o_p(1) \\
=: \quad & W_{k1} + o_p(1)
\end{aligned}$$

and

$$\begin{aligned}
& \tfrac{1}{\sqrt{k}} \textstyle\sum_{i=1}^{k} G_2(i) \\
= \quad & \tfrac{1}{\sqrt{k}} \textstyle\sum_{i=1}^{k} \{ F(\epsilon_i(\alpha_0);\beta_0) + F(\tilde{\epsilon}_i(\alpha_0);\beta_0) - 2F^3(\epsilon_i(\alpha_0);\beta_0) - 2F^3(\tilde{\epsilon}_i(\alpha_0);\beta_0) \} \\
& + E\{ 2F'(\epsilon_1(\alpha_0);\beta_0) - 12F^2(\epsilon_1(\alpha_0);\beta_0)F'(\epsilon_1(\alpha_0);\beta_0) \} \times \\
& E\{ \tfrac{\partial}{\partial \alpha^T} m(X_1;\alpha_0) \} \{ E \tfrac{\partial}{\partial \alpha^T} h_1(\alpha_0) \}^{-1} \tfrac{1}{\sqrt{k}} \textstyle\sum_{i=1}^{k} h_i(\alpha_0) + o_p(1) \\
=: \quad & W_{k2} + o_p(1).
\end{aligned}$$

Proof. For simplicity we write $F(x)$, $m(x)$, ϵ_i and $\tilde{\epsilon}_i$ instead of $F(x;\beta_0)$, $m(x;\alpha_0)$,

$\epsilon_i(\alpha_0)$ and $\tilde{\epsilon}_i(\alpha_0)$, respectively. So

$$
\begin{aligned}
& F(\epsilon_j^*(\hat{\alpha})) - F(\epsilon_j^*(\hat{\alpha}_i)) \\
= \; & F(\epsilon_j(\hat{\alpha})) - F(\epsilon_j(\hat{\alpha}_i)) \\
& + \{F(\tilde{\epsilon}_j(\hat{\alpha})) - F(\epsilon_j(\hat{\alpha})) - F(\tilde{\epsilon}_j(\hat{\alpha}_i)) + F(\epsilon_j(\hat{\alpha}_i))\} I(\epsilon_j(\hat{\alpha}) \le \tilde{\epsilon}_j(\hat{\alpha})) \\
& + \{F(\tilde{\epsilon}_j(\hat{\alpha}_i)) - F(\epsilon_j(\hat{\alpha}_i))\}\{I(\epsilon_j(\hat{\alpha}) \le \tilde{\epsilon}_j(\hat{\alpha})) - I(\epsilon_j(\hat{\alpha}_i) \le \tilde{\epsilon}_j(\hat{\alpha}_i))\} \\
= \; & I_1(j,i) + I_2(j,i) + I_3(j,i).
\end{aligned}
$$

Since $\max_{1 \le i \le k} |\hat{\alpha} - \hat{\alpha}_i| = O_p(k^{-\delta})$ for some $\delta > 1/2$, by conditions A1)– A3), there are some $\delta' \in (1/2, \delta)$ and some $M > 0$ such that

$$
\begin{aligned}
& \tfrac{1}{\sqrt{k}} \textstyle\sum_{i=1}^k \sum_{j \ne i} I_3(j,i) \\
= \; & O_p\Big(|\tfrac{1}{\sqrt{k}} \textstyle\sum_{i=1}^k \sum_{j \ne i} \{F(\tilde{\epsilon}_j(\hat{\alpha}_i)) - F(\epsilon_j(\hat{\alpha}_i))\} \times \\
& I(|\epsilon_j(\hat{\alpha}_i) - \tilde{\epsilon}_j(\hat{\alpha}_i)| \le M||(\tfrac{1}{k} \textstyle\sum_{l=1}^k \tfrac{\partial}{\partial \alpha^T} h_l(\hat{\alpha}))^{-1} h_i(\hat{\alpha})||k^{-1})| \Big) \\
= \; & O_p\Big(\tfrac{1}{\sqrt{k}} \textstyle\sum_{i=1}^k \sum_{j \ne i} M^2 ||(\tfrac{1}{k} \sum_{l=1}^k \tfrac{\partial}{\partial \alpha^T} h_l(\hat{\alpha}))^{-1} h_i(\hat{\alpha})||k^{-1} \times \\
& I(|\epsilon_j(\hat{\alpha}) - \tilde{\epsilon}_j(\hat{\alpha})| \le k^{-\delta'}) \Big) \\
= \; & o_p(1).
\end{aligned} \tag{42}
$$

Since $\sum_{i=1}^k h_i(\hat{\alpha}) = 0$, it follows from Taylor expansions that

$$
\begin{aligned}
& \tfrac{1}{\sqrt{k}} \textstyle\sum_{i=1}^k \sum_{j \ne i} I_1(j,i) \\
= \; & \tfrac{1}{\sqrt{k}} \textstyle\sum_{i=1}^k \sum_{j \ne i} F'(\epsilon_j(\hat{\alpha})) \{\tfrac{\partial}{\partial \alpha^T} m(X_j; \hat{\alpha})\} \{\tfrac{1}{k} \sum_{l=1}^k \tfrac{\partial}{\partial \alpha^T} h_l(\hat{\alpha})\}^{-1} \tfrac{h_i(\hat{\alpha})}{k} + o_p(1) \\
= \; & -\tfrac{1}{\sqrt{k}} \textstyle\sum_{j=1}^k F'(\epsilon_j(\hat{\alpha})) \{\tfrac{\partial}{\partial \alpha^T} m(X_j; \hat{\alpha})\} \{\tfrac{1}{k} \sum_{l=1}^k \tfrac{\partial}{\partial \alpha^T} h_l(\hat{\alpha})\}^{-1} \tfrac{h_j(\hat{\alpha})}{k} + o_p(1) \\
= \; & o_p(1).
\end{aligned}
$$

Similarly we have

$$\frac{1}{\sqrt{k}} \sum_{i=1}^k \sum_{j \ne i} I_2(j,i) = o_p(1).$$

Therefore,

$$\frac{1}{\sqrt{k}} \sum_{i=1}^k \sum_{j \ne i} \{F(\epsilon_j^*(\hat{\alpha})) - F(\epsilon_j^*(\hat{\alpha}_i))\} = o_p(1). \tag{43}$$

Using $\sum_{i=1}^k h_i(\hat{\alpha}) = 0$ again, we have

$$
\begin{aligned}
& \tfrac{1}{\sqrt{k}} \textstyle\sum_{i=1}^k \sum_{j\neq i}\{F^2(\epsilon_j(\hat{\alpha})) + F^2(\tilde{\epsilon}_j(\hat{\alpha})) - F^2(\epsilon_j(\hat{\alpha}_i)) - F^2(\tilde{\epsilon}_j(\hat{\alpha}_i))\} \\
= \; & \tfrac{1}{\sqrt{k}} \textstyle\sum_{i=1}^k \sum_{j\neq i} 2F(\epsilon_j(\hat{\alpha}))F'(\epsilon_j(\hat{\alpha}))\{\tfrac{\partial}{\partial\alpha^T} m(X_j;\hat{\alpha})\}\{\tfrac{1}{k}\sum_{l=1}^k \tfrac{\partial}{\partial\alpha^T} h_l(\hat{\alpha})\}^{-1}\tfrac{h_i(\hat{\alpha})}{k} \\
& + \tfrac{1}{\sqrt{k}} \textstyle\sum_{i=1}^k \sum_{j\neq i} 2F(\tilde{\epsilon}_j(\hat{\alpha}))F'(\tilde{\epsilon}_j(\hat{\alpha}))\{\tfrac{\partial}{\partial\alpha^T} m(X_{k+j};\hat{\alpha})\}\{\tfrac{1}{k}\sum_{l=1}^k \tfrac{\partial}{\partial\alpha^T} h_l(\hat{\alpha})\}^{-1}\tfrac{h_i(\hat{\alpha})}{k} \\
& + o_p(1) \\
= \; & -\tfrac{1}{\sqrt{k}} \textstyle\sum_{j=1}^k 2F(\epsilon_j(\hat{\alpha}))F'(\epsilon_j(\hat{\alpha}))\{\tfrac{\partial}{\partial\alpha^T} m(X_j;\hat{\alpha})\}\{\tfrac{1}{k}\sum_{l=1}^k \tfrac{\partial}{\partial\alpha^T} h_l(\hat{\alpha})\}^{-1}\tfrac{h_j(\hat{\alpha})}{k} \\
& + \tfrac{1}{\sqrt{k}} \textstyle\sum_{j=1}^k 2F(\tilde{\epsilon}_j(\hat{\alpha}))F'(\tilde{\epsilon}_j(\hat{\alpha}))\{\tfrac{\partial}{\partial\alpha^T} m(X_{k+j};\hat{\alpha})\}\{\tfrac{1}{k}\sum_{l=1}^k \tfrac{\partial}{\partial\alpha^T} h_l(\hat{\alpha})\}^{-1}\tfrac{h_j(\hat{\alpha})}{k} \\
& + o_p(1) \\
= \; & o_p(1).
\end{aligned} \tag{44}
$$

Thus, it follows from (43) and (44) that

$$
\frac{1}{\sqrt{k}}\sum_{i=1}^k G_1(i) = \frac{1}{\sqrt{k}}\sum_{i=1}^k \{\frac{F^2(\epsilon_i(\hat{\alpha})) + F^2(\tilde{\epsilon}_i(\hat{\alpha}))}{2} - F(\epsilon_i^*(\hat{\alpha})) + 1/3\} + o_p(1). \tag{45}
$$

Similar to (44), we can show that

$$
\begin{aligned}
& \tfrac{1}{\sqrt{k}} \textstyle\sum_{i=1}^k G_2(i) \\
= \; & \tfrac{1}{\sqrt{k}} \textstyle\sum_{i=1}^k \{F(\epsilon_i(\hat{\alpha})) + F(\tilde{\epsilon}_i(\hat{\alpha})) - 2F^3(\epsilon_i(\hat{\alpha})) - 2F^3(\tilde{\epsilon}_i(\hat{\alpha}))\} \\
& + \tfrac{1}{\sqrt{k}} \textstyle\sum_{i=1}^k \sum_{j\neq i}\{F(\epsilon_j(\hat{\alpha})) + F(\tilde{\epsilon}_j(\hat{\alpha})) - F(\epsilon_j(\hat{\alpha}_i)) - F(\tilde{\epsilon}_j(\hat{\alpha}_i))\} \\
& - \tfrac{1}{\sqrt{k}} \textstyle\sum_{i=1}^k \sum_{j\neq i}\{F^3(\epsilon_j(\hat{\alpha})) + F^3(\tilde{\epsilon}_j(\hat{\alpha})) - F^3(\epsilon_j(\hat{\alpha}_i)) - F^3(\tilde{\epsilon}_j(\hat{\alpha}_i))\} \\
= \; & \tfrac{1}{\sqrt{k}} \textstyle\sum_{i=1}^k \{F(\epsilon_i(\hat{\alpha})) + F(\tilde{\epsilon}_i(\hat{\alpha})) - 2F^3(\epsilon_i(\hat{\alpha})) - 2F^3(\tilde{\epsilon}_i(\hat{\alpha}))\} + o_p(1).
\end{aligned} \tag{46}
$$

It is easy to show that

$$
\sqrt{k}\{\hat{\alpha} - \alpha_0\} = -\{E\frac{\partial}{\partial\alpha^T}h_1(\alpha_0)\}^{-1}\frac{1}{\sqrt{k}}\sum_{i=1}^k h_i(\alpha_0) + o_p(1). \tag{47}
$$

Like the proof of (42), we have

$$\begin{aligned}
& \tfrac{1}{\sqrt{k}}\textstyle\sum_{i=1}^{k}\{F(\epsilon_i^*(\hat{\alpha}))-2/3\} \\
= \;& \tfrac{1}{\sqrt{k}}\textstyle\sum_{i=1}^{k}F(\epsilon_i(\hat{\alpha}))-\tfrac{2}{3}\sqrt{k} \\
& +\tfrac{1}{\sqrt{k}}\textstyle\sum_{i=1}^{k}\{F(\tilde{\epsilon}_i(\hat{\alpha}))-F(\epsilon_i(\hat{\alpha}))\}I(\epsilon_i(\hat{\alpha})<\tilde{\epsilon}_i(\hat{\alpha})) \\
= \;& \tfrac{1}{\sqrt{k}}\textstyle\sum_{i=1}^{k}F(\epsilon_i(\hat{\alpha}))-\tfrac{2}{3}\sqrt{k} \\
& +\tfrac{1}{\sqrt{k}}\textstyle\sum_{i=1}^{k}\{F(\tilde{\epsilon}_i(\hat{\alpha}))-F(\epsilon_i(\hat{\alpha}))\}I(\epsilon_i<\tilde{\epsilon}_i)+o_p(1) \\
= \;& \tfrac{1}{\sqrt{k}}\textstyle\sum_{i=1}^{k}\{F(\epsilon_i)-F'(\epsilon_i)\tfrac{\partial}{\partial\alpha^T}m(X_i;\alpha_0)(\hat{\alpha}-\alpha_0)\}-\tfrac{2}{3}\sqrt{k} \\
& +\tfrac{1}{\sqrt{k}}\textstyle\sum_{i=1}^{k}\{F(\tilde{\epsilon}_i)-F'(\tilde{\epsilon})\tfrac{\partial}{\partial\alpha^T}m(X_{k+i};\alpha_0)(\hat{\alpha}-\alpha_0)-F(\epsilon_i) \qquad (48) \\
& +F'(\epsilon_i)\tfrac{\partial}{\partial\alpha^T}m(X_i;\alpha_0)(\hat{\alpha}-\alpha_0)\}I(\epsilon_i<\tilde{\epsilon}_i)+o_p(1) \\
= \;& \tfrac{1}{\sqrt{k}}\textstyle\sum_{i=1}^{k}\{F(\epsilon_i\vee\tilde{\epsilon}_i)-2/3\} \\
& -E\{F'(\epsilon_1)\tfrac{\partial}{\partial\alpha^T}m(X_1;\alpha_0)I(\epsilon_1>\tilde{\epsilon}_1) \\
& +F'(\tilde{\epsilon}_1)\tfrac{\partial}{\partial\alpha^T}m(X_{k+1};\alpha_0)I(\epsilon_1\le\tilde{\epsilon}_1)\}\sqrt{k}(\hat{\alpha}-\alpha_0)+o_p(1) \\
= \;& \tfrac{1}{\sqrt{k}}\textstyle\sum_{i=1}^{k}\{F(\epsilon_i\vee\tilde{\epsilon}_i)-2/3\} \\
& -2E\{F(\epsilon_1)F'(\epsilon_1)\}E\{\tfrac{\partial}{\partial\alpha^T}m(X_1;\alpha_0)\}\sqrt{k}(\hat{\alpha}-\alpha_0)+o_p(1).
\end{aligned}$$

It is easy to verify that

$$\begin{aligned}
& \tfrac{1}{\sqrt{k}}\textstyle\sum_{i=1}^{k}\{F(\epsilon_i(\hat{\alpha}))-1/2\} \\
= \;& \tfrac{1}{\sqrt{k}}\textstyle\sum_{i=1}^{k}\{F(\epsilon_i)-1/2\} \qquad (49) \\
& -E\{F'(\epsilon_1)\}E\{\tfrac{\partial}{\partial\alpha^T}m(X_1;\alpha_0)\}\sqrt{k}(\hat{\alpha}-\alpha_0)+o_p(1),
\end{aligned}$$

$$\begin{aligned}
& \tfrac{1}{\sqrt{k}}\textstyle\sum_{i=1}^{k}\{F(\tilde{\epsilon}_i(\hat{\alpha}))-1/2\} \\
= \;& \tfrac{1}{\sqrt{k}}\textstyle\sum_{i=1}^{k}\{F(\tilde{\epsilon}_i)-1/2\} \qquad (50) \\
& -E\{F'(\epsilon_1)\}E\{\tfrac{\partial}{\partial\alpha^T}m(X_1;\alpha_0)\}\sqrt{k}(\hat{\alpha}-\alpha_0)+o_p(1),
\end{aligned}$$

$$\begin{aligned}
&\tfrac{1}{\sqrt{k}}\textstyle\sum_{i=1}^{k}\{F^2(\epsilon_i(\hat\alpha))-1/3\} \\
= \;&\tfrac{1}{\sqrt{k}}\textstyle\sum_{i=1}^{k}\{F^2(\epsilon_i)-1/3\} \qquad (51)\\
&-2E\{F(\epsilon_1)F'(\epsilon_1)\}E\{\tfrac{\partial}{\partial\alpha^T}m(X_1;\alpha_0)\}\sqrt{k}(\hat\alpha-\alpha_0)+o_p(1),
\end{aligned}$$

$$\begin{aligned}
&\tfrac{1}{\sqrt{k}}\textstyle\sum_{i=1}^{k}\{F^2(\tilde\epsilon_i(\hat\alpha))-1/3\} \\
= \;&\tfrac{1}{\sqrt{k}}\textstyle\sum_{i=1}^{k}\{F^2(\tilde\epsilon_i)-1/3\} \qquad (52)\\
&-2E\{F(\epsilon_1)F'(\epsilon_1)\}E\{\tfrac{\partial}{\partial\alpha^T}m(X_1;\alpha_0)\}\sqrt{k}(\hat\alpha-\alpha_0)+o_p(1),
\end{aligned}$$

$$\begin{aligned}
&\tfrac{1}{\sqrt{k}}\textstyle\sum_{i=1}^{k}\{F^3(\epsilon_i(\hat\alpha))-1/4\} \\
= \;&\tfrac{1}{\sqrt{k}}\textstyle\sum_{i=1}^{k}\{F^3(\epsilon_i)-1/4\} \qquad (53)\\
&-3E\{F^2(\epsilon_1)F'(\epsilon_1)\}E\{\tfrac{\partial}{\partial\alpha^T}m(X_1;\alpha_0)\}\sqrt{k}(\hat\alpha-\alpha_0)+o_p(1)
\end{aligned}$$

and

$$\begin{aligned}
&\tfrac{1}{\sqrt{k}}\textstyle\sum_{i=1}^{k}\{F^3(\tilde\epsilon_i(\hat\alpha))-1/4\} \\
= \;&\tfrac{1}{\sqrt{k}}\textstyle\sum_{i=1}^{k}\{F^3(\tilde\epsilon_i)-1/4\} \qquad (54)\\
&-3E\{F^2(\epsilon_1)F'(\epsilon_1)\}E\{\tfrac{\partial}{\partial\alpha^T}m(X_1;\alpha_0)\}\sqrt{k}(\hat\alpha-\alpha_0)+o_p(1).
\end{aligned}$$

Hence the lemma follows from (45)–(54).

Lemma 2. Under conditions of Theorem 4, we have

$$\frac{1}{k}\sum_{i=1}^{k}G_j(i)G_l(i)\xrightarrow{p}\lim_{n\to\infty}E(W_{kj}W_{kl})$$

for $j,l=1,2$ as $n\to\infty$.

Proof. Put

$$A_{i1}=\sum_{j\neq i}\frac{F^2(\epsilon_j(\hat\alpha))+F^2(\tilde\epsilon_j(\hat\alpha))-F^2(\epsilon_j(\hat\alpha_i))-F^2(\tilde\epsilon_j(\hat\alpha_i))}{2}$$

and

$$A_{i2}=\sum_{j\neq i}\{F(\epsilon_j^*(\hat\alpha))-F(\epsilon_j^*(\hat\alpha_i))\}.$$

Like the proof of (42), we have

$$
\begin{aligned}
A_{i2} &= \sum_{j\neq i}\{F(\epsilon_j(\hat\alpha)) - F(\epsilon_j(\hat\alpha_i))\}I(\epsilon_j(\hat\alpha) > \tilde\epsilon_j(\hat\alpha)) \\
&\quad + \sum_{j\neq i}\{F(\tilde\epsilon_j(\hat\alpha)) - F(\tilde\epsilon_j(\hat\alpha_i))\}I(\epsilon_j(\hat\alpha) \le \tilde\epsilon_j(\hat\alpha)) + o_p(1) \\
&= \sum_{j\neq i} F'(\epsilon_j)\{\tfrac{\partial}{\partial\alpha^T}m(X_j)\}\{\tfrac{1}{k}\sum_{l=1}^k \tfrac{\partial}{\partial\alpha^T}h_l(\alpha_0)\}^{-1}h_i(\alpha_0)k^{-1}I(\epsilon_j > \tilde\epsilon_j) \\
&\quad + \sum_{j\neq i} F'(\tilde\epsilon_j)\{\tfrac{\partial}{\partial\alpha^T}m(X_{k+j})\}\{\tfrac{1}{k}\sum_{l=1}^k \tfrac{\partial}{\partial\alpha^T}h_l(\alpha_0)\}^{-1}h_i(\alpha_0)k^{-1}I(\epsilon_j \le \tilde\epsilon_j) \\
&\quad + o_p(1) \\
&= 2E\{F(\epsilon_1)F'(\epsilon_1)\}E\{\tfrac{\partial}{\partial\alpha^T}m(X_1)\}\{E\tfrac{\partial}{\partial\alpha^T}h_1(\alpha_0)\}^{-1}h_i(\alpha_0) + o_p(1).
\end{aligned}
\tag{55}
$$

It is easy to check that

$$
A_{i1} = 2E\{F(\epsilon_1)F'(\epsilon_1)\}E\{\frac{\partial}{\partial\alpha^T}m(X_1)\}\{E\frac{\partial}{\partial\alpha^T}h_1(\alpha_0)\}^{-1}h_i(\alpha_0) + o_p(1). \tag{56}
$$

Thus, it follows from (55) and (56) that

$$
\begin{aligned}
& \tfrac{1}{k}\sum_{i=1}^k G_1^2(i) \\
= \; & \tfrac{1}{k}\sum_{i=1}^k (A_{i1} - A_{i2})^2 + \tfrac{1}{k}\sum_{i=1}^k (\tfrac{F^2(\epsilon_i(\hat\alpha)) + F^2(\tilde\epsilon_i(\hat\alpha))}{2} - F(\epsilon_i^*(\hat\alpha)) + \tfrac{1}{3})^2 \\
& + \tfrac{2}{k}\sum_{i=1}^k (A_{i1} - A_{i2})(\tfrac{F^2(\epsilon_i(\hat\alpha)) + F^2(\tilde\epsilon_i(\hat\alpha))}{2} - F(\epsilon_i^*(\hat\alpha)) + \tfrac{1}{3}) \\
= \; & \tfrac{1}{k}\sum_{i=1}^k (\tfrac{F^2(\epsilon_i) + F^2(\tilde\epsilon_i)}{2} - F(\epsilon_i \vee \tilde\epsilon_i) + \tfrac{1}{3})^2 + o_p(1) \\
= \; & \lim_{n\to\infty} W_{k1}^2 + o_p(1).
\end{aligned}
$$

The rest can be shown in a similar way.

Proof of Theorem 4. It follows from Lemmas 1, 2 and some standard arguments in the empirical likelihood method (see Chapter 11 of Owen (2001)).

Proof of Theorem 5. This can be shown in a similar way to the proof of Theorem 4 although some more tedious expansions are needed.

3.6 Tables

Table 11: Powers of the proposed jackknife empirical likelihood test(JEL) and Cramér-von-Mises test (CM) are reported for the case of $n = 200$ and $\nu = 3$. Define $\mathcal{F}^n = \{N(0, \sigma^2) : \sigma > 0\}$ and $\mathcal{F}^t = \{t(\nu) : \nu > 2\}$.

δ	H_0	JEL Level 5%	CM Level 5%	JEL Level 10%	CM Level 10%
0	N(0,1)	0.0500	0.0479	0.0999	0.0944
	t(3)	0.0592	0.0541	0.1128	0.1031
	$\mathcal{F}^n$	0.0597	0.0512	0.1115	0.1007
	$\mathcal{F}^t$	0.0584	0.0451	0.1095	0.0974
0.5	N(0,1)	0.0760	0.0667	0.1405	0.1297
	t(3)	0.0814	0.0497	0.1423	0.1006
	$\mathcal{F}^n$	0.0651	0.0490	0.1170	0.0996
	$\mathcal{F}^t$	0.0663	0.0504	0.1210	0.1063
1	N(0,1)	0.1542	0.1384	0.2426	0.2193
	t(3)	0.1247	0.0531	0.1989	0.1164
	$\mathcal{F}^n$	0.0620	0.0541	0.1157	0.1039
	$\mathcal{F}^t$	0.1007	0.0547	0.1702	0.1227
2	N(0,1)	0.4158	0.3793	0.5404	0.5129
	t(3)	0.3238	0.1067	0.4340	0.2424
	$\mathcal{F}^n$	0.0609	0.0604	0.1104	0.1096
	$\mathcal{F}^t$	0.2496	0.1289	0.3559	0.2601
3	N(0,1)	0.6677	0.6397	0.7709	0.7466
	t(3)	0.6092	0.2737	0.7181	0.4959
	$\mathcal{F}^n$	0.0819	0.0877	0.1464	0.1498
	$\mathcal{F}^t$	0.4995	0.3133	0.6227	0.5199
$\sqrt{n}$	$\mathcal{F}^n$	0.9580	0.9649	0.9752	0.9817

Table 12: Powers of the proposed jackknife empirical likelihood test(JEL) and Cramér-von-Mises test (CM) are reported for the case of $n = 200$ and $\nu = 8$. Define $\mathcal{F}^n = \{N(0, \sigma^2) : \sigma > 0\}$ and $\mathcal{F}^t = \{t(\nu) : \nu > 2\}$.

δ	H_0	JEL Level 5%	CM Level 5%	JEL Level 10%	CM Level 10%
0	N(0,1)	0.0527	0.0478	0.1044	0.0969
	t(8)	0.0550	0.0538	0.1062	0.1040
	$\mathcal{F}^n$	0.0632	0.0551	0.1185	0.1011
	$\mathcal{F}^t$	0.0548	0.0494	0.1043	0.0934
0.5	N(0,1)	0.0819	0.0724	0.1445	0.1357
	t(8)	0.0784	0.0671	0.1363	0.1276
	$\mathcal{F}^n$	0.0628	0.0505	0.1152	0.0998
	$\mathcal{F}^t$	0.0717	0.0632	0.1335	0.1198
1	N(0,1)	0.1593	0.1394	0.2555	0.2292
	t(8)	0.1289	0.1216	0.2391	0.2100
	$\mathcal{F}^n$	0.0593	0.0499	0.1169	0.1028
	$\mathcal{F}^t$	0.1299	0.1204	0.2158	0.2016
2	N(0,1)	0.5104	0.4574	0.6399	0.5908
	t(8)	0.4636	0.3894	0.5907	0.5283
	$\mathcal{F}^n$	0.0624	0.0508	0.1169	0.1014
	$\mathcal{F}^t$	0.4043	0.3809	0.5350	0.5158
3	N(0,1)	0.8368	0.7973	0.9040	0.8786
	t(8)	0.8042	0.7508	0.8810	0.8469
	$\mathcal{F}^n$	0.0593	0.0527	0.1140	0.1033
	$\mathcal{F}^t$	0.7598	0.7399	0.8480	0.8369
$\sqrt{n}$	$\mathcal{F}^n$	0.2728	0.2528	0.3721	0.3655

Table 13: Powers of the proposed jackknife empirical likelihood test(JEL) and Cramér-von-Mises test (CM) are reported for the case of $n = 500$ and $\nu = 3$. Define $\mathcal{F}^n = \{N(0, \sigma^2) : \sigma > 0\}$ and $\mathcal{F}^t = \{t(\nu) : \nu > 2\}$.

δ	H_0	JEL Level 5%	CM Level 5%	JEL Level 10%	CM Level 10%
0	N(0,1)	0.0505	0.0544	0.1061	0.1018
	t(3)	0.0532	0.0517	0.1040	0.0994
	$\mathcal{F}^n$	0.0545	0.0518	0.1023	0.1036
	$\mathcal{F}^t$	0.0561	0.0483	0.1054	0.0991
0.5	N(0,1)	0.0751	0.0692	0.1370	0.1296
	t(3)	0.0675	0.0487	0.1458	0.1018
	$\mathcal{F}^n$	0.0528	0.0505	0.1027	0.1009
	$\mathcal{F}^t$	0.0615	0.0510	0.1183	0.1016
1	N(0,1)	0.1543	0.1400	0.2491	0.2284
	t(3)	0.1252	0.0589	0.1913	0.1234
	$\mathcal{F}^n$	0.0527	0.0490	0.1021	0.0967
	$\mathcal{F}^t$	0.0921	0.0650	0.1636	0.1304
2	N(0,1)	0.4609	0.4191	0.5903	0.5483
	t(3)	0.2982	0.1173	0.4189	0.2567
	$\mathcal{F}^n$	0.0496	0.0556	0.1005	0.1061
	$\mathcal{F}^t$	0.2378	0.1249	0.3450	0.2525
3	N(0,1)	0.7846	0.7387	0.8616	0.8361
	t(3)	0.5919	0.3000	0.7065	0.5217
	$\mathcal{F}^n$	0.0570	0.0592	0.1117	0.1095
	$\mathcal{F}^t$	0.4695	0.3151	0.5948	0.5180
$\sqrt{n}$	$\mathcal{F}^n$	0.9999	1	1	1

Table 14: Powers of the proposed jackknife empirical likelihood test(JEL) and Cramér-von-Mises test (CM) are reported for the case of $n = 500$ and $\nu = 8$. Define $\mathcal{F}^n = \{N(0, \sigma^2) : \sigma > 0\}$ and $\mathcal{F}^t = \{t(\nu) : \nu > 2\}$.

δ	H_0	JEL Level 5%	CM Level 5%	JEL Level 10%	CM Level 10%
0	N(0,1)	0.0507	0.0503	0.1015	0.0999
	t(8)	0.0541	0.0485	0.1040	0.1010
	$\mathcal{F}^n$	0.0551	0.0527	0.1088	0.1032
	$\mathcal{F}^t$	0.0511	0.0503	0.0984	0.0982
0.5	N(0,1)	0.0776	0.0708	0.1441	0.1355
	t(8)	0.0779	0.0686	0.1347	0.1303
	$\mathcal{F}^n$	0.0527	0.0510	0.1045	0.0985
	$\mathcal{F}^t$	0.0691	0.0642	0.1282	0.1250
1	N(0,1)	0.1584	0.1433	0.2484	0.2340
	t(8)	0.1474	0.1232	0.2381	0.2078
	$\mathcal{F}^n$	0.0553	0.0536	0.1048	0.1024
	$\mathcal{F}^t$	0.1266	0.1275	0.2134	0.2086
2	N(0,1)	0.5191	0.4777	0.6462	0.6103
	t(8)	0.4508	0.4014	0.5821	0.5322
	$\mathcal{F}^n$	0.0573	0.0552	0.1068	0.1026
	$\mathcal{F}^t$	0.3926	0.3897	0.5230	0.5232
3	N(0,1)	0.8720	0.8316	0.9279	0.9041
	t(8)	0.8209	0.7696	0.8949	0.8615
	$\mathcal{F}^n$	0.0544	0.0514	0.1010	0.0995
	$\mathcal{F}^t$	0.7559	0.7572	0.8411	0.8533
$\sqrt{n}$	$\mathcal{F}^n$	0.5899	0.5452	0.6937	0.6699

Table 15: Powers of the proposed jackknife empirical likelihood test(JEL), its bootstrap calibrated version (BCJEL) and Cramér-von-Mises test (CM) are reported for the case of $n = 50$ and $\nu = 8$. Define $\mathcal{F}^n = \{N(0, \sigma^2) : \sigma > 0\}$ and $\mathcal{F}^t = \{t(\nu) : \nu > 2\}$.

δ	H_0	JEL Level 5%	BCJEL Level 5%	CM Level 5%	JEL Level 10%	BCJEL Level 10%	CM Level 10%
0	N(0,1)	0.0836	0.0442	0.0510	0.1437	0.0901	0.1000
	t(8)	0.0877	0.0423	0.0506	0.1450	0.0893	0.0981
	$\mathcal{F}^n$	0.1133	0.0460	0.0539	0.1743	0.0934	0.1039
	$\mathcal{F}^t$	0.0816	0.0430	0.0376	0.1380	0.0895	0.0802
0.5	N(0,1)	0.1327	0.0786	0.0671	0.1999	0.1368	0.1319
	t(8)	0.1273	0.0733	0.0644	0.1899	0.1263	0.1218
	$\mathcal{F}^n$	0.1118	0.0458	0.0476	0.1730	0.0914	0.0916
	$\mathcal{F}^t$	0.1036	0.0586	0.0580	0.1666	0.1072	0.1104
1	N(0,1)	0.2289	0.1430	0.1335	0.3153	0.2308	0.2174
	t(8)	0.2118	0.1386	0.1171	0.2952	0.2136	0.1977
	$\mathcal{F}^n$	0.1170	0.0468	0.0509	0.1782	0.0930	0.0995
	$\mathcal{F}^t$	0.1619	0.0975	0.1031	0.2294	0.1620	0.1775
2	N(0,1)	0.5123	0.3665	0.3669	0.6150	0.5038	0.4957
	t(8)	0.5128	0.3685	0.3329	0.6146	0.5027	0.4691
	$\mathcal{F}^n$	0.1107	0.0515	0.0425	0.1729	0.0879	0.1044
	$\mathcal{F}^t$	0.4047	0.2723	0.3058	0.5123	0.3929	0.4302
3	N(0,1)	0.6505	0.4868	0.5327	0.7474	0.6349	0.6659
	t(8)	0.7566	0.5973	0.5724	0.8348	0.7374	0.7105
	$\mathcal{F}^n$	0.1101	0.0399	0.0530	0.1772	0.0838	0.1047
	$\mathcal{F}^t$	0.6576	0.4966	0.5395	0.7558	0.6370	0.6576

Table 16: Powers of the proposed jackknife empirical likelihood test(JEL), its bootstrap calibrated version (BCJEL) and Cramér-von-Mises test (CM) are reported for the case of $n = 100$ and $\nu = 8$. Define $\mathcal{F}^n = \{N(0, \sigma^2) : \sigma > 0\}$ and $\mathcal{F}^t = \{t(\nu) : \nu > 2\}$.

δ	H_0	JEL Level 5%	BCJEL Level 5%	CM Level 5%	JEL Level 10%	BCJEL Level 10%	CM Level 10%
0	N(0,1)	0.0615	0.0446	0.0518	0.1153	0.0905	0.1024
	t(8)	0.0618	0.0434	0.0480	0.1108	0.0861	0.1006
	$\mathcal{F}^n$	0.0709	0.0393	0.0487	0.1302	0.0819	0.0990
	$\mathcal{F}^t$	0.0620	0.0431	0.0393	0.1159	0.0934	0.0820
0.5	N(0,1)	0.0934	0.0694	0.0716	0.1586	0.1279	0.1331
	t(8)	0.0931	0.0702	0.0611	0.1562	0.1308	0.1178
	$\mathcal{F}^n$	0.0739	0.0408	0.0494	0.1341	0.0865	0.1000
	$\mathcal{F}^t$	0.0696	0.0522	0.0542	0.1251	0.0999	0.1054
1	N(0,1)	0.1748	0.1424	0.1426	0.2634	0.2246	0.2361
	t(8)	0.1649	0.1335	0.1224	0.2542	0.2142	0.2041
	$\mathcal{F}^n$	0.0738	0.0428	0.0533	0.1335	0.0839	0.1032
	$\mathcal{F}^t$	0.1135	0.0868	0.0989	0.1838	0.1517	0.1670
2	N(0,1)	0.4978	0.4370	0.4276	0.6211	0.5703	0.5642
	t(8)	0.4729	0.4068	0.3741	0.5965	0.5442	0.5078
	$\mathcal{F}^n$	0.0717	0.0392	0.0539	0.1317	0.0827	0.1037
	$\mathcal{F}^t$	0.3392	0.2794	0.3086	0.4570	0.3990	0.4359
3	N(0,1)	0.7753	0.7153	0.7165	0.8662	0.8264	0.8221
	t(8)	0.7948	0.7374	0.6990	0.8727	0.8374	0.8161
	$\mathcal{F}^n$	0.0751	0.0383	0.0548	0.1299	0.0810	0.1037
	$\mathcal{F}^t$	0.6630	0.5852	0.6317	0.7731	0.7169	0.7500

CHAPTER IV

INTERVAL ESTIMATION FOR A SIMPLE BILINEAR MODEL

This chapter is base on the submitted paper:

H. Feng, L. Peng and F. Zhu (2012). Interval Estimation for a Simple Bilinear Model.

4.1 Introduction

As a kind of nonlinear time series models, bilinear models have been widely studied in the literature, see Subba Rao (1981), Pham and Tran (1981), Kim et al. (1990), Basrak et al. (1999) and Giordano (2004), among others. It is well known that the general bilinear models are difficult to deal with because of their complex probabilistic structure. Consider the following simple bilinear time series model

$$X_t = b\epsilon_{t-1}X_{t-2} + \epsilon_t, \tag{57}$$

for $t = 1, 2, ..., n$, where $\epsilon_t's$ are independent and identically distributed random variables with zero mean and variance σ^2. As pointed out by Giordano (2000), model (57) is appealing because it looks like a white noise if one considers only the first and second moments and it can be fitted to residuals of some other linear or nonlinear time series models in order to capture, for example, the skewness or kurtosis. In other words, model (57) may be used as a first step tool for building much more complex nonlinear time series models.

Write

$$\begin{cases} X_t^2 = \sigma^2 + b^2\sigma^2 X_{t-2}^2 + 2bX_{t-2}\epsilon_t\epsilon_{t-1} + b^2X_{t-2}^2(\epsilon_{t-1}^2 - \sigma^2) + (\epsilon_t^2 - \sigma^2), \\ X_tX_{t-1} = b\sigma^2 X_{t-2} + b^2X_{t-3}X_{t-2}\epsilon_{t-2}\epsilon_{t-1} + bX_{t-2}(\epsilon_{t-1}^2 - \sigma^2) + X_{t-1}\epsilon_t. \end{cases} \tag{58}$$

By noting that $E(X_t^2|X_s, s \le t-2) = \sigma^2 + b^2\sigma^2X_{t-2}^2$ and $E(X_tX_{t-1}|X_s, s \le t-2) = b\sigma^2X_{t-2}$, Grahn (1995) proposed a conditional least squares (CLS) estimator

$$\hat{b} = \frac{\hat{\beta}_2}{\hat{\sigma}^2} = \frac{\sum_{t=3}^n X_tX_{t-1}X_{t-2}}{\hat{\sigma}^2 \sum_{t=3}^n X_{t-2}^2}$$

for estimating b in model (57), where $(\hat{\sigma}^2, \hat{\beta}_1)^T$ and $\hat{\beta}_2$ are obtained by minimizing

$$\sum_{t=3}^n \{X_t^2 - (\sigma^2 + \beta_1 X_{t-2}^2)\}^2 \quad \text{and} \quad \sum_{t=3}^n \{X_tX_{t-1} - \beta_2X_{t-2}\}^2,$$

respectively. Further, Grahn (1995) showed the strong convergence and derived the normality of $\hat{b}$ under some regularity conditions without an explicit formula for the asymptotic variance. When ϵ_t is normally distributed, Giordano (2000) showed $\hat{b}$ and the following quantity

$$\frac{1}{n\sigma^2\mu_2}\sum_{t=3}^n X_tX_{t-1}X_{t-2}$$

have the same asymptotic variance, which equals

$$V_1^2 = \frac{1}{n^2\sigma^4\mu_2^2}\mathrm{Var}\left(\sum_{t=3}^n X_tX_{t-1}X_{t-2}\right),$$

where $\mu_2 = E(X_t^2) = \sigma^2/(1-b^2\sigma^2)$. In order to construct a confidence interval for b, when ϵ_t is normally distributed, Giordano and Vitale (2003) further showed that

$$V_1^2 = \frac{1}{n^2}\left\{\frac{1}{\mu_2}\left[\frac{\left(45b^8\sigma^8\frac{1+b^2\sigma^2}{1-3b^4\sigma^4} + 54b^6\sigma^4\mu_2 + 15b^4\sigma^4\right)}{1-3b^6\sigma^6} + 9b^2\sigma^2\frac{1+b^2\sigma^2}{1-3b^4\sigma^4} + 2b^2\mu_2 + 1\right]\right.$$
$$\left. + 6b^2(1+2b^2\sigma^2) + \frac{4b^2}{1-b^2\sigma^2}\left(2 + 3b^2\sigma^2\left(1 + \frac{1}{1-3b^4\sigma^4}\right)\right)\right\}$$

and also proposed another estimator for b with a slightly simple asymptotic variance.

Unfortunately, as far as we know, it still remains unknown on how to derive the explicit asymptotic variance of the conditional least squares estimator without assuming normality for errors. Although naive bootstrap method is the simply way to construct a confidence interval for b without estimating the asymptotic variance, it is known that naive bootstrap method generally gives a poor coverage probability. Moreover, a bootstrap method for time series models requires sampling from the estimated errors and refitting the models, which is computationally intensive. Here we investigate the possibility of using the empirical likelihood method. Recently the empirical likelihood method has been extended to many different fields including time series since Owen (1988, 1990) introduced the empirical likelihood method for a mean vector. For an overview of the empirical likelihood method and related studies, see Owen (2001) and Chen and Van Keilegom (2009). Although the empirical likelihood method has been applied to some nonlinear time series (see Chan and Ling (2006) for GARCH models) and non-stationary time series (see Chuang and Chan (2002) for unit root processes), as far as we know, this chapter is the first time to explore the possible application of empirical likelihood method to bilinear time series models.

We organize this chapter as follows. Section 2 presents the empirical likelihood methods. A simulation study is given in Section 3. Some conclusions are summarized in Section 4. All proofs are provided in Section 5.

4.2 Methodology and main results

Motivated by the way of constructing the conditional least squares estimator $\hat{b}$, one may apply the profile empirical likelihood method in Qin and Lawless (1994) to the following estimating equations

$$\begin{cases} \sum_{t=3}^n \frac{\partial}{\partial b}\{(X_t^2-\sigma^2-b^2\sigma^2X_{t-2}^2)^2+(X_tX_{t-1}-b\sigma^2X_{t-2})^2\}=0, \\ \sum_{t=3}^n \frac{\partial}{\partial \sigma^2}\{(X_t^2-\sigma^2-b^2\sigma^2X_{t-2}^2)^2+(X_tX_{t-1}-b\sigma^2X_{t-2})^2\}=0, \end{cases}$$

or

$$\begin{cases} \sum_{t=3}^n (X_t^2-\sigma^2-b^2\sigma^2X_{t-2}^2)=0, \\ \sum_{t=3}^n (X_tX_{t-1}-b\sigma^2X_{t-2})X_{t-2}=0, \end{cases}$$

or

$$\begin{cases} \sum_{t=3}^n (X_t^2-\sigma^2-b^2\sigma^2X_{t-2}^2)=0, \\ \sum_{t=3}^n (X_t^2-\sigma^2-b^2\sigma^2X_{t-2}^2)X_{t-2}^2=0, \\ \sum_{t=3}^n (X_tX_{t-1}-b\sigma^2X_{t-2})X_{t-2}=0. \end{cases}$$

However, a brief simulation study shows that such ideas lead to poor finite sample behavior. The reason may be that the terms $X_t^2 - E(X_t^2|X_s, s \le t-2)$ and $X_tX_{t-1} - E(X_tX_{t-1}|X_s, s \le t-2)$ in (58) depend on X_{t-2}. This motivates us to apply the profile empirical likelihood method to some weighted estimating equations. More specifically, with respect to the above three estimating equations, we define the empirical likelihood functions as

$$L_1(b,\sigma^2) = \sup\left\{\prod_{t=3}^n((n-2)p_t) : p_3 \ge 0, \cdots, p_n \ge 0, \sum_{t=3}^n p_t = 1, \sum_{t=3}^n p_tY_t(b,\sigma^2) = 0\right\},$$

$$\begin{aligned} L_2(b,\sigma^2) = \sup\Bigg\{&\prod_{t=3}^n((n-2)p_t) : p_3 \ge 0, \cdots, p_t \ge 0, \sum_{t=3}^n p_t = 1, \\ &\sum_{t=3}^n p_t\frac{X_t^2-\sigma^2-b^2\sigma^2X_{t-2}^2}{d+X_{t-2}^2} = 0, \sum_{t=3}^n p_t\frac{(X_tX_{t-1}-b\sigma^2X_{t-2})X_{t-2}}{d+X_{t-2}^2} = 0\Bigg\}, \end{aligned}$$

$$\begin{aligned} L_3(b,\sigma^2) = \sup\Bigg\{&\prod_{t=3}^n((n-2)p_t) : p_3 \ge 0, \cdots, p_n \ge 0, \sum_{t=3}^n p_t = 1, \\ &\sum_{t=3}^n p_t\frac{X_t^2-\sigma^2-b^2\sigma^2X_{t-2}^2}{d+X_{t-2}^2} = 0, \sum_{t=3}^n p_t\frac{(X_t^2-\sigma^2-b^2\sigma^2X_{t-2}^2)X_{t-2}^2}{d+X_{t-2}^2} = 0, \\ &\sum_{t=3}^n p_t\frac{(X_tX_{t-1}-b\sigma^2X_{t-2})X_{t-2}}{d+X_{t-2}^2} = 0\Bigg\}, \end{aligned}$$

where $Y_t(b,\sigma^2) = (Y_{t1}(b,\sigma^2), Y_{t2}(b,\sigma^2))^T$,

$$Y_{t1}(b,\sigma^2) = \frac{(X_t^2 - \sigma^2 - b^2\sigma^2 X_{t-2}^2)2bX_{t-2}^2 + (X_tX_{t-1} - b\sigma^2 X_{t-2})X_{t-2}}{(d + X_{t-2}^2)},$$
$$Y_{t2}(b,\sigma^2) = \frac{(X_t^2 - \sigma^2 - b^2\sigma^2 X_{t-2}^2)(1 + b^2 X_{t-2}^2) + (X_tX_{t-1} - b\sigma^2 X_{t-2})bX_{t-2}}{(d + X_{t-2}^2)},$$

and d is a positive constant. In practice, one can simply choose $d = 1$ as we do in the simulation study.

For proving the Wilks's theorem for the above empirical likelihood methods, we focus on the first one since the other two can be shown similarly.

By the Lagrange multiplier techniques, we have

$$l_1(b,\sigma^2) = -2\log L_1(b,\sigma^2) = 2\sum_{t=3}^{n} \log\{1 + \lambda^T Y_t(b,\sigma^2)\},$$

where $\lambda = \lambda(b,\sigma^2)$ satisfies

$$\sum_{t=3}^{n} \frac{Y_t(b,\sigma^2)}{1 + \lambda^T Y_t(b,\sigma^2)} = 0.$$

Since we are only interested in constructing confidence intervals for b, we consider the profile empirical likelihood function $l_1^P(b) = \min_{\sigma^2>0} l_1(b,\sigma^2)$. Throughout b_0 and σ_0^2 denote the true values of b and σ^2, respectively. Following the procedure in Qin and Lawless (1994), we first show the following proposition.

Proposition 4.2.1. *Suppose model* (57) *has a strictly stationary, causal and ergodic solution with* $E|X_t|^{4\delta} < \infty$ *for some* $\delta > 1$. *Then, with probability tending to 1,* $l_1(b_0,\sigma^2)$ *attains its minimum value at some point* $\tilde\sigma^2$ *in the interior of the ball* $U_n = \{\sigma^2 : |\sigma^2 - \sigma_0^2| \le Cn^{-1/(2\gamma)}\}$ *for some given* $\gamma \in (1,\delta)$ *and* $C > 0$. *Moreover* $\tilde\sigma^2$ *and* $\tilde\lambda = \tilde\lambda(b_0,\tilde\sigma^2)$ *satisfy*

$$Q_{1n}(\tilde\sigma^2,\tilde\lambda) = 0, \quad Q_{2n}(\tilde\sigma^2,\tilde\lambda) = 0,$$

where

$$Q_{1n}(\sigma^2, \lambda) = \frac{1}{n-2} \sum_{t=3}^{n} \frac{Y_t(b_0, \sigma^2)}{1 + \lambda^T Y_t(b_0, \sigma^2)},$$

$$Q_{2n}(\sigma^2, \lambda) = \frac{1}{n-2} \sum_{t=3}^{n} \frac{\lambda^T}{1 + \lambda^T Y_t(b_0, \sigma^2)} \frac{\partial Y_t(b_0, \sigma^2)}{\partial \sigma^2}.$$

Theorem 6. *Under conditions of Proposition 4.2.1, $l_1^P(b_0)$ converges in distribution to a chi-square distribution with one degree of freedom as $n \to \infty$.*

Remark 1. *Theorem 3.1 of Grahn (1995) gives conditions to ensure that there exists a strictly stationary, ergodic and causal solution. When $\epsilon_t \sim N(0, \sigma^2)$, then $b^4\sigma^4 < 1/3$ is a sufficient condition for $E(X_t^4) < \infty$, and $b^2\sigma^2 < 1$ is a sufficient condition for the existence of a strictly stationary, ergodic and causal solution.*

Based on the above theorem, we can construct empirical likelihood confidence intervals for b as

$$I_\alpha^{(1)} = \{\theta : l_1^P(\theta) \le \chi^2_{1,\alpha}\},$$

where $\chi^2_{1,\alpha}$ denotes the α quantile of a chi-square distribution with one degree of freedom.

Similarly we can consider the following profile empirical likelihood functions

$$L_2^P(b) = \max_{\sigma^2} L_2(b, \sigma^2), \quad L_3^P(b) = \max_{b,\sigma^2} L_3(b, \sigma^2) - \max_{\sigma^2} L_3(b, \sigma^2),$$

and prove that both $-2\log L_2^P(b_0)$ and $-2\log L_3^P(b_0)$ converge in distribution to a chi-square distribution with one degree of freedom as $n \to \infty$ under the same regularity conditions as in Theorem 6. Hence the corresponding empirical likelihood confidence intervals for b with level α

$$I_\alpha^{(2)} = \{\theta : -2\log L_2^P(\theta) \le \chi^2_{1,\alpha}\} \quad \text{and} \quad I_\alpha^{(3)} = \{\theta : -2\log L_3^P(\theta) \le \chi^2_{1,\alpha}\}.$$

4.3 Simulations

In this section, we investigate the finite sample performance of the proposed empirical likelihood confidence intervals and compare them with the bootstrap method in terms of coverage accuracy.

We draw $10,000$ random samples with sizes $n = 100, 200, 1,000$ from model (57) with $\epsilon_t \sim N(0,1)$. We choose $b = 0.1, 0.2, 0.3, 0.4$ and calculate coverage probabilities for the proposed four intervals $I_\alpha^*, I_\alpha^{(1)}, I_\alpha^{(2)}, I_\alpha^{(3)}$ with levels $\alpha = 0.9$ and 0.95. For calculating the empirical likelihood confidence intervals, we choose $d = 1$ and employ the R package 'emplik' to obtain the empirical likelihood function and use the R package 'nlm' to find the profile empirical likelihood function. For computing the bootstrap confidence interval I_α^*, we employ the following naive bootstrap method.

For each sample $X_1, \cdots, X_n$, we first estimate ϵ_t by $\hat{\epsilon}_t = X_t - \hat{b} X_{t-2} \hat{\epsilon}_{t-1}$ recursively. Next draw $1,000$ resamples with size n from $\hat{\epsilon}_1, \cdots, \hat{\epsilon}_n$, say $\hat{\epsilon}_1^{*(j)}, \cdots, \hat{\epsilon}_n^{*(j)}$ for $j = 1, \cdots, 1000$. For each j, we generate bootstrap sample $X_t^{*(j)} = \hat{b} X_{t-2}^{*(j)} \hat{\epsilon}_{t-1}^{*(j)} + \hat{\epsilon}_t^{*(j)}$. Based on the bootstrap samples, we compute the conditional least squares estimator so that we have 1000 bootstrapped estimators $\hat{b}^{*(1)}, \cdots, \hat{b}^{*(1000)}$. Let $c_{(1-\alpha)/2}$ and $c_{(1+\alpha)/2}$ denote the largest $[1000(1-\alpha)/2]$th and $[1000(1+\alpha)/2]$th values of $\hat{b}^{*(1)} - \hat{b}, \cdots, \hat{b}^{*(1000)} - \hat{b}$. Therefore the bootstrap confidence interval for b with level α is defined as

$$I_\alpha^* = [\hat{b} - c_{(1+\alpha)/2}, \quad \hat{b} - c_{(1-\alpha)/2}].$$

Table below reports the empirical coverage probabilities for the above four intervals, which show that the proposed three profile empirical likelihood methods perform better than the bootstrap method especially for the case of $b = 0.4$ and small sample

Table 17: Empirical coverage probabilities for the proposed profile empirical likelihood confidence intervals $I_\alpha^{(1)}, I_\alpha^{(2)}, I_\alpha^{(3)}$ and the bootstrap confidence interval I_α^* with levels $\alpha = 0.9$ and 0.95.

(n, b)	$I_{0.9}^*$	$I_{0.9}^{(1)}$	$I_{0.9}^{(2)}$	$I_{0.9}^{(3)}$	$I_{0.95}^*$	$I_{0.95}^{(1)}$	$I_{0.95}^{(2)}$	$I_{0.95}^{(3)}$
$(100, 0.1)$	0.9612	0.8719	0.8695	0.8536	0.9924	0.9256	0.9252	0.9116
$(100, 0.2)$	0.9252	0.8706	0.8618	0.8525	0.9724	0.9279	0.9260	0.9146
$(100, 0.3)$	0.8587	0.8644	0.8697	0.8545	0.9343	0.9226	0.9273	0.9131
$(100, 0.4)$	0.7674	0.8607	0.8689	0.8501	0.8568	0.9191	0.9277	0.9147
$(200, 0.1)$	0.9357	0.8902	0.8932	0.8881	0.9805	0.9395	0.9400	0.9373
$(200, 0.2)$	0.9161	0.8903	0.8913	0.8873	0.9597	0.9423	0.9421	0.9383
$(200, 0.3)$	0.8808	0.8875	0.8908	0.8822	0.9238	0.9420	0.9423	0.9390
$(200, 0.4)$	0.8151	0.8833	0.8868	0.8814	0.8657	0.9385	0.9408	0.9417
$(1000, 0.1)$	0.9308	0.9028	0.8974	0.8965	0.9575	0.9497	0.9494	0.9494
$(1000, 0.2)$	0.9026	0.9008	0.8962	0.8973	0.9353	0.9513	0.9493	0.9497
$(1000, 0.3)$	0.9016	0.8976	0.8972	0.8968	0.9467	0.9487	0.9508	0.9487
$(1000, 0.4)$	0.8844	0.8933	0.8979	0.8925	0.9231	0.9464	0.9499	0.9449

size. When the sample size is small, the profile empirical likelihood interval $I_\alpha^{(3)}$ performs worse than the other two profile empirical likelihood intervals since it involves one more estimating equation.

4.4 Conclusions

The coefficient parameters in bilinear time series models are easily estimated by the conditional least squares estimators. Unfortunately, it remains unknown on how to explicitly estimate the asymptotic variances of the conditional least squares estimators without assuming normality for errors. This chapter proposes some profile empirical likelihood methods based on some weighted score equations to construct confidence intervals for the coefficient parameter without estimating the asymptotic variance, and shows that Wilks's theorem holds. The proposed methods are easy to implement by using the R package 'emplik' and have good finite sample behavior.

4.5 Proofs

Before proving Proposition 4.2.1, we need some preliminary lemmas.

Lemma 1. *Under conditions of Proposition 4.2.1, we have*

$$\sup_{3\le t\le n}\sup_{\sigma^2\in U_n}\|Y_t(b_0,\sigma^2)\| = o_p(n^{\frac{1}{2}\gamma}),$$

where $\|\cdot\|$ denotes the L_2 norm.

Proof. Write

$$\begin{aligned}
Y_{t1}(b_0,\sigma^2) &= \frac{(X_t^2-\sigma_0^2-b_0^2\sigma_0^2X_{t-2}^2)2b_0X_{t-2}^2+(X_tX_{t-1}-b_0\sigma_0^2X_{t-2})X_{t-2}}{d+X_{t-2}^2} \\
&\quad -(\sigma^2-\sigma_0^2)\{(1+b_0^2X_{t-2}^2)2b_0+b_0\}\frac{X_{t-2}^2}{d+X_{t-2}^2} \\
&= I_1(t)-I_2(t). \qquad (59)
\end{aligned}$$

Using the inequalities $x^2/(d+x^2)<1$ and $|x|/(d+x^2)\le 1/(2\sqrt{d})$ for any x, we can show that there exist constants C_1 and C_2 such that for any $\varepsilon>0$,

$$\begin{aligned}
P\left\{\sup_{3\le t\le n}|I_1(t)|\ge\varepsilon n^{\frac{1}{2}\gamma}\right\} &\le \sum_{t=3}^{n}\varepsilon^{-2\delta}n^{-\frac{\delta}{\gamma}}E|I_1(t)|^{2\delta} \\
&\le \sum_{t=3}^{n}\varepsilon^{-2\delta}n^{-\frac{\delta}{\gamma}}C_1(1+E|X_t|^{4\delta}+E|X_t|^{2\delta}) \\
&\longrightarrow 0, \qquad (60) \\
P\left\{\sup_{3\le t\le}|I_2(t)|\ge\varepsilon n^{\frac{1}{2}\gamma}\right\} &\le \sum_{t=3}^{n}\varepsilon^{-2\gamma}n^{-2}E|I_2(t)|^{2\gamma} \\
&\le \sum_{t=3}^{n}\varepsilon^{-2\gamma}n^{-2}C_2(1+E|X_t|^{4\gamma}+E|X_t|^{2\gamma}) \\
&\longrightarrow 0 \qquad (61)
\end{aligned}$$

as $n\to\infty$. Hence, (60) and (61) imply that $\sup_{3\le t\le n}\sup_{\sigma^2\in U_n}|Y_{t1}(b_0,\sigma^2)| = o_p(n^{\frac{1}{2}\gamma})$. Similarly, we have $\sup_{3\le t\le n}\sup_{\sigma^2\in U_n}|Y_{t2}(b_0,\sigma^2)| = o_p(n^{\frac{1}{2}\gamma})$. Thus the lemma holds. □

Lemma 2. *Under conditions of Proposition 4.2.1, we have*

$$\text{(i)} \quad \sup_{\sigma^2 \in U_n} \left\| \frac{1}{n-2} \sum_{t=3}^{n} \frac{\partial}{\partial \sigma^2} Y_t(b_0, \sigma^2) - E\left\{ \frac{\partial}{\partial \sigma^2} Y_3(b_0, \sigma_0^2) \right\} \right\| = o_p(1);$$

$$\text{(ii)} \quad \sup_{\sigma^2 \in U_n} \left\| \frac{1}{n-2} \sum_{t=3}^{n} Y_t(b_0, \sigma^2) Y_t^T(b_0, \sigma^2) - E\{Y_3(b_0, \sigma_0^2) Y_3^T(b_0, \sigma_0^2)\} \right\| = o_p(1);$$

$$\text{(iii)} \quad \frac{1}{\sqrt{n-2}} \sum_{t=3}^{n} Y_t(b_0, \sigma_0^2) \xrightarrow{d} N(0, E\{Y_3(b_0, \sigma_0^2) Y_3^T(b_0, \sigma_0^2)\}).$$

Proof. Since $\partial Y_t(b_0, \sigma^2)/\partial \sigma^2$ is independent of σ^2, (i) simply follows from the ergodic theorem. From (59) we know that $Y_{t1}^2(b_0, \sigma^2) = (I_1(t) - I_2(t))^2 = I_1^2(t) - 2I_1(t)I_2(t) + I_2^2(t)$. It is easy to show that

$$\begin{aligned} &\frac{1}{n-2} \sum_{t=3}^{n} I_1^2(t) \xrightarrow{P} E\{Y_{31}^2(b_0, \sigma_0^2)\}, \\ &\frac{1}{n-2} \sum_{t=3}^{n} I_2^2(t) \leq n^{-\frac{1}{\gamma}} \frac{1}{n-2} \sum_{t=3}^{n} \{(1 + b_0^2 X_{t-2}^2) 2b_0 + b_0\}^2 \xrightarrow{P} 0, \end{aligned}$$

which imply that

$$\sup_{\sigma^2 \in U_n} \left| \frac{1}{n-2} \sum_{t=3}^{n} Y_{t1}^2(b_0, \sigma^2) - E\{Y_{31}^2(b_0, \sigma_0^2)\} \right| = o_p(1).$$

Similarly, we can show that

$$\sup_{\sigma^2 \in U_n} \left| \frac{1}{n-2} \sum_{t=3}^{n} Y_{t2}^2(b_0, \sigma^2) - E\{Y_{32}^2(b_0, \sigma_0^2)\} \right| = o_p(1)$$

and

$$\sup_{\sigma^2 \in U_n} \left| \frac{1}{n-2} \sum_{t=3}^{n} Y_{t1}(b_0, \sigma^2) Y_{t2}(b_0, \sigma^2) - E\{Y_{31}(b_0, \sigma_0^2) Y_{32}(b_0, \sigma_0^2)\} \right| = o_p(1).$$

Thus (ii) follows. Since $Y_t(b_0, \sigma_0^2)$ is a martingale difference sequence, (iii) follows from the central limit theorem for martingales (see Hall and Heyde, 1980). This completes the proof of Lemma 2. □

Proof of Proposition 4.2.1. This can be proved in the same way as Lemma 1 of Qin and Lawless (1994) by using Lemmas 1 and 2.

Proof of Theorem 6. It follows from the same arguments in the proof of Theorem 1 of Qin and Lawless (1994) by using Proposition 4.2.1, Lemmas 1 and 2.

CHAPTER V

R CODES

5.1 Codes for Chapter II

```
{
library(SparseM)
library(quantreg)
library(emplik)

critical1=0.347 #90% for  CS test
critical1a=0.461#95% for CS test
#Calculate Cramer-Smirnov  test
CS_Norm=function(sample,theta){
n=length(sample);
X=pnorm(sample,mean=theta[1],sd=theta[2]);
Y=sort(X);
tmp=sum((Y-(2*(1:n)-1)/2/n)^2)+1/12/n;
return (c(as.integer(tmp>critical1),as.integer(tmp>critical1a)));}

CS_T=function(sample,theta){
n=length(sample)
X=pt(sample,df=theta[1])
Y=sort(X)
tmp=sum((Y-(2*(1:n)-1)/2/n)^2)+1/12/n
return(c(as.integer(tmp>critical1),as.integer(tmp>critical1a)))}

CS_Gamma=function(sample,theta){
n=length(sample);
X=pgamma(sample,shape=theta[1],scale=theta[2]);
Y=sort(X);
tmp=sum((Y-(2*(1:n)-1)/2/n)^2)+1/12/n;
return (c(as.integer(tmp>critical1),as.integer(tmp>critical1a)));}

CS_chisq=function(sample,beta){
n=length(sample)
X=pchisq(sample,df=beta)
Y=sort(X)
tmp=sum((Y-(2*(1:n)-1)/2/n)^2)+1/12/n
return(c(as.integer(tmp>critical1),as.integer(tmp>critical1a)))}

#################
CS_Norm_Est=function(sample,BN){
```

```
n=length(sample)
theta1=mean(sample)
theta2=(mean((sample-theta1)^2))^0.5
X=pnorm(sample,mean=theta1,sd=theta2)
Y=sort(X)
tmp=sum((Y-(2*(1:n)-1)/2/n)^2)+1/12/n
tmp1=0*(1:BN)
for(j in 1:BN){
U=runif(n)
Bsample=qnorm(U,mean=theta1,sd=theta2)
Btheta1=mean(Bsample)
Btheta2=(mean((Bsample-Btheta1)^2))^0.5
BX=pnorm(Bsample,mean=Btheta1,sd=Btheta2)
BY=sort(BX)
tmp1[j]=sum((BY-(2*(1:n)-1)/2/n)^2)+1/12/n
}
Z=sort(tmp1)
return(c(as.integer(tmp>Z[as.integer(0.9*BN)]),as.integer(tmp>Z[as.integer(0.95*BN)])))
}

CS_T_Est=function(sample,BN){
n=length(sample)
sigma2=mean(sample^2)
v=max(2*sigma2/(sigma2-1),2.001)
X=pt(sample,df=v)
Y=sort(X)
tmp=sum((Y-(2*(1:n)-1)/2/n)^2)+1/12/n
tmp1=0*(1:BN)
for(j in 1:BN){
U=runif(n)
Bsample=qt(U,df=v)
Bsigma2=mean(Bsample^2)
Bv=max(2*Bsigma2/(Bsigma2-1),2.001)
BX=pt(Bsample,df=Bv)
BY=sort(BX)
tmp1[j]=sum((BY-(2*(1:n)-1)/2/n)^2)+1/12/n
}
Z=sort(tmp1)
return(c(as.integer(tmp>Z[as.integer(0.9*BN)]),as.integer(tmp>Z[as.integer(0.95*BN)])))}

CS_Gamma_Est=function(sample,BN){
n=length(sample)
theta1=max(mean(sample),0.0000001)
theta2=var(sample)
theta2=theta2/theta1
theta1=theta1/theta2
X=pgamma(sample,shape=theta1,scale=theta2)
Y=sort(X)
tmp=sum((Y-(2*(1:n)-1)/2/n)^2)+1/12/n
```

```
tmp1=0*(1:BN)
for(j in 1:BN){
Bsample=rgamma(n,shape=theta1,scale=theta2)
Btheta1=mean(Bsample)
Btheta2=var(Bsample)
Btheta2=Btheta2/Btheta1
Btheta1=Btheta1/Btheta2
BX=pgamma(Bsample,shape=Btheta1,scale=Btheta2)
BY=sort(BX)
tmp1[j]=sum((BY-(2*(1:n)-1)/2/n)^2)+1/12/n
}
Z=sort(tmp1)
return(c(as.integer(tmp>Z[as.integer(0.9*BN)]),as.integer(tmp>Z[as.integer(0.95*BN)])))
}

CS_chisq_Est=function(sample,BN){
n=length(sample)
v=mean(sample)
X=pchisq(sample,df=v)
Y=sort(X)
tmp=sum((Y-(2*(1:n)-1)/2/n)^2)+1/12/n
tmp1=0*(1:BN)
for(j in 1:BN){
Bsample=rchisq(n,df=v)
Bv=mean(Bsample)

BX=pt(Bsample,df=Bv)
BY=sort(BX)
tmp1[j]=sum((BY-(2*(1:n)-1)/2/n)^2)+1/12/n
}
Z=sort(tmp1)
return(c(as.integer(tmp>Z[as.integer(0.9*BN)]),as.integer(tmp>Z[as.integer(0.95*BN)])))}

#JELM based Cramer-von Mises test
JEL_CM1_Norm=function(sample,theta){
n=length(sample)
X=pnorm(sample,mean=theta[1],sd=theta[2])
OX=sort(X)
T1=1-sum((2*(1:n)-1)*OX)/n^2-1/3
T2=1/2-sum(OX^2)/2/n-1/3
#T3=mean(OX)-1/2
T3=mean((1-OX^2)^0.5)-pi/4
JT1=0*(1:n)
JT2=0*(1:n)
JT3=0*(1:n)
for(i in 1:n){
JOX=sort(X[-i])
tmp1=1-sum((2*(1:(n-1))-1)*JOX)/(n-1)^2-1/3
tmp2=1/2-sum(JOX^2)/2/(n-1)-1/3
```

```
#tmp3=mean(JOX)-1/2
tmp3=mean((1-JOX^2)^0.5)-pi/4
JT1[i]=tmp1
JT2[i]=tmp2
JT3[i]=tmp3
}
W1=n*T1-(n-1)*JT1
W2=n*(T2+T3)-(n-1)*(JT2+JT3)
ratio1=el.test(cbind(W1,W2),c(0,0))$'-2LLR'
W2=n*(T2-T3)-(n-1)*(JT2-JT3)
ratio2=el.test(cbind(W1,W2),c(0,0))$'-2LLR'
return (c(as.integer(ratio1>qchisq(0.9,df=2)),as.integer(ratio1>qchisq(0.95,df=2)),
as.integer(ratio2>qchisq(0.9,df=2)),as.integer(ratio2>qchisq(0.95,df=2))))
}

JEL_CM1_T=function(sample,theta){
n=length(sample)
X=pt(sample,df=theta[1])
OX=sort(X)
T1=1-sum((2*(1:n)-1)*OX)/n^2-1/3
T2=1/2-sum(OX^2)/2/n-1/3
#T3=mean(OX)-1/2
T3=mean((1-OX^2)^0.5)-pi/4
JT1=0*(1:n)
JT2=0*(1:n)
JT3=0*(1:n)
for(i in 1:n){
JOX=sort(X[-i])
tmp1=1-sum((2*(1:(n-1))-1)*JOX)/(n-1)^2-1/3
tmp2=1/2-sum(JOX^2)/2/(n-1)-1/3
#tmp3=mean(JOX)-1/2
tmp3=mean((1-JOX^2)^0.5)-pi/4
JT1[i]=tmp1
JT2[i]=tmp2
JT3[i]=tmp3
}
W1=n*T1-(n-1)*JT1
W2=n*(T2+T3)-(n-1)*(JT2+JT3)
ratio1=el.test(cbind(W1,W2),c(0,0))$'-2LLR'
W2=n*(T2-T3)-(n-1)*(JT2-JT3)
ratio2=el.test(cbind(W1,W2),c(0,0))$'-2LLR'
return (c(as.integer(ratio1>qchisq(0.9,df=2)),as.integer(ratio1>qchisq(0.95,df=2)),
as.integer(ratio2>qchisq(0.9,df=2)),as.integer(ratio2>qchisq(0.95,df=2))))
}

JEL_CM1_Gamma=function(sample,theta){
n=length(sample)
X=pgamma(sample,shape=theta[1],scale=theta[2])
OX=sort(X)
```

```
T1=1-sum((2*(1:n)-1)*OX)/n^2-1/3
T2=1/2-sum(OX^2)/2/n-1/3
#T3=mean(OX)-1/2
T3=mean((1-OX^2)^0.5)-pi/4
JT1=0*(1:n)
JT2=0*(1:n)
JT3=0*(1:n)
for(i in 1:n){
JOX=sort(X[-i])
tmp1=1-sum((2*(1:(n-1))-1)*JOX)/(n-1)^2-1/3
tmp2=1/2-sum(JOX^2)/2/(n-1)-1/3
#tmp3=mean(JOX)-1/2
tmp3=mean((1-JOX^2)^0.5)-pi/4
JT1[i]=tmp1
JT2[i]=tmp2
JT3[i]=tmp3
}
W1=n*T1-(n-1)*JT1
W2=n*(T2+T3)-(n-1)*(JT2+JT3)
ratio1=el.test(cbind(W1,W2),c(0,0))$'-2LLR'
W2=n*(T2-T3)-(n-1)*(JT2-JT3)
ratio2=el.test(cbind(W1,W2),c(0,0))$'-2LLR'
return (c(as.integer(ratio1>qchisq(0.9,df=2)),as.integer(ratio1>qchisq(0.95,df=2)),
as.integer(ratio2>qchisq(0.9,df=2)),as.integer(ratio2>qchisq(0.95,df=2))))
}
JEL_CM1_chisq=function(sample,theta){
n=length(sample)
X=pchisq(sample,df=theta[1])
OX=sort(X)
T1=1-sum((2*(1:n)-1)*OX)/n^2-1/3
T2=1/2-sum(OX^2)/2/n-1/3
#T3=mean(OX)-1/2
T3=mean((1-OX^2)^0.5)-pi/4
JT1=0*(1:n)
JT2=0*(1:n)
JT3=0*(1:n)
for(i in 1:n){
JOX=sort(X[-i])
tmp1=1-sum((2*(1:(n-1))-1)*JOX)/(n-1)^2-1/3
tmp2=1/2-sum(JOX^2)/2/(n-1)-1/3
#tmp3=mean(JOX)-1/2
tmp3=mean((1-JOX^2)^0.5)-pi/4
JT1[i]=tmp1
JT2[i]=tmp2
JT3[i]=tmp3
}
W1=n*T1-(n-1)*JT1
W2=n*(T2+T3)-(n-1)*(JT2+JT3)
ratio1=el.test(cbind(W1,W2),c(0,0))$'-2LLR'
```

```
W2=n*(T2-T3)-(n-1)*(JT2-JT3)
ratio2=el.test(cbind(W1,W2),c(0,0))$'-2LLR'
return (c(as.integer(ratio1>qchisq(0.9,df=2)),as.integer(ratio1>qchisq(0.95,df=2)),
as.integer(ratio2>qchisq(0.9,df=2)),as.integer(ratio2>qchisq(0.95,df=2))))
}

##############
JEL_CM1_Norm_Est=function(sample){
n=length(sample)
theta1=mean(sample)
theta2=(mean((sample-theta1)^2))^0.5
X=pnorm(sample,mean=theta1,sd=theta2)
OX=sort(X)
T1=1-sum((2*(1:n)-1)*OX)/n^2-1/3
T2=1/2-sum(OX^2)/2/n-1/3
#T3=mean(OX)-1/2
T3=mean((1-OX^2)^0.5)-pi/4
JT1=0*(1:n)
JT2=0*(1:n)
JT3=0*(1:n)
for(i in 1:n){
theta1=mean(sample[-i])
theta2=(mean((sample[-i]-theta1)^2))^0.5
JX=pnorm(sample[-i],mean=theta1,sd=theta2)
JOX=sort(JX)
tmp1=1-sum((2*(1:(n-1))-1)*JOX)/(n-1)^2-1/3
tmp2=1/2-sum(JOX^2)/2/(n-1)-1/3
#tmp3=mean(JOX)-1/2
tmp3=mean((1-JOX^2)^0.5)-pi/4
JT1[i]=tmp1
JT2[i]=tmp2
JT3[i]=tmp3
}
W1=n*T1-(n-1)*JT1
W2=n*(T2+T3)-(n-1)*(JT2+JT3)
ratio1=el.test(cbind(W1,W2),c(0,0))$'-2LLR'
W2=n*(T2-T3)-(n-1)*(JT2-JT3)
ratio2=el.test(cbind(W1,W2),c(0,0))$'-2LLR'
return (c(as.integer(ratio1>qchisq(0.9,df=2)),as.integer(ratio1>qchisq(0.95,df=2)),
as.integer(ratio2>qchisq(0.9,df=2)),as.integer(ratio2>qchisq(0.95,df=2))))
}

JEL_CM1_T_Est=function(sample){
n=length(sample)
tmp0=mean(sample^2)
theta=max(2*tmp0/(tmp0-1),2.001)
X=pt(sample,df=theta)
OX=sort(X)
T1=1-sum((2*(1:n)-1)*OX)/n^2-1/3
```

```
T2=1/2-sum(OX^2)/2/n-1/3
#T3=mean(OX)-1/2
T3=mean((1-OX^2)^0.5)-pi/4
JT1=0*(1:n)
JT2=0*(1:n)
JT3=0*(1:n)
for(i in 1:n){
tmp0=mean((sample[-i])^2)
theta=max(2*tmp0/(tmp0-1),2.001)
JX=pt(sample[-i],df=theta)
JOX=sort(JX)
tmp1=1-sum((2*(1:(n-1))-1)*JOX)/(n-1)^2-1/3
tmp2=1/2-sum(JOX^2)/2/(n-1)-1/3
#tmp3=mean(JOX)-1/2
tmp3=mean((1-JOX^2)^0.5)-pi/4
JT1[i]=tmp1
JT2[i]=tmp2
JT3[i]=tmp3
}
W1=n*T1-(n-1)*JT1
W2=n*(T2+T3)-(n-1)*(JT2+JT3)
ratio1=el.test(cbind(W1,W2),c(0,0))$'-2LLR'
W2=n*(T2-T3)-(n-1)*(JT2-JT3)
ratio2=el.test(cbind(W1,W2),c(0,0))$'-2LLR'
return (c(as.integer(ratio1>qchisq(0.9,df=2)),as.integer(ratio1>qchisq(0.95,df=2)),
as.integer(ratio2>qchisq(0.9,df=2)),as.integer(ratio2>qchisq(0.95,df=2))))
}
JEL_CM1_Gamma_Est=function(sample){
n=length(sample)
theta1=max(mean(sample),0.0000001)
theta2=var(sample)
theta2=theta2/theta1
theta1=theta1/theta2
X=pgamma(sample,shape=theta1,scale=theta2)
OX=sort(X)
T1=1-sum((2*(1:n)-1)*OX)/n^2-1/3
T2=1/2-sum(OX^2)/2/n-1/3
#T3=mean(OX)-1/2
T3=mean((1-OX^2)^0.5)-pi/4
JT1=0*(1:n)
JT2=0*(1:n)
JT3=0*(1:n)
for(i in 1:n){
theta1=max(mean(sample[-i]),0.0000001)
theta2=var(sample[-i])
theta2=theta2/theta1
theta1=theta1/theta2
JX=pgamma(sample[-i],shape=theta1,scale=theta2)
JOX=sort(JX)
```

```
tmp1=1-sum((2*(1:(n-1))-1)*JOX)/(n-1)^2-1/3
tmp2=1/2-sum(JOX^2)/2/(n-1)-1/3
#tmp3=mean(JOX)-1/2
tmp3=mean((1-JOX^2)^0.5)-pi/4
JT1[i]=tmp1
JT2[i]=tmp2
JT3[i]=tmp3
}
W1=n*T1-(n-1)*JT1
W2=n*(T2+T3)-(n-1)*(JT2+JT3)
ratio1=el.test(cbind(W1,W2),c(0,0))$'-2LLR'
W2=n*(T2-T3)-(n-1)*(JT2-JT3)
ratio2=el.test(cbind(W1,W2),c(0,0))$'-2LLR'
return (c(as.integer(ratio1>qchisq(0.9,df=2)),as.integer(ratio1>qchisq(0.95,df=2)),
as.integer(ratio2>qchisq(0.9,df=2)),as.integer(ratio2>qchisq(0.95,df=2))))
}

JEL_CM1_chisq_Est=function(sample){
n=length(sample)
theta=mean(sample)
X=pchisq(sample,df=theta)
OX=sort(X)
T1=1-sum((2*(1:n)-1)*OX)/n^2-1/3
T2=1/2-sum(OX^2)/2/n-1/3
#T3=mean(OX)-1/2
T3=mean((1-OX^2)^0.5)-pi/4
JT1=0*(1:n)
JT2=0*(1:n)
JT3=0*(1:n)
for(i in 1:n){
theta=mean((sample[-i]))

JX=pchisq(sample[-i],df=theta)
JOX=sort(JX)
tmp1=1-sum((2*(1:(n-1))-1)*JOX)/(n-1)^2-1/3
tmp2=1/2-sum(JOX^2)/2/(n-1)-1/3
#tmp3=mean(JOX)-1/2
tmp3=mean((1-JOX^2)^0.5)-pi/4
JT1[i]=tmp1
JT2[i]=tmp2
JT3[i]=tmp3
}
W1=n*T1-(n-1)*JT1
W2=n*(T2+T3)-(n-1)*(JT2+JT3)
ratio1=el.test(cbind(W1,W2),c(0,0))$'-2LLR'
W2=n*(T2-T3)-(n-1)*(JT2-JT3)
ratio2=el.test(cbind(W1,W2),c(0,0))$'-2LLR'
return (c(as.integer(ratio1>qchisq(0.9,df=2)),as.integer(ratio1>qchisq(0.95,df=2)),
as.integer(ratio2>qchisq(0.9,df=2)),as.integer(ratio2>qchisq(0.95,df=2))))}
```

```
###############
#simple
Einmahl_Norm=function(sample,theta,BN){
n=length(sample);
X=pnorm(sample,theta[1],theta[2])
X=pmax(X,0.00000001)
X=pmin(X,0.99999999)
Y=sort(X);
tmp=(2*n*((1-Y[1])*log(1-Y[1])+Y[1]+1+Y[n]*log(Y[n])-Y[n])
    +sum(-2*(1:(n-1))*log(n/(1:(n-1)))*(Y[2:n]-Y[1:(n-1)])
    -2*(1:(n-1))*(Y[2:n]*log(Y[2:n])-Y[2:n]-Y[1:(n-1)]*log(Y[1:(n-1)])
    +Y[1:(n-1)])+2*(n-(1:(n-1)))*log(1-(1:(n-1))/n)*(Y[2:n]-Y[1:(n-1)])
    +2*(n-(1:(n-1)))*((1-Y[2:n])*log(1-Y[2:n])+Y[2:n]-Y[1:(n-1)]
    -(1-Y[1:(n-1)])*log(1-Y[1:(n-1)]))))

tmp1=0*(1:BN)
for(j in 1:BN){
Bsample=rnorm(n,theta[1],theta[2])
BX=pnorm(Bsample,theta[1],theta[2])
BX=pmax(BX,0.0000001)
BX=pmin(BX,0.9999999)
BY=sort(BX)
tmp1[j]=(2*n*((1-BY[1])*log(1-BY[1])+BY[1]+1+BY[n]*log(BY[n])-BY[n])
        +sum(-2*(1:(n-1))*log(n/(1:(n-1)))*(BY[2:n]-BY[1:(n-1)])
        -2*(1:(n-1))*(BY[2:n]*log(BY[2:n])-BY[2:n]
        -BY[1:(n-1)]*log(BY[1:(n-1)])+BY[1:(n-1)])
        +2*(n-(1:(n-1)))*log(1-(1:(n-1))/n)*(BY[2:n]-BY[1:(n-1)])
        +2*(n-(1:(n-1)))*((1-BY[2:n])*log(1-BY[2:n])+BY[2:n]-BY[1:(n-1)]
        -(1-BY[1:(n-1)])*log(1-BY[1:(n-1)]))))
               }
Z=sort(tmp1)
return(c(as.integer(tmp>Z[as.integer(0.9*BN)]),as.integer(tmp>Z[as.integer(0.95*BN)])))}

Einmahl_T=function(sample,beta,BN){
n=length(sample)
X=pt(sample,df=beta)
Y=sort(X)
tmp=(2*n*((1-Y[1])*log(1-Y[1])+Y[1]+1+Y[n]*log(Y[n])-Y[n])
    +sum(-2*(1:(n-1))*log(n/(1:(n-1)))*(Y[2:n]-Y[1:(n-1)])
    -2*(1:(n-1))*(Y[2:n]*log(Y[2:n])-Y[2:n]-Y[1:(n-1)]*log(Y[1:(n-1)])
    +Y[1:(n-1)])+2*(n-(1:(n-1)))*log(1-(1:(n-1))/n)*(Y[2:n]-Y[1:(n-1)])
    +2*(n-(1:(n-1)))*((1-Y[2:n])*log(1-Y[2:n])+Y[2:n]-Y[1:(n-1)]
    -(1-Y[1:(n-1)])*log(1-Y[1:(n-1)]))))

tmp1=0*(1:BN)
for(j in 1:BN){
Bsample=rchisq(n,beta)
BX=pt(Bsample,beta)
```

```
BY=sort(BX)
tmp1[j]=(2*n*((1-BY[1])*log(1-BY[1])+BY[1]+1+BY[n]*log(BY[n])-BY[n])
        +sum(-2*(1:(n-1))*log(n/(1:(n-1)))*(BY[2:n]-BY[1:(n-1)])
        -2*(1:(n-1))*(BY[2:n]*log(BY[2:n])-BY[2:n]
        -BY[1:(n-1)]*log(BY[1:(n-1)])+BY[1:(n-1)])
        +2*(n-(1:(n-1)))*log(1-(1:(n-1))/n)*(BY[2:n]-BY[1:(n-1)])
        +2*(n-(1:(n-1)))*((1-BY[2:n])*log(1-BY[2:n])+BY[2:n]-BY[1:(n-1)]
        -(1-BY[1:(n-1)])*log(1-BY[1:(n-1)]))))
               }
Z=sort(tmp1)
return(c(as.integer(tmp>Z[as.integer(0.9*BN)]),as.integer(tmp>Z[as.integer(0.95*BN)])))
}

Einmahl_Gamma=function(sample,theta,BN){
n=length(sample);
X=pgamma(sample,shape=theta[1],scale=theta[2]);
X=pmax(X,0.00000001)
X=pmin(X,0.99999999)
Y=sort(X);
tmp=(2*n*((1-Y[1])*log(1-Y[1])+Y[1]+1+Y[n]*log(Y[n])-Y[n])
    +sum(-2*(1:(n-1))*log(n/(1:(n-1)))*(Y[2:n]-Y[1:(n-1)])
    -2*(1:(n-1))*(Y[2:n]*log(Y[2:n])-Y[2:n]-Y[1:(n-1)]*log(Y[1:(n-1)])
    +Y[1:(n-1)])+2*(n-(1:(n-1)))*log(1-(1:(n-1))/n)*(Y[2:n]-Y[1:(n-1)])
    +2*(n-(1:(n-1)))*((1-Y[2:n])*log(1-Y[2:n])+Y[2:n]-Y[1:(n-1)]
    -(1-Y[1:(n-1)])*log(1-Y[1:(n-1)]))))

tmp1=0*(1:BN)
for(j in 1:BN){
Bsample=rgamma(n,shape=theta[1],scale=theta[2])
BX=pgamma(Bsample,shape=theta[1],scale=theta[2])
BX=pmax(BX,0.00000001)
BX=pmin(BX,0.99999999)
BY=sort(BX)
tmp1[j]=(2*n*((1-BY[1])*log(1-BY[1])+BY[1]+1+BY[n]*log(BY[n])-BY[n])
        +sum(-2*(1:(n-1))*log(n/(1:(n-1)))*(BY[2:n]-BY[1:(n-1)])
        -2*(1:(n-1))*(BY[2:n]*log(BY[2:n])-BY[2:n]
        -BY[1:(n-1)]*log(BY[1:(n-1)])+BY[1:(n-1)])
        +2*(n-(1:(n-1)))*log(1-(1:(n-1))/n)*(BY[2:n]-BY[1:(n-1)])
        +2*(n-(1:(n-1)))*((1-BY[2:n])*log(1-BY[2:n])+BY[2:n]-BY[1:(n-1)]
        -(1-BY[1:(n-1)])*log(1-BY[1:(n-1)]))))
               }
Z=sort(tmp1)
#print(c(Y[1],Y[n]))
return(c(as.integer(tmp>Z[as.integer(0.9*BN)]),as.integer(tmp>Z[as.integer(0.95*BN)])))
}
Einmahl_chisq=function(sample,beta,BN){
n=length(sample)
X=pchisq(sample,df=beta)
Y=sort(X)
```

```
tmp=(2*n*((1-Y[1])*log(1-Y[1])+Y[1]+1+Y[n]*log(Y[n])-Y[n])
    +sum(-2*(1:(n-1))*log(n/(1:(n-1)))*(Y[2:n]-Y[1:(n-1)])
    -2*(1:(n-1))*(Y[2:n]*log(Y[2:n])-Y[2:n]-Y[1:(n-1)]*log(Y[1:(n-1)])
    +Y[1:(n-1)])+2*(n-(1:(n-1)))*log(1-(1:(n-1))/n)*(Y[2:n]-Y[1:(n-1)])
    +2*(n-(1:(n-1)))*((1-Y[2:n])*log(1-Y[2:n])+Y[2:n]-Y[1:(n-1)]
    -(1-Y[1:(n-1)])*log(1-Y[1:(n-1)]))))

tmp1=0*(1:BN)
for(j in 1:BN){
Bsample=rchisq(n,beta)
BX=pchisq(Bsample,beta)
BY=sort(BX)
tmp1[j]=(2*n*((1-BY[1])*log(1-BY[1])+BY[1]+1+BY[n]*log(BY[n])-BY[n])
        +sum(-2*(1:(n-1))*log(n/(1:(n-1)))*(BY[2:n]-BY[1:(n-1)])
        -2*(1:(n-1))*(BY[2:n]*log(BY[2:n])-BY[2:n]
        -BY[1:(n-1)]*log(BY[1:(n-1)])+BY[1:(n-1)])
        +2*(n-(1:(n-1)))*log(1-(1:(n-1))/n)*(BY[2:n]-BY[1:(n-1)])
        +2*(n-(1:(n-1)))*((1-BY[2:n])*log(1-BY[2:n])+BY[2:n]-BY[1:(n-1)]
        -(1-BY[1:(n-1)])*log(1-BY[1:(n-1)]))))
               }
Z=sort(tmp1)
return(c(as.integer(tmp>Z[as.integer(0.9*BN)]),as.integer(tmp>Z[as.integer(0.95*BN)])))
}
##################
Jager_Norm=function(sample,theta,BN){
n=length(sample);
X=pnorm(sample,theta[1],theta[2]);
Y=sort(X);
tmp=sum(-4/3-8/3*sqrt(((1:n)-1)/n)*Y^(3/2)+8/3*sqrt((1:n)/n)*Y^(3/2)
+8/3*sqrt(1+1/n-(1:n)/n)*(1-Y)^(3/2)-8/3*sqrt(1-(1:n)/n)*(1-Y)^(3/2))
tmp1=0*(1:BN)
for(j in 1:BN){
Bsample=rnorm(n,theta[1],theta[2])
BX=pnorm(Bsample,theta[1],theta[2])
BY=sort(BX)
tmp1[j]=sum(-4/3-8/3*sqrt(((1:n)-1)/n)*BY^(3/2)+8/3*sqrt((1:n)/n)*BY^(3/2)
+8/3*sqrt(1+1/n-(1:n)/n)*(1-BY)^(3/2)-8/3*sqrt(1-(1:n)/n)*(1-BY)^(3/2))
               }
Z=sort(tmp1)
return(c(as.integer(tmp>Z[as.integer(0.9*BN)]),as.integer(tmp>Z[as.integer(0.95*BN)])))
}
Jager_T=function(sample,beta,BN){
n=length(sample)
X=pt(sample,df=beta)
Y=sort(X)
tmp=sum(-4/3-8/3*sqrt(((1:n)-1)/n)*Y^(3/2)+8/3*sqrt((1:n)/n)*Y^(3/2)
+8/3*sqrt(1+1/n-(1:n)/n)*(1-Y)^(3/2)-8/3*sqrt(1-(1:n)/n)*(1-Y)^(3/2))
tmp1=0*(1:BN)
for(j in 1:BN){
```

```
Bsample=rt(n,beta)
BX=pt(Bsample,beta)
BY=sort(BX)
tmp1[j]=sum(-4/3-8/3*sqrt(((1:n)-1)/n)*BY^(3/2)+8/3*sqrt((1:n)/n)*BY^(3/2)
+8/3*sqrt(1+1/n-(1:n)/n)*(1-BY)^(3/2)-8/3*sqrt(1-(1:n)/n)*(1-BY)^(3/2))
              }
Z=sort(tmp1)
return(c(as.integer(tmp>Z[as.integer(0.9*BN)]),as.integer(tmp>Z[as.integer(0.95*BN)])))
}

Jager_Gamma=function(sample,theta,BN){
n=length(sample);
X=pgamma(sample,shape=theta[1],scale=theta[2]);
Y=sort(X);
tmp=sum(-4/3-8/3*sqrt(((1:n)-1)/n)*Y^(3/2)+8/3*sqrt((1:n)/n)*Y^(3/2)
+8/3*sqrt(1+1/n-(1:n)/n)*(1-Y)^(3/2)-8/3*sqrt(1-(1:n)/n)*(1-Y)^(3/2))
tmp1=0*(1:BN)
for(j in 1:BN){
Bsample=rgamma(n,shape=theta[1],scale=theta[2])
BX=pgamma(Bsample,shape=theta[1],scale=theta[2])
BY=sort(BX)
tmp1[j]=sum(-4/3-8/3*sqrt(((1:n)-1)/n)*BY^(3/2)+8/3*sqrt((1:n)/n)*BY^(3/2)
+8/3*sqrt(1+1/n-(1:n)/n)*(1-BY)^(3/2)-8/3*sqrt(1-(1:n)/n)*(1-BY)^(3/2))
               }
Z=sort(tmp1)
return(c(as.integer(tmp>Z[as.integer(0.9*BN)]),as.integer(tmp>Z[as.integer(0.95*BN)])))
}
Jager_chisq=function(sample,beta,BN){
n=length(sample)
X=pchisq(sample,df=beta)
Y=sort(X)
tmp=sum(-4/3-8/3*sqrt(((1:n)-1)/n)*Y^(3/2)+8/3*sqrt((1:n)/n)*Y^(3/2)
+8/3*sqrt(1+1/n-(1:n)/n)*(1-Y)^(3/2)-8/3*sqrt(1-(1:n)/n)*(1-Y)^(3/2))
tmp1=0*(1:BN)
for(j in 1:BN){
Bsample=rchisq(n,beta)
BX=pchisq(Bsample,beta)
BY=sort(BX)
tmp1[j]=sum(-4/3-8/3*sqrt(((1:n)-1)/n)*BY^(3/2)+8/3*sqrt((1:n)/n)*BY^(3/2)
+8/3*sqrt(1+1/n-(1:n)/n)*(1-BY)^(3/2)-8/3*sqrt(1-(1:n)/n)*(1-BY)^(3/2))
              }
Z=sort(tmp1)
return(c(as.integer(tmp>Z[as.integer(0.9*BN)]),as.integer(tmp>Z[as.integer(0.95*BN)])))
}

##################
#composite
Einmahl_Norm_Est=function(sample,BN){
n=length(sample)
```

```
theta1=mean(sample)
theta2=(mean((sample-theta1)^2))^0.5
X=pnorm(sample,mean=theta1,sd=theta2)
X=pmax(X,0.000000001)
X=pmin(X,0.999999999)
Y=sort(X)
tmp=(2*n*((1-Y[1])*log(1-Y[1])+Y[1]+1+Y[n]*log(Y[n])-Y[n])
    +sum(-2*(1:(n-1))*log(n/(1:(n-1)))*(Y[2:n]-Y[1:(n-1)])
    -2*(1:(n-1))*(Y[2:n]*log(Y[2:n])-Y[2:n]-Y[1:(n-1)]*log(Y[1:(n-1)])
    +Y[1:(n-1)])+2*(n-(1:(n-1)))*log(1-(1:(n-1))/n)*(Y[2:n]-Y[1:(n-1)])
    +2*(n-(1:(n-1)))*((1-Y[2:n])*log(1-Y[2:n])+Y[2:n]-Y[1:(n-1)]
    -(1-Y[1:(n-1)])*log(1-Y[1:(n-1)]))))
tmp1=0*(1:BN)
for(j in 1:BN){
Bsample=rnorm(n,mean=theta1,sd=theta2)
Btheta1=mean(Bsample)
Btheta2=(mean((Bsample-Btheta1)^2))^0.5
BX=pnorm(Bsample,mean=Btheta1,sd=Btheta2)
BY=sort(BX)
tmp1[j]=(2*n*((1-BY[1])*log(1-BY[1])+BY[1]+1+BY[n]*log(BY[n])-BY[n])
        +sum(-2*(1:(n-1))*log(n/(1:(n-1)))*(BY[2:n]-BY[1:(n-1)])
        -2*(1:(n-1))*(BY[2:n]*log(BY[2:n])-BY[2:n]
        -BY[1:(n-1)]*log(BY[1:(n-1)])+BY[1:(n-1)])
        +2*(n-(1:(n-1)))*log(1-(1:(n-1))/n)*(BY[2:n]-BY[1:(n-1)])
        +2*(n-(1:(n-1)))*((1-BY[2:n])*log(1-BY[2:n])+BY[2:n]-BY[1:(n-1)]
        -(1-BY[1:(n-1)])*log(1-BY[1:(n-1)]))))

}
Z=sort(tmp1)
return(c(as.integer(tmp>Z[as.integer(0.9*BN)]),as.integer(tmp>Z[as.integer(0.95*BN)])))
}

Einmahl_T_Est=function(sample,BN){
n=length(sample)
sigma2=mean(sample^2)
v=max(2*sigma2/(sigma2-1),2.001)
X=pt(sample,df=v)
Y=sort(X)
tmp=(2*n*((1-Y[1])*log(1-Y[1])+Y[1]+1+Y[n]*log(Y[n])-Y[n])
    +sum(-2*(1:(n-1))*log(n/(1:(n-1)))*(Y[2:n]-Y[1:(n-1)])
    -2*(1:(n-1))*(Y[2:n]*log(Y[2:n])-Y[2:n]-Y[1:(n-1)]*log(Y[1:(n-1)])
    +Y[1:(n-1)])+2*(n-(1:(n-1)))*log(1-(1:(n-1))/n)*(Y[2:n]-Y[1:(n-1)])
    +2*(n-(1:(n-1)))*((1-Y[2:n])*log(1-Y[2:n])+Y[2:n]-Y[1:(n-1)]
    -(1-Y[1:(n-1)])*log(1-Y[1:(n-1)]))))

tmp1=0*(1:BN)
for(j in 1:BN){
Bsample=rt(n,df=v)
Bsigma2=mean(Bsample^2)
```

```
Bv=max(2*Bsigma2/(Bsigma2-1),2.001)
BX=pt(Bsample,df=Bv)
BY=sort(BX)
tmp1[j]=(2*n*((1-BY[1])*log(1-BY[1])+BY[1]+1+BY[n]*log(BY[n])-BY[n])
        +sum(-2*(1:(n-1))*log(n/(1:(n-1)))*(BY[2:n]-BY[1:(n-1)])
        -2*(1:(n-1))*(BY[2:n]*log(BY[2:n])-BY[2:n]
        -BY[1:(n-1)]*log(BY[1:(n-1)])+BY[1:(n-1)])
        +2*(n-(1:(n-1)))*log(1-(1:(n-1))/n)*(BY[2:n]-BY[1:(n-1)])
        +2*(n-(1:(n-1)))*((1-BY[2:n])*log(1-BY[2:n])+BY[2:n]-BY[1:(n-1)]
        -(1-BY[1:(n-1)])*log(1-BY[1:(n-1)]))))

}
Z=sort(tmp1)
return(c(as.integer(tmp>Z[as.integer(0.9*BN)]),as.integer(tmp>Z[as.integer(0.95*BN)])))}

Einmahl_Gamma_Est=function(sample,BN){
n=length(sample)
theta1=max(mean(sample),0.0000001)
theta2=var(sample)
theta2=theta2/theta1
theta1=theta1/theta2
X=pgamma(sample,shape=theta1,scale=theta2)
X=pmax(X,0.000000001)
X=pmin(X,0.999999999)
Y=sort(X)
tmp=(2*n*((1-Y[1])*log(1-Y[1])+Y[1]+1+Y[n]*log(Y[n])-Y[n])
    +sum(-2*(1:(n-1))*log(n/(1:(n-1)))*(Y[2:n]-Y[1:(n-1)])
    -2*(1:(n-1))*(Y[2:n]*log(Y[2:n]+as.integer(Y[2:n]==0))
    -Y[2:n]-Y[1:(n-1)]*log(Y[1:(n-1)]+as.integer(Y[1:(n-1)]==0))
    +Y[1:(n-1)])+2*(n-(1:(n-1)))*log(1-(1:(n-1))/n)*(Y[2:n]-Y[1:(n-1)])
    +2*(n-(1:(n-1)))*((1-Y[2:n])*log(1-Y[2:n]+as.integer((1-Y[2:n])==0))
    +Y[2:n]-Y[1:(n-1)]
    -(1-Y[1:(n-1)])*log(1-Y[1:(n-1)]+as.integer((1-Y[1:(n-1)])==0)))))

tmp1=0*(1:BN)
for(j in 1:BN){
Bsample=rgamma(n,shape=theta1,scale=theta2)
Btheta1=max(mean(Bsample),0.00000001)
Btheta2=var(Bsample)
Btheta2=Btheta2/Btheta1
Btheta1=Btheta1/Btheta2
BX=pgamma(Bsample,shape=Btheta1,scale=Btheta2)
BX=pmax(BX,0.00000001)
BX=pmin(BX,0.99999999)
BY=sort(BX)
tmp1[j]=(2*n*((1-BY[1])*log(1-BY[1])+BY[1]+1+BY[n]*log(BY[n])-BY[n])
    +sum(-2*(1:(n-1))*log(n/(1:(n-1)))*(BY[2:n]-BY[1:(n-1)])
    -2*(1:(n-1))*(BY[2:n]*log(BY[2:n]+as.integer((BY[2:n])==0))
    -BY[2:n]-BY[1:(n-1)]*log(BY[1:(n-1)]+as.integer((BY[1:(n-1)])==0))
```

```
    +BY[1:(n-1)])+2*(n-(1:(n-1)))*log(1-(1:(n-1))/n)*(BY[2:n]-BY[1:(n-1)])
    +2*(n-(1:(n-1)))*((1-BY[2:n])*log(1-BY[2:n]+as.integer((1-BY[2:n])==0))
    +BY[2:n]-BY[1:(n-1)]
    -(1-BY[1:(n-1)])*log(1-BY[1:(n-1)]+as.integer((1-BY[1:(n-1)])==0)))))
                }
Z=sort(tmp1)
return(c(as.integer(tmp>Z[as.integer(0.9*BN)]),as.integer(tmp>Z[as.integer(0.95*BN)])))
}

Einmahl_chisq_Est=function(sample,BN){
n=length(sample)
v=mean(sample)
X=pchisq(sample,v)
Y=sort(X)
tmp=(2*n*((1-Y[1])*log(1-Y[1])+Y[1]+1+Y[n]*log(Y[n])-Y[n])
    +sum(-2*(1:(n-1))*log(n/(1:(n-1)))*(Y[2:n]-Y[1:(n-1)])
    -2*(1:(n-1))*(Y[2:n]*log(Y[2:n]+as.integer((Y[2:n])==0))
    -Y[2:n]-Y[1:(n-1)]*log(Y[1:(n-1)]+as.integer((Y[1:(n-1)])==0))
    +Y[1:(n-1)])+2*(n-(1:(n-1)))*log(1-(1:(n-1))/n)*(Y[2:n]-Y[1:(n-1)])
    +2*(n-(1:(n-1)))*((1-Y[2:n])*log(1-Y[2:n]+as.integer((1-Y[2:n])==0))
    +Y[2:n]-Y[1:(n-1)]
    -(1-Y[1:(n-1)])*log(1-Y[1:(n-1)]+as.integer((1-Y[1:(n-1)])==0)))))

tmp1=0*(1:BN)
for(j in 1:BN){
Bsample=rchisq(n,v)
Bv=mean(Bsample)
BX=pchisq(Bsample,Bv)
BY=sort(BX)
tmp1[j]=(2*n*((1-BY[1])*log(1-BY[1])+BY[1]+1+BY[n]*log(BY[n])-BY[n])
    +sum(-2*(1:(n-1))*log(n/(1:(n-1)))*(BY[2:n]-BY[1:(n-1)])
    -2*(1:(n-1))*(BY[2:n]*log(BY[2:n]+as.integer((BY[2:n])==0))
    -BY[2:n]-BY[1:(n-1)]*log(BY[1:(n-1)]+as.integer((BY[1:(n-1)])==0))
    +BY[1:(n-1)])+2*(n-(1:(n-1)))*log(1-(1:(n-1))/n)*(BY[2:n]-BY[1:(n-1)])
    +2*(n-(1:(n-1)))*((1-BY[2:n])*log(1-BY[2:n]+as.integer((1-BY[2:n])==0))
    +BY[2:n]-BY[1:(n-1)]
    -(1-BY[1:(n-1)])*log(1-BY[1:(n-1)]+as.integer((1-BY[1:(n-1)])==0)))))
                }
Z=sort(tmp1)
return(c(as.integer(tmp>Z[as.integer(0.9*BN)]),as.integer(tmp>Z[as.integer(0.95*BN)])))
}

#################
Jager_Norm_Est=function(sample,BN){
n=length(sample)
theta1=mean(sample)
theta2=(mean((sample-theta1)^2))^0.5
X=pnorm(sample,mean=theta1,sd=theta2)
Y=sort(X)
```

```
tmp=sum(-4/3-8/3*sqrt(((1:n)-1)/n)*Y^(3/2)+8/3*sqrt((1:n)/n)*Y^(3/2)
+8/3*sqrt(1+1/n-(1:n)/n)*(1-Y)^(3/2)-8/3*sqrt(1-(1:n)/n)*(1-Y)^(3/2))
tmp1=0*(1:BN)
for(j in 1:BN){
Bsample=rnorm(n,mean=theta1,sd=theta2)
Btheta1=mean(Bsample)
Btheta2=(mean((Bsample-Btheta1)^2))^0.5
BX=pnorm(Bsample,mean=Btheta1,sd=Btheta2)
BY=sort(BX)
tmp1[j]=sum(-4/3-8/3*sqrt(((1:n)-1)/n)*BY^(3/2)+8/3*sqrt((1:n)/n)*BY^(3/2)
+8/3*sqrt(1+1/n-(1:n)/n)*(1-BY)^(3/2)-8/3*sqrt(1-(1:n)/n)*(1-BY)^(3/2))
}
Z=sort(tmp1)
return(c(as.integer(tmp>Z[as.integer(0.9*BN)]),as.integer(tmp>Z[as.integer(0.95*BN)])))
}

Jager_T_Est=function(sample,BN){
n=length(sample)
sigma2=mean(sample^2)
v=max(2*sigma2/(sigma2-1),2.001)
X=pt(sample,df=v)
Y=sort(X)
tmp=tmp=sum(-4/3-8/3*sqrt(((1:n)-1)/n)*Y^(3/2)+8/3*sqrt((1:n)/n)*Y^(3/2)
+8/3*sqrt(1+1/n-(1:n)/n)*(1-Y)^(3/2)-8/3*sqrt(1-(1:n)/n)*(1-Y)^(3/2))
tmp1=0*(1:BN)
for(j in 1:BN){
Bsample=rt(n,df=v)
Bsigma2=mean(Bsample^2)
Bv=max(2*Bsigma2/(Bsigma2-1),2.001)
BX=pt(Bsample,df=Bv)
BY=sort(BX)
tmp1[j]=tmp=sum(-4/3-8/3*sqrt(((1:n)-1)/n)*BY^(3/2)+8/3*sqrt((1:n)/n)*BY^(3/2)
+8/3*sqrt(1+1/n-(1:n)/n)*(1-BY)^(3/2)-8/3*sqrt(1-(1:n)/n)*(1-BY)^(3/2))
}
Z=sort(tmp1)
return(c(as.integer(tmp>Z[as.integer(0.9*BN)]),as.integer(tmp>Z[as.integer(0.95*BN)])))}

Jager_Gamma_Est=function(sample,BN){
n=length(sample);
theta1=max(mean(sample),0.0000001)
theta2=var(sample)
theta2=theta2/theta1
theta1=theta1/theta2
X=pgamma(sample,shape=theta1,scale=theta2)
Y=sort(X);
tmp=sum(-4/3-8/3*sqrt(((1:n)-1)/n)*Y^(3/2)+8/3*sqrt((1:n)/n)*Y^(3/2)
+8/3*sqrt(1+1/n-(1:n)/n)*(1-Y)^(3/2)-8/3*sqrt(1-(1:n)/n)*(1-Y)^(3/2))
tmp1=0*(1:BN)
for(j in 1:BN){
```

```
Bsample=rgamma(n,shape=theta1,scale=theta2)
Btheta1=max(mean(Bsample),0.0000001)
Btheta2=var(Bsample)
Btheta2=Btheta2/Btheta1
Btheta1=Btheta1/Btheta2
BX=pgamma(Bsample,shape=Btheta1,scale=Btheta2)
BY=sort(BX)
tmp1[j]=sum(-4/3-8/3*sqrt(((1:n)-1)/n)*BY^(3/2)+8/3*sqrt((1:n)/n)*BY^(3/2)
+8/3*sqrt(1+1/n-(1:n)/n)*(1-BY)^(3/2)-8/3*sqrt(1-(1:n)/n)*(1-BY)^(3/2))
            }
Z=sort(tmp1)
return(c(as.integer(tmp>Z[as.integer(0.9*BN)]),as.integer(tmp>Z[as.integer(0.95*BN)])))
}

Jager_chisq_Est=function(sample,BN){
n=length(sample)
v=mean(sample)
X=pchisq(sample,df=v)
Y=sort(X)
tmp=sum(-4/3-8/3*sqrt(((1:n)-1)/n)*Y^(3/2)+8/3*sqrt((1:n)/n)*Y^(3/2)
+8/3*sqrt(1+1/n-(1:n)/n)*(1-Y)^(3/2)-8/3*sqrt(1-(1:n)/n)*(1-Y)^(3/2))

tmp1=0*(1:BN)
for(j in 1:BN){
Bsample=rchisq(n,v)
Bv=mean(Bsample)
BX=pchisq(Bsample,Bv)
BY=sort(BX)
tmp1[j]=sum(-4/3-8/3*sqrt(((1:n)-1)/n)*BY^(3/2)+8/3*sqrt((1:n)/n)*BY^(3/2)
+8/3*sqrt(1+1/n-(1:n)/n)*(1-BY)^(3/2)-8/3*sqrt(1-(1:n)/n)*(1-BY)^(3/2))
               }
Z=sort(tmp1)
return(c(as.integer(tmp>Z[as.integer(0.9*BN)]),as.integer(tmp>Z[as.integer(0.95*BN)])))
}

#################
NT=function(r,n,delta,beta,theta){
res1=0*matrix(1:(4*r),ncol=4)
res2=0*matrix(1:(4*r),ncol=4)
res3=0*matrix(1:(4*r),ncol=4)
res4=0*matrix(1:(4*r),ncol=4)

for( i in 1:r){
U=runif(n)
X=(U<=1-delta/n^0.5)*rnorm(n,mean=theta[1],sd=theta[2])+(U>1-delta/n^0.5)*rt(n,df=beta)
res1[i,1:2]=CS_Norm(X,theta)
res1[i,3:4]=res1[i,1:2]
res2[i,]=JEL_CM1_Norm(X,theta)
res3[i,3:4]=res3[i,1:2]=Einmahl_Norm(X,theta,BN)
```

```
res4[i,3:4]=res4[i,1:2]=Jager_Norm(X,theta,BN)}
res=rbind(c(mean(res1[,1]),mean(res1[,2]),mean(res1[,3]),mean(res1[,4])),
          c(mean(res2[,1]),mean(res2[,2]),mean(res2[,3]),mean(res2[,4])),
          c(mean(res3[,1]),mean(res3[,2]),mean(res3[,3]),mean(res3[,4])),
          c(mean(res4[,1]),mean(res4[,2]),mean(res4[,3]),mean(res4[,4])))
return(res)
}

NN=function(r,n,delta,beta,theta){
res1=0*matrix(1:(4*r),ncol=4)
res2=0*matrix(1:(4*r),ncol=4)
res3=0*matrix(1:(4*r),ncol=4)
res4=0*matrix(1:(4*r),ncol=4)

for( i in 1:r){
U=runif(n)
X=(U<=1-delta/n^0.5)*rnorm(n,mean=theta[1],sd=theta[2])+(U>1-delta/n^0.5)*rnorm(n,mean=beta[1],sd=beta[2])
 res1[i,1:2]=CS_Norm(X,theta)
res1[i,3:4]=res1[i,1:2]
res2[i,]=JEL_CM1_Norm(X,theta)
res3[i,3:4]=res3[i,1:2]=Einmahl_Norm(X,theta)
res4[i,3:4]=res4[i,1:2]=Jager_Norm(X,theta)
}
res=rbind(c(mean(res1[,1]),mean(res1[,2]),mean(res1[,3]),mean(res1[,4])),
          c(mean(res2[,1]),mean(res2[,2]),mean(res2[,3]),mean(res2[,4])),
          c(mean(res3[,1]),mean(res3[,2]),mean(res3[,3]),mean(res3[,4])),
          c(mean(res4[,1]),mean(res4[,2]),mean(res4[,3]),mean(res4[,4])))
return(res)
}

TN=function(r,n,delta,beta,theta){
res1=0*matrix(1:(4*r),ncol=4)
res2=0*matrix(1:(4*r),ncol=4)
res3=0*matrix(1:(4*r),ncol=4)
res4=0*matrix(1:(4*r),ncol=4)

for( i in 1:r){
U=runif(n)
X=(U<=1-delta/n^0.5)*rt(n,df=theta[1])+(U>1-delta/n^0.5)*rnorm(n,mean=beta[1],sd=beta[2])
 res1[i,1:2]=CS_T(X,theta)
res1[i,3:4]=res1[i,1:2]
res2[i,]=JEL_CM1_T(X,theta)
res3[i,3:4]=res3[i,1:2]=Einmahl_T(X,theta)
res4[i,3:4]=res4[i,1:2]=Jager_T(X,theta)
}
res=rbind(c(mean(res1[,1]),mean(res1[,2]),mean(res1[,3]),mean(res1[,4])),
          c(mean(res2[,1]),mean(res2[,2]),mean(res2[,3]),mean(res2[,4])),
          c(mean(res3[,1]),mean(res3[,2]),mean(res3[,3]),mean(res3[,4])),
          c(mean(res4[,1]),mean(res4[,2]),mean(res4[,3]),mean(res4[,4])))
```

```
return(res)
}

TT=function(r,n,delta,beta,theta){
res1=0*matrix(1:(4*r),ncol=4)
res2=0*matrix(1:(4*r),ncol=4)
res3=0*matrix(1:(4*r),ncol=4)
res4=0*matrix(1:(4*r),ncol=4)

for( i in 1:r){
U=runif(n)
X=(U<=1-delta/n^0.5)*rt(n,df=theta[1])+(U>1-delta/n^0.5)*rt(n,df=beta[1])
 res1[i,1:2]=CS_T(X,theta)
res1[i,3:4]=res1[i,1:2]
res2[i,]=JEL_CM1_T(X,theta)
res3[i,3:4]=res3[i,1:2]=Einmahl_T(X,theta)
res4[i,3:4]=res4[i,1:2]=Jager_T(X,theta)
}
res=rbind(c(mean(res1[,1]),mean(res1[,2]),mean(res1[,3]),mean(res1[,4])),
          c(mean(res2[,1]),mean(res2[,2]),mean(res2[,3]),mean(res2[,4])),
          c(mean(res3[,1]),mean(res3[,2]),mean(res3[,3]),mean(res3[,4])),
          c(mean(res4[,1]),mean(res4[,2]),mean(res4[,3]),mean(res4[,4])))
return(res)}

GC=function(r,n,delta,beta,theta,BN){
res1=0*matrix(1:(4*r),ncol=4)
res2=0*matrix(1:(4*r),ncol=4)
res3=0*matrix(1:(4*r),ncol=4)
res4=0*matrix(1:(4*r),ncol=4)

for( i in 1:r){
U=runif(n)
X=(U<=1-delta/n^0.5)*rgamma(n,shape=theta[1],scale=theta[2])+(U>1-delta/n^0.5)*rchisq(n,df=beta)
res1[i,1:2]=CS_Gamma(X,theta)
res1[i,3:4]=res1[i,1:2]
res2[i,]=JEL_CM1_Gamma(X,theta)
res3[i,3:4]=res3[i,1:2]=Einmahl_Gamma(X,theta,BN)
res4[i,3:4]=res4[i,1:2]=Jager_Gamma(X,theta,BN)

}
res=rbind(c(mean(res1[,1]),mean(res1[,2]),mean(res1[,3]),mean(res1[,4])),
          c(mean(res2[,1]),mean(res2[,2]),mean(res2[,3]),mean(res2[,4])),
          c(mean(res3[,1]),mean(res3[,2]),mean(res3[,3]),mean(res3[,4])),
          c(mean(res4[,1]),mean(res4[,2]),mean(res4[,3]),mean(res4[,4])))
return(res)}

GG=function(r,n,delta,beta,theta){
res1=0*matrix(1:(4*r),ncol=4)
res2=0*matrix(1:(4*r),ncol=4)
```

```
res3=0*matrix(1:(4*r),ncol=4)
res4=0*matrix(1:(4*r),ncol=4)

for( i in 1:r){
U=runif(n)
X=(U<=1-delta/n^0.5)*rgamma(n,shape=theta[1],scale=theta[2])+(U>1-delta/n^0.5)*rgamma(n,shape=beta[1],scale=beta[2])
res1[i,1:2]=CS_Gamma(X,theta)
res1[i,3:4]=res1[i,1:2]
res2[i,]=JEL_CM1_Gamma(X,theta)
res3[i,3:4]=res3[i,1:2]=Einmahl_Gamma(X,theta)
res4[i,3:4]=res4[i,1:2]=Jager_Gamma(X,theta)}
res=rbind(c(mean(res1[,1]),mean(res1[,2]),mean(res1[,3]),mean(res1[,4])),
          c(mean(res2[,1]),mean(res2[,2]),mean(res2[,3]),mean(res2[,4])),
          c(mean(res3[,1]),mean(res3[,2]),mean(res3[,3]),mean(res3[,4])),
          c(mean(res4[,1]),mean(res4[,2]),mean(res4[,3]),mean(res4[,4])))
return(res)
}

CG=function(r,n,delta,beta,theta){
res1=0*matrix(1:(4*r),ncol=4)
res2=0*matrix(1:(4*r),ncol=4)
res3=0*matrix(1:(4*r),ncol=4)
res4=0*matrix(1:(4*r),ncol=4)

for( i in 1:r){
U=runif(n)
X=(U<=1-delta/n^0.5)*rchisq(n,df=theta[1])+(U>1-delta/n^0.5)*rgamma(n,shape=beta[1],scale=beta[2])
res1[i,1:2]=CS_chisq(X,theta)
res1[i,3:4]=res1[i,1:2]
res2[i,]=JEL_CM1_chisq(X,theta)
res3[i,3:4]=res3[i,1:2]=Einmahl_chisq(X,theta)
res4[i,3:4]=res4[i,1:2]=Jager_chisq(X,theta)
}
res=rbind(c(mean(res1[,1]),mean(res1[,2]),mean(res1[,3]),mean(res1[,4])),
          c(mean(res2[,1]),mean(res2[,2]),mean(res2[,3]),mean(res2[,4])),
          c(mean(res3[,1]),mean(res3[,2]),mean(res3[,3]),mean(res3[,4])),
          c(mean(res4[,1]),mean(res4[,2]),mean(res4[,3]),mean(res4[,4])))
return(res)
}

CC=function(r,n,delta,beta,theta){
res1=0*matrix(1:(4*r),ncol=4)
res2=0*matrix(1:(4*r),ncol=4)
res3=0*matrix(1:(4*r),ncol=4)
res4=0*matrix(1:(4*r),ncol=4)

for( i in 1:r){
U=runif(n)
X=(U<=1-delta/n^0.5)*rchisq(n,df=theta[1])+(U>1-delta/n^0.5)*rchisq(n,df=beta[1])
```

```
res1[i,1:2]=CS_chisq(X,theta)
res1[i,3:4]=res1[i,1:2]
res2[i,]=JEL_CM1_chisq(X,theta)
res3[i,3:4]=res3[i,1:2]=Einmahl_chisq(X,theta)
res4[i,3:4]=res4[i,1:2]=Jager_chisq(X,theta)
}
res=rbind(c(mean(res1[,1]),mean(res1[,2]),mean(res1[,3]),mean(res1[,4])),
          c(mean(res2[,1]),mean(res2[,2]),mean(res2[,3]),mean(res2[,4])),
          c(mean(res3[,1]),mean(res3[,2]),mean(res3[,3]),mean(res3[,4])),
          c(mean(res4[,1]),mean(res4[,2]),mean(res4[,3]),mean(res4[,4])))
return(res)
}

############
NT_Est=function(r,n,delta,beta,theta,BN){
res1=0*matrix(1:(4*r),ncol=4)
res2=0*matrix(1:(4*r),ncol=4)
res3=0*matrix(1:(4*r),ncol=4)
res4=0*matrix(1:(4*r),ncol=4)

for( i in 1:r){
U=runif(n)
X=(U<=1-delta/n^0.5)*rnorm(n,mean=theta[1],sd=theta[2])+(U>1-delta/n^0.5)*rt(n,df=beta)
res1[i,1:2]=CS_Norm_Est(X,BN)
res1[i,3:4]=res1[i,1:2]
res2[i,]=JEL_CM1_Norm_Est(X)
res3[i,3:4]=res3[i,1:2]=Einmahl_Norm_Est(X,BN)
res4[i,3:4]=res4[i,1:2]=Jager_Norm_Est(X,BN)

}
res=rbind(c(mean(res1[,1]),mean(res1[,2]),mean(res1[,3]),mean(res1[,4])),
          c(mean(res2[,1]),mean(res2[,2]),mean(res2[,3]),mean(res2[,4])),
          c(mean(res3[,1]),mean(res3[,2]),mean(res3[,3]),mean(res3[,4])),
          c(mean(res4[,1]),mean(res4[,2]),mean(res4[,3]),mean(res4[,4])))
return(res)
}
TN_Est=function(r,n,delta,beta,theta,BN){
res1=0*matrix(1:(4*r),ncol=4)
res2=0*matrix(1:(4*r),ncol=4)
res3=0*matrix(1:(4*r),ncol=4)
res4=0*matrix(1:(4*r),ncol=4)

for( i in 1:r){
U=runif(n)
X=(U<=1-delta/n^0.5)*rt(n,df=theta[1])+(U>1-delta/n^0.5)*rnorm(n,mean=beta[1],sd=beta[2])
 res1[i,1:2]=CS_T_Est(X,BN)
res1[i,3:4]=res1[i,1:2]
res2[i,]=JEL_CM1_T_Est(X)
res3[i,3:4]=res3[i,1:2]=Einmahl_T_Est(X,BN)
```

```
res4[i,3:4]=res4[i,1:2]=Jager_T_Est(X,BN)
}
res=rbind(c(mean(res1[,1]),mean(res1[,2]),mean(res1[,3]),mean(res1[,4])),
          c(mean(res2[,1]),mean(res2[,2]),mean(res2[,3]),mean(res2[,4])),
          c(mean(res3[,1]),mean(res3[,2]),mean(res3[,3]),mean(res3[,4])),
          c(mean(res4[,1]),mean(res4[,2]),mean(res4[,3]),mean(res4[,4])))
return(res)
}

GC_Est=function(r,n,delta,beta,theta,BN){
res1=0*matrix(1:(4*r),ncol=4)
res2=0*matrix(1:(4*r),ncol=4)
res3=0*matrix(1:(4*r),ncol=4)
res4=0*matrix(1:(4*r),ncol=4)

for( i in 1:r){
U=runif(n)
X=(U<=1-delta/n^0.5)*rgamma(n,shape=theta[1],scale=theta[2])+(U>1-delta/n^0.5)*rchisq(n,df=beta)
res1[i,1:2]=CS_Gamma_Est(X,BN)
res1[i,3:4]=res1[i,1:2]
res2[i,]=JEL_CM1_Gamma_Est(X)
res3[i,3:4]=res3[i,1:2]=Einmahl_Gamma_Est(X,BN)
res4[i,3:4]=res4[i,1:2]=Jager_Gamma_Est(X,BN)
}
res=rbind(c(mean(res1[,1]),mean(res1[,2]),mean(res1[,3]),mean(res1[,4])),
          c(mean(res2[,1]),mean(res2[,2]),mean(res2[,3]),mean(res2[,4])),
          c(mean(res3[,1]),mean(res3[,2]),mean(res3[,3]),mean(res3[,4])),
          c(mean(res4[,1]),mean(res4[,2]),mean(res4[,3]),mean(res4[,4])))
return(res)
}

CG_Est=function(r,n,delta,beta,theta,BN){
res1=0*matrix(1:(4*r),ncol=4)
res2=0*matrix(1:(4*r),ncol=4)
res3=0*matrix(1:(4*r),ncol=4)
res4=0*matrix(1:(4*r),ncol=4)

for( i in 1:r){
U=runif(n)
X=(U<=1-delta/n^0.5)*rchisq(n,df=theta[1])+(U>1-delta/n^0.5)*rgamma(n,shape=beta[1],scale=beta[2])
res1[i,1:2]=CS_chisq_Est(X,BN)
res1[i,3:4]=res1[i,1:2]
res2[i,]=JEL_CM1_chisq_Est(X)
res3[i,3:4]=res3[i,1:2]=Einmahl_chisq_Est(X,BN)
res4[i,3:4]=res4[i,1:2]=Jager_chisq_Est(X,BN)
}
res=rbind(c(mean(res1[,1]),mean(res1[,2]),mean(res1[,3]),mean(res1[,4])),
          c(mean(res2[,1]),mean(res2[,2]),mean(res2[,3]),mean(res2[,4])),
          c(mean(res3[,1]),mean(res3[,2]),mean(res3[,3]),mean(res3[,4])),
```

```
        c(mean(res4[,1]),mean(res4[,2]),mean(res4[,3]),mean(res4[,4])))
return(res)
}
#################
#set.seed(__MQSUB1__)
r=50
n=200
BN=1000
truedeltaN=1
truethetaN=c(-1,0.2)
truebetaT=c(1)
#myname=paste("NT_n",n,"_delta","truedeltaN","_theta",truethetaN[1],truethetaN[2],"_beta",truebetaT,"_",__MQSUB2__,".Rdata",sep="")
resN=NT(r,n,0,truebetaT,truethetaN)
resA=NT(r,n,1,truebetaT,truethetaN)
resN_Est=NT_Est(r,n,0,truebetaT,truethetaN,BN)
resA_Est=NT_Est(r,n,1,truebetaT,truethetaN,BN)
#truedeltaG=0
#truethetaG=c(5,2)
#truebetaC=c(3)
#myname=paste("GC_n",n,"_delta","truedeltaG","_theta",truethetaG[1],truethetaG[2],"_beta",truebetaC,"_",__MQSUB2__,".Rdata",sep="")
#resN=GC(r,n,0,truebetaC,truethetaG,BN)
#resA=GC(r,n,1,truebetaC,truethetaG,BN)
#resN_Est=GC_Est(r,n,0,truebetaC,truethetaG,BN)
#resA_Est=GC_Est(r,n,1,truebetaC,truethetaG,BN)
#save.image(file="myname")
    }
```

5.2 Codes for Chapter III

```
{
library(emplik)

#Test Normal(0, beta=sigma^2) against mixture with t(nu)#
ff_normal_Composite=function(n,alpha,mu,sigma,nu,delta)
{
x=2*(runif(n)+1)
m=function(x,alpha) exp(alpha*x)
e=(1-delta/sqrt(n))*rnorm(n,mu,sigma)+delta/sqrt(n)*rt(n,nu)
z=function(y,x,alpha) y-m(x,alpha)
y=exp(alpha*x)+e
k=n/2
zalpha1star=matrix(rep(0,k^2),k,k)
zalpha2star=matrix(rep(0,k^2),k,k)
zalpha0star=matrix(rep(0,k^2),k,k)
G1=numeric(k)
G2=numeric(k)
G3=numeric(k)
```

```
G4=numeric(k)
f11=matrix(rep(0,k^2),k,k)
f22=matrix(rep(0,k^2),k,k)
f00=matrix(rep(0,k^2),k,k)
x1=x[1:k]
x2=x[(k+1):(2*k)]
y1=y[1:k]
y2=y[(k+1):(2*k)]
h0=function(v) mean((z(y,x,v))^2)
tmp=try(optimize(h0,interval=c(-1,1)),silent=TRUE)
if(length(tmp)>1){
alpha_hat=tmp$minimum
}
else{
alpha_hat=0
}
beta_hat=mean((y-exp(alpha_hat*x))^2)
alpha_hathat=alpha_hat+1/sum(x^2*y*exp(alpha_hat*x)-2*x^2*exp(2*alpha_hat*x))*(x1*y1*exp(alpha_hat*x1)
             -x1*exp(2*alpha_hat*x1)+(x2*y2*exp(alpha_hat*x2)-x2*exp(2*alpha_hat*x2)))
#beta_hathat=beta_hat+1/sum(-(y-exp(alpha_hat*x))^2/(beta_hat^3)
             +1/2*beta_hat^(-2))*((y1-exp(alpha_hat*x1))^2/(2*beta_hat^2)
             +(y2-exp(alpha_hat*x2))^2/(2*beta_hat^2)-1/beta_hat)+1/sum(-(y-exp(alpha_hat*x))^2/(beta_hat^3)
             +1/2*beta_hat^(-2))*sum(x*exp(alpha_hat*x)*(y-exp(alpha_hat*x))/(beta_hat^2))*(alpha_hathat-alpha_hat)

beta_hathat=beta_hat-((y1-exp(alpha_hat*x1))^2+(y2-exp(alpha_hat*x2))^2-2*beta_hat)/2/k
            +mean(-2*(y1-exp(alpha_hat*x1))*x1*exp(alpha_hat*x1)
            -2*(y2-exp(alpha_hat*x2))*x2*exp(alpha_hat*x2))/2*(alpha_hathat-alpha_hat)
zalpha1=z(y1,x1,alpha_hat)
zalpha2=z(y2,x2,alpha_hat)
f1=pnorm(zalpha1,mean=0,sd=sqrt(max(beta_hat,0.001)))
f2=pnorm(zalpha2,mean=0,sd=sqrt(max(beta_hat,0.001)))
f0=pmax(f1,f2)
for(j in 1:k)
{
zalpha1star[,j]=z(y1,x1,alpha_hathat[j])
zalpha2star[,j]=z(y2,x2,alpha_hathat[j])
f11[,j]=pnorm(zalpha1star[,j],mean=0,sd=sqrt(max(beta_hathat[j],0.001)))
f22[,j]=pnorm(zalpha2star[,j],mean=0,sd=sqrt(max(beta_hathat[j],0.001)))
}
f00=pmax(f11,f22)
for(j in 1:k)
{
#G1[j]=sum(2/3-f0)-sum(2/3-f00[,j])+2/3-f00[j,j]
#G2[j]=sum(2/3-f1^2-f2^2)-sum(2/3-f11[,j]^2-f22[,j]^2)+2/3-f11[j,j]^2-f22[j,j]^2
#G3[j]=sum(f1+f2-1)-sum(f11[,j]+f22[,j]-1)+f11[j,j]+f22[j,j]-1
#G4[j]=sum(sqrt(1-f1^2)+sqrt(1-f2^2)-pi/2)-sum(sqrt(1-f11[,j]^2)+sqrt(1-f22[,j]^2)-pi/2)+sqrt(1-f11[j,j]^2)+sqrt(1-f22[j,j]^2)-pi/2
G1[j]=sum(1-f0)-sum(1-f00[,j])+1-f00[j,j]
G2[j]=sum(2-f1^2-f2^2)-sum(2-f11[,j]^2-f22[,j]^2)+2-f11[j,j]^2-f22[j,j]^2
#G3[j]=sum(f1+f2-1)-sum(f11[,j]+f22[,j]-1)+f11[j,j]+f22[j,j]-1
```

```
G4[j]=sum(sqrt(1-f1^2)+sqrt(1-f2^2)-pi/2)-sum(sqrt(1-f11[,j]^2)+sqrt(1-f22[,j]^2)-pi/2)+sqrt(1-f11[j,j]^2)+sqrt(1-f22[j,j]^2)-pi/2
G3[j]=2*(sum(f1+f2)-sum(f11[,j]+f22[,j])+f11[j,j]+f22[j,j])-4*(sum(f1^3+f2^3)-sum(f11[,j]^3+f22[,j]^3)+f11[j,j]^3+f22[j,j]^3)
}
#G=cbind(G1,G2,G3,G4)
#ELM=el.test(G,c(0,0,0,0))$'-2LLR'
#a1=as.integer(ELM>qchisq(0.9,df=4))
#a2=as.integer(ELM>qchisq(0.95,df=4))
G=cbind(G1-G2/2+1/3,G3)
ELM=el.test(G,c(0,0))$'-2LLR'
a1=as.integer(ELM>qchisq(0.9,df=2))
a2=as.integer(ELM>qchisq(0.95,df=2))
#bootstrap
boot_r=1000
T=numeric(boot_r)
q=c(0,sort(pnorm(y-exp(alpha_hat*x),0,sqrt(beta_hat))),1)
T0=sum(((0:n)/n-q[-(n+2)])^3-((0:n)/n-q[-1])^3)/3
for (i in 1:boot_r)
{
err=rnorm(n,0,sqrt(beta_hat))
ystar=exp(alpha_hat*x)+err
h1=function(v) mean((z(ystar,x,v))^2)
tmpstar=try(optimize(h1,interval=c(-1,1)),silent=TRUE)
if(length(tmpstar)>1){
alpha_star=tmpstar$minimum
}
else{
alpha_star=0
}
beta_star=mean((ystar-exp(alpha_star*x))^2)
err_star=ystar-exp(alpha_star*x)
p=sort(pnorm(err_star,0,sqrt(max(beta_star,0.001))))
p=c(0,p,1)
T[i]=sum(((0:n)/n-p[-(n+2)])^3-((0:n)/n-p[-1])^3)/3
}
quant=quantile(T,probs=c(0.9,0.95))
aa=c((T0>quant[1]),(T0>quant[2]))
return(c(a1,a2,aa))
}


#Test t(beta=nu) against mixture with N(mu,sigma^2)#
ff_t_Composite=function(n,alpha,mu,sigma,nu,delta)
{
x=2*(runif(n)+1)
m=function(x,alpha) exp(alpha*x)
e=(1-delta/sqrt(n))*rt(n,nu)+delta/sqrt(n)*rnorm(n,mu,sigma)
z=function(y,x,alpha) y-m(x,alpha)
y=exp(alpha*x)+e
k=n/2
zalpha1star=matrix(rep(0,k^2),k,k)
```

```
zalpha2star=matrix(rep(0,k^2),k,k)
zalpha0star=matrix(rep(0,k^2),k,k)
G1=numeric(k)
G2=numeric(k)
G3=numeric(k)
G4=numeric(k)
f11=matrix(rep(0,k^2),k,k)
f22=matrix(rep(0,k^2),k,k)
f00=matrix(rep(0,k^2),k,k)
x1=x[1:k]
x2=x[(k+1):(2*k)]
y1=y[1:k]
y2=y[(k+1):(2*k)]
h0=function(v) mean((z(y,x,v))^2)
tmp=try(optimize(h0,interval=c(-1,1)),silent=TRUE)
if(length(tmp)>1){
alpha_hat=tmp$minimum
}
else{
alpha_hat=0
}
z0=z(y,x,alpha_hat)^2
z1=z(y1,x1,alpha_hat)^2
z2=z(y2,x2,alpha_hat)^2
h1=function(theta) (mean(digamma((exp(theta)+1)/2)-1/nu-digamma((exp(theta))/2)-log(1+z0/(exp(theta)))
+(exp(theta)+1)*z0/((exp(theta))^2+(exp(theta))*z0)))^2
tmp1=try(optimize(h1,interval=c(0.1,log(15))),silent=TRUE)
if(length(tmp1)>1){
beta_hat=exp(tmp1$minimum)
}
else{
beta_hat=0
}
alpha_hathat=alpha_hat+1/sum(x^2*y*exp(alpha_hat*x)-2*x^2*exp(2*alpha_hat*x))*(x1*y1*exp(alpha_hat*x1)
-x1*exp(2*alpha_hat*x1)+x2*y2*exp(alpha_hat*x2)-x2*exp(2*alpha_hat*x2))
h=1/2*digamma((beta_hat+1)/2)-1/(2*beta_hat)-1/2*digamma(beta_hat/2)-1/2*log(1+z0/beta_hat)
+(beta_hat+1)/2*z0/(beta_hat^2+z0*beta_hat)
h_beta=sum(1/4*trigamma((beta_hat+1)/2)+1/(2*beta_hat^2)-1/4*trigamma(beta_hat/2)+z0/(beta_hat^2+beta_hat*z0)
-(beta_hat+1)/2*z0*(2*beta_hat+z0)/(beta_hat^2+z0*beta_hat)^2)
h_alpha=sum((y-exp(alpha_hat*x))*x*exp(alpha_hat*x)/(beta_hat+z0)
+((beta_hat+1)*(-x*exp(alpha_hat*x)*(y-exp(alpha_hat*x))*(beta_hat+z0)
+x*exp(alpha_hat*x)*(y-exp(alpha_hat*x))^3)/(beta_hat*(beta_hat+z0)^2)))
beta_hathat=beta_hat+1/h_beta*h-1/h_beta*h_alpha*(alpha_hathat-alpha_hat)
zalpha1=z(y1,x1,alpha_hat)
zalpha2=z(y2,x2,alpha_hat)
f1=pt(zalpha1,df=max(beta_hat,0.001))
f2=pt(zalpha2,df=max(beta_hat,0.001))
f0=pmax(f1,f2)
for(j in 1:k)
```

```
{
zalpha1star[,j]=z(y1,x1,alpha_hathat[j])
zalpha2star[,j]=z(y2,x2,alpha_hathat[j])
f11[,j]=pt(zalpha1star[,j],df=max(beta_hathat[j],0.001))
f22[,j]=pt(zalpha2star[,j],df=max(beta_hathat[j],0.001))
}
f00=pmax(f11,f22)
for(j in 1:k)
{
#G1[j]=sum(2/3-f0)-sum(2/3-f00[,j])+2/3-f00[j,j]
#G2[j]=sum(2/3-f1^2-f2^2)-sum(2/3-f11[,j]^2-f22[,j]^2)+2/3-f11[j,j]^2-f22[j,j]^2
#G3[j]=sum(f1+f2-1)-sum(f11[,j]+f22[,j]-1)+f11[j,j]+f22[j,j]-1
#G4[j]=sum(sqrt(1-f1^2)+sqrt(1-f2^2)-pi/2)-sum(sqrt(1-f11[,j]^2)+sqrt(1-f22[,j]^2)-pi/2)
+sqrt(1-f11[j,j]^2)+sqrt(1-f22[j,j]^2)-pi/2
G1[j]=sum(1-f0)-sum(1-f00[,j])+1-f00[j,j]
G2[j]=sum(2-f1^2-f2^2)-sum(2-f11[,j]^2-f22[,j]^2)+2-f11[j,j]^2-f22[j,j]^2
G3[j]=sum(f1+f2-1)-sum(f11[,j]+f22[,j]-1)+f11[j,j]+f22[j,j]-1
G4[j]=sum(sqrt(1-f1^2)+sqrt(1-f2^2)-pi/2)-sum(sqrt(1-f11[,j]^2)+sqrt(1-f22[,j]^2)-pi/2)
+sqrt(1-f11[j,j]^2)+sqrt(1-f22[j,j]^2)-pi/2
G3[j]=2*(sum(f1+f2)-sum(f11[,j]+f22[,j])+f11[j,j]+f22[j,j])-4*(sum(f1^3+f2^3)-sum(f11[,j]^3
+f22[,j]^3)+f11[j,j]^3+f22[j,j]^3)
}
#G=cbind(G1,G2,G3,G4)
#ELM=el.test(G,c(0,0,0,0))$'-2LLR'
#a1=as.integer(ELM>qchisq(0.9,df=4))
#a2=as.integer(ELM>qchisq(0.95,df=4))
G=cbind(G1-G2/2+1/3,G3)
ELM=el.test(G,c(0,0))$'-2LLR'
a1=as.integer(ELM>qchisq(0.9,df=2))
a2=as.integer(ELM>qchisq(0.95,df=2))
#bootstrap
boot_r=1000
T=numeric(boot_r)
q=c(0,sort(pt(y-exp(alpha_hat*x),df=beta_hat)),1)
T0=sum(((0:n)/n-q[-(n+2)])^3-((0:n)/n-q[-1])^3)/3
for (i in 1:boot_r)
{
err=rt(n,df=beta_hat)
ystar=exp(alpha_hat*x)+err
h2=function(v) mean((z(ystar,x,v))^2)
tmpstar=try(optimize(h2,interval=c(-1,1)),silent=TRUE)
if(length(tmpstar)>1){
alpha_star=tmpstar$minimum
}
else{
alpha_star=0
}
z00=z(y,x,alpha_star)^2
h3=function(theta) (mean(digamma((exp(theta)+1)/2)-1/nu-digamma((exp(theta))/2)-log(1+z00/(exp(theta)))
```

```
+(exp(theta)+1)*z00/((exp(theta))^2+(exp(theta))*z00)))^2
tmp1star=try(optimize(h3,interval=c(0.1,log(15))),silent=TRUE)
if(length(tmp1star)>1){
beta_star=exp(tmp1star$minimum)
}
else{
beta_star=0
}
err_star=ystar-exp(alpha_star*x)
p=sort(pt(err_star,df=max(beta_star,0.001)))
p=c(0,p,1)
T[i]=sum(((0:n)/n-p[-(n+2)])^3-((0:n)/n-p[-1])^3)/3
}
quant=quantile(T,probs=c(0.9,0.95))
aa=c((T0>quant[1]),(T0>quant[2]))
return(c(a1,a2,aa))
}

#test N(mu=0,sigma=1) against mixture with t(nu)#
ff_normal_Simple=function(n,alpha,mu,sigma,nu,delta)
{
x=2*(runif(n)+1)
m=function(x,alpha) exp(alpha*x)
e=(1-delta/sqrt(n))*rnorm(n,mu,sigma)+delta/sqrt(n)*rt(n,nu)
z=function(y,x,alpha) y-exp(alpha*x)
y=m(x,alpha)+e
k=n/2
zalpha1star=matrix(rep(0,k^2),k,k)
zalpha2star=matrix(rep(0,k^2),k,k)
zalpha0star=matrix(rep(0,k^2),k,k)
G1=numeric(k)
G2=numeric(k)
G3=numeric(k)
G4=numeric(k)
f11=matrix(rep(0,k^2),k,k)
f22=matrix(rep(0,k^2),k,k)
f00=matrix(rep(0,k^2),k,k)
x1=x[1:k]
x2=x[(k+1):(2*k)]
y1=y[1:k]
y2=y[(k+1):(2*k)]
h=function(v) mean(z(y,x,v)^2)
a=try(optimize(h,interval=c(-1,1)),silent=TRUE)
if(length(a)>1){
alpha_hat=a$minimum
}
else{
alpha_hat=0
}
```

```
alpha_hathat=alpha_hat+
(sum(y*x^2*exp(alpha_hat*x)-2*x^2*exp(2*alpha_hat*x)))^(-1)*(x1*exp(alpha_hat*x1)*(y1-exp(alpha_hat*x1))
+x2*exp(alpha_hat*x2)*(y2-exp(alpha_hat*x2)))
zalpha1=z(y1,x1,alpha_hat)
zalpha2=z(y2,x2,alpha_hat)

for (j in 1:k)
{
zalpha1star[,j]=z(y1,x1,alpha_hathat[j])
zalpha2star[,j]=z(y2,x2,alpha_hathat[j])
}
f1=pnorm(zalpha1,mean=mu,sd=sigma)
f2=pnorm(zalpha2,mean=mu,sd=sigma)
f0=pmax(f1,f2)
f11=pnorm(zalpha1star,mean=mu,sd=sigma)
f22=pnorm(zalpha2star,mean=mu,sd=sigma)
f00=pmax(f11,f22)
for(j in 1:k)
{
#G1[j]=sum(2/3-f0)-sum(2/3-f00[,j])+2/3-f00[j,j]
#G2[j]=sum(2/3-f1^2-f2^2)-sum(2/3-f11[,j]^2-f22[,j]^2)+2/3-f11[j,j]^2-f22[j,j]^2
#G3[j]=sum(f1+f2-1)-sum(f11[,j]+f22[,j]-1)+f11[j,j]+f22[j,j]-1
#G4[j]=sum(sqrt(1-f1^2)+sqrt(1-f2^2)-pi/2)-sum(sqrt(1-f11[,j]^2)+sqrt(1-f22[,j]^2)-pi/2)
+sqrt(1-f11[j,j]^2)+sqrt(1-f22[j,j]^2)-pi/2
G1[j]=sum(1-f0)-sum(1-f00[,j])+1-f00[j,j]
G2[j]=sum(2-f1^2-f2^2)-sum(2-f11[,j]^2-f22[,j]^2)+2-f11[j,j]^2-f22[j,j]^2
G3[j]=sum(f1+f2-1)-sum(f11[,j]+f22[,j]-1)+f11[j,j]+f22[j,j]-1
G4[j]=sum(sqrt(1-f1^2)+sqrt(1-f2^2)-pi/2)-sum(sqrt(1-f11[,j]^2)+sqrt(1-f22[,j]^2)-pi/2)
+sqrt(1-f11[j,j]^2)+sqrt(1-f22[j,j]^2)-pi/2
G3[j]=2*(sum(f1+f2)-sum(f11[,j]+f22[,j])+f11[j,j]+f22[j,j])-4*(sum(f1^3+f2^3)-sum(f11[,j]^3
+f22[,j]^3)+f11[j,j]^3+f22[j,j]^3)
}
#G=cbind(G1,G2,G3,G4)
#ELM=el.test(G,c(0,0,0,0))$'-2LLR'
#a1=as.integer(ELM>qchisq(0.9,df=4))
#a2=as.integer(ELM>qchisq(0.95,df=4))
G=cbind(G1-G2/2+1/3,G3)
ELM=el.test(G,c(0,0))$'-2LLR'
a1=as.integer(ELM>qchisq(0.9,df=2))
a2=as.integer(ELM>qchisq(0.95,df=2))
#bootstrap#
boot_r=1000
T=numeric(boot_r)
q=c(0,sort(pnorm(y-exp(alpha_hat*x),mean=mu,sd=sigma)),1)
T0=sum(((0:n)/n-q[-(n+2)])^3-((0:n)/n-q[-1])^3)/3
for (i in 1:boot_r)
{
err=rnorm(n,mean=mu,sd=sigma)
ystar=exp(alpha_hat*x)+err
```

```
h2=function(v) mean(z(ystar,x,v)^2)
tmpstar=try(optimize(h2,interval=c(-1,1)),silent=TRUE)
if(length(tmpstar)>1){
alpha_star=tmpstar$minimum
}
else{
alpha_star=0
}
err_star=ystar-exp(alpha_star*x)
p=sort(pnorm(err_star,mean=mu,sd=sigma))
p=c(0,p,1)
T[i]=sum(((0:n)/n-p[-(n+2)])^3-((0:n)/n-p[-1])^3)/3
}
quant=quantile(T,probs=c(0.9,0.95))
aa=c((T0>quant[1]),(T0>quant[2]))
return(c(a1,a2,aa))
}

#test t(nu=4) against mixture with N(mu,sigma^2)#
ff_t_Simple=function(n,alpha,mu,sigma,nu,delta)
{
x=2*(runif(n)+1)
m=function(x,alpha) exp(alpha*x)
e=(1-delta/sqrt(n))*rt(n,nu)+delta/sqrt(n)*rnorm(n,mu,sigma)
z=function(y,x,alpha) y-m(x,alpha)
y=exp(alpha*x)+e
k=n/2
x1=x[1:k]
x2=x[(k+1):(2*k)]
y1=y[1:k]
y2=y[(k+1):(2*k)]
zalpha1star=matrix(rep(0,k^2),k,k)
zalpha2star=matrix(rep(0,k^2),k,k)
zalpha0star=matrix(rep(0,k^2),k,k)
G1=numeric(k)
G2=numeric(k)
G3=numeric(k)
G4=numeric(k)
f11=matrix(rep(0,k^2),k,k)
f22=matrix(rep(0,k^2),k,k)
f00=matrix(rep(0,k^2),k,k)
h0=function(v) mean((z(y,x,v))^2)
tmp=try(optimize(h0,interval=c(-1,1)),silent=TRUE)
if(length(tmp)>1){
alpha_hat=tmp$minimum
}
else{
alpha_hat=0
}
```

```
z0=z(y,x,alpha_hat)^2
z1=z(y1,x1,alpha_hat)^2
z2=z(y2,x2,alpha_hat)^2
alpha_hathat=alpha_hat+1/sum(x^2*y*exp(alpha_hat*x)-2*x^2*exp(2*alpha_hat*x))*(x1*y1*exp(alpha_hat*x1)
-x1*exp(2*alpha_hat*x1)+x2*y2*exp(alpha_hat*x2)-x2*exp(2*alpha_hat*x2))
zalpha1=z(y1,x1,alpha_hat)
zalpha2=z(y2,x2,alpha_hat)
f1=pt(zalpha1,df=nu)
f2=pt(zalpha2,df=nu)
f0=pmax(f1,f2)
for(j in 1:k)
{
zalpha1star[,j]=z(y1,x1,alpha_hathat[j])
zalpha2star[,j]=z(y2,x2,alpha_hathat[j])
f11[,j]=pt(zalpha1star[,j],df=nu)
f22[,j]=pt(zalpha2star[,j],df=nu)
}
f00=pmax(f11,f22)

for(j in 1:k)
{
#G1[j]=sum(2/3-f0)-sum(2/3-f00[,j])+2/3-f00[j,j]
#G2[j]=sum(2/3-f1^2-f2^2)-sum(2/3-f11[,j]^2-f22[,j]^2)+2/3-f11[j,j]^2-f22[j,j]^2
#G3[j]=sum(f1+f2-1)-sum(f11[,j]+f22[,j]-1)+f11[j,j]+f22[j,j]-1
#G4[j]=sum(sqrt(1-f1^2)+sqrt(1-f2^2)-pi/2)-sum(sqrt(1-f11[,j]^2)+sqrt(1-f22[,j]^2)-pi/2)
+sqrt(1-f11[j,j]^2)+sqrt(1-f22[j,j]^2)-pi/2
G1[j]=sum(1-f0)-sum(1-f00[,j])+1-f00[j,j]
G2[j]=sum(2-f1^2-f2^2)-sum(2-f11[,j]^2-f22[,j]^2)+2-f11[j,j]^2-f22[j,j]^2
G3[j]=sum(f1+f2-1)-sum(f11[,j]+f22[,j]-1)+f11[j,j]+f22[j,j]-1
G4[j]=sum(sqrt(1-f1^2)+sqrt(1-f2^2)-pi/2)-sum(sqrt(1-f11[,j]^2)+sqrt(1-f22[,j]^2)-pi/2)
+sqrt(1-f11[j,j]^2)+sqrt(1-f22[j,j]^2)-pi/2
G3[j]=2*(sum(f1+f2)-sum(f11[,j]+f22[,j])+f11[j,j]+f22[j,j])-4*(sum(f1^3+f2^3)-sum(f11[,j]^3
+f22[,j]^3)+f11[j,j]^3+f22[j,j]^3)
}
#G=cbind(G1,G2,G3,G4)
#ELM=el.test(G,c(0,0,0,0))$'-2LLR'
#a1=as.integer(ELM>qchisq(0.9,df=4))
#a2=as.integer(ELM>qchisq(0.95,df=4))
G=cbind(G1-G2/2+1/3,G3)
ELM=el.test(G,c(0,0))$'-2LLR'
a1=as.integer(ELM>qchisq(0.9,df=2))
a2=as.integer(ELM>qchisq(0.95,df=2))
#bootstrap #
boot_r=1000
T=numeric(boot_r)
q=c(0,sort(pt(y-exp(alpha_hat*x),df=nu)),1)
T0=sum(((0:n)/n-q[-(n+2)])^3-((0:n)/n-q[-1])^3)/3
for (i in 1:boot_r){
err=rt(n,df=nu)
```

```
ystar=exp(alpha_hat*x)+err
h2=function(v) mean(z(ystar,x,v)^2)
tmpstar=try(optimize(h2,interval=c(-1,1)),silent=TRUE)
if(length(tmpstar)>1){
alpha_star=tmpstar$minimum
}
else{
alpha_star=0
}
err_star=ystar-exp(alpha_star*x)
p=sort(pt(err_star,df=nu))
p=c(0,p,1)
T[i]=sum(((0:n)/n-p[-(n+2)])^3-((0:n)/n-p[-1])^3)/3
}
quant=quantile(T,probs=c(0.9,0.95))
aa=c((T0>quant[1]),(T0>quant[2]))
return(c(a1,a2,aa))
}

set.seed(__MQSUB1__)
n=500
r=1000
alpha=0.25
delta=3
nu=8
res1=0*matrix(1:(r*4),ncol=4)
res2=0*matrix(1:(r*4),ncol=4)
res3=0*matrix(1:(r*4),ncol=4)
res4=0*matrix(1:(r*4),ncol=4)
res5=res1
res6=res1
for(i in 1:r){
res1[i,]=ff_normal_Simple(n,alpha,0,1,nu,delta)
res2[i,]=ff_t_Simple(n,alpha,0,1,nu,delta)
res3[i,]=ff_normal_Composite(n,alpha,0,1,nu,delta)
res4[i,]=ff_t_Composite(n,alpha,0,1,nu,delta)
res5[i,]=ff_normal_Composite(n,alpha,0,1,nu,n^0.5)
res6[i,]=ff_t_Composite(n,alpha,0,1,nu,n^0.5)
#print(c(res1[i,],res2[i,],res3[i,],res4[i,],res5[i,],res6[i,]))
}
out1=c(mean(res1[,1]),mean(res1[,2]),mean(res1[,3]),mean(res1[,4]))
out2=c(mean(res2[,1]),mean(res2[,2]),mean(res2[,3]),mean(res2[,4]))
out3=c(mean(res3[,1]),mean(res3[,2]),mean(res3[,3]),mean(res3[,4]))
out4=c(mean(res4[,1]),mean(res4[,2]),mean(res4[,3]),mean(res4[,4]))
out5=c(mean(res5[,1]),mean(res5[,2]),mean(res5[,3]),mean(res5[,4]))
out6=c(mean(res6[,1]),mean(res6[,2]),mean(res6[,3]),mean(res6[,4]))
#print(out)
myname=paste("n",n,"_v",nu,"_delta",delta,"_",__MQSUB2__,".Rdata",sep="")
save.image(file=myname)
```

```
}
```

5.3 Codes for Chapter IV

```
{
library(emplik)
library(stats)
##To run entire process, just run compare(n) for some specified n value
##bootstrapFun() is only one iteration
bootstrapFun <- function(b,sigma,df0,s,BN,DIS) {
xColumns <- function() {
BUP=500
if(DIS==1){
    epsilon <- rnorm(s+BUP, mean=0, sd=sigma) #s values for epsilon}
if(DIS==2){
epsilon=rt(s+BUP,df=df0)}
    x <- numeric(s+BUP) #create vector to store x values
    x[1] <- 0
    x[2] <- 0
    for (i in 3:(s+BUP)) {
      x[i] <- b * epsilon[i-1] * x[i-2] + epsilon[i]}
    return(x[(BUP+1):(BUP+s)])
 }
xSample <- xColumns() #x values
listMatFun <- function(col) { #this function finds one b estimate
  s=length(col)
  outOne<-(sum(col[(3:s)-2]^2)*sum(col[3:s]^2*col[(3:s)-2]^2)-sum(col[3:s]^2)
      *sum(col[(3:s)-2]^4))/((sum(col[(3:s)-2]^2))^2-(s-2)*sum(col[(3:s)-2]^4))
  #outT<-((s-2)*sum(col[3:s]^2*col[3:s-2]^2)-sum(col[3:s]^2)*sum(col[3:s-2]^2)
  #       )/((s-2)*sum(col[3:s-2]^4)-(sum(col[3:s-2]^2))^2)
  #after burn up, good.
  outTwo<- sum(col[3:s]*col[(3:s)-1]*col[(3:s)-2])/sum(col[(3:s)-2]^2)
  bH <- outTwo/outOne #this is the b estimate for the x sample
  #bH <- min(bH, 1/sigma-0.0000001)
  return(bH)
}
bHat <- listMatFun(xSample) #what you need as input for part 2


#####Part 2#####
#function to do this for each x matrix from the list?
eFun <- function(matIn) {
  s=length(matIn)
   eH <- numeric(s)
   eH[1] <- 0
   eH[2] <- 0
   for (i in 3:s) eH[i] <- matIn[i] - bHat * eH[i-1] * matIn[i-2]
```

```
   return(eH[3:s])
}
eHat <- eFun(xSample) ##creates eHat values based on xSample
uFun <- function(eH) {
  s=length(eH)
   eStar<-matrix(NA,s,BN)
  filleStar<-function()
      {
       uInt<-sample(1:s,s,replace=TRUE)
       eStar=eH[uInt]
       return(eStar)
      }
   eStar<-replicate(BN, filleStar())
   xStar <- matrix(NA, s, BN)
   xStar[1,] <- 0
   xStar[2,] <- 0
   for (i in 3:s) xStar[i,] <- bHat * eStar[i-1,] * xStar[i-2,] + eStar[i,]
   return(xStar[3:s,])
}
bootFun <- function() {
xStar <- uFun(eHat) #new matrix sxBN xStar values
##now use same funtion from above to get BN new bStar
bStar<-apply(xStar,2,listMatFun) #BN estimate
        return(bStar)}
bootOut <- bootFun() # vector of BN bStar values
covFun <- function() { #see if initial b value falls in this interval
aa=c(0,0)
if(is.na(sum(bootOut))==F && is.na(bHat)==F){
tStar <- bootOut - bHat
tStarOrdered <- sort(tStar, decreasing=FALSE)
c1 <- tStarOrdered[as.integer(BN*.025)]
c2 <- tStarOrdered[as.integer(BN*.975)]
c3 <- tStarOrdered[as.integer(BN*.05)]
c4 <- tStarOrdered[as.integer(BN*.95)]
l1 <- bHat-c2
r1 <- bHat-c1
  l2 <- bHat-c4
r2 <- bHat-c3
if (b>=l2 && b<=r2) {aa[1]=1}
if (b>=l1 && b<=r1) {aa[2]=1}
}
        return(aa)
}
coverage <- covFun() #b will be the initial value (.5)

##############################
#PELM
x0=xSample[3:s]
x1=xSample[3:s-1]
```

```
x2=xSample[3:s-2]
f=function(sig2,b_hat)
   {
   T1=2*(x0^2-exp(sig2)-b_hat^2*exp(sig2)*x2^2)*b_hat*x2^2/(1+x2^2)+(x0*x1-b_hat*exp(sig2)*x2)*x2/(1+x2^2)
   T2=(x0^2-exp(sig2)-b_hat^2*exp(sig2)*x2^2)*(1+b_hat^2*x2^2)/(1+x2^2)+
      (x0*x1-b_hat*exp(sig2)*x2)*b_hat*x2/(1+x2^2)
ratio=el.test(cbind(T1,T2),c(0,0))$'-2LLR'
#T1= (x0^2-exp(sig2)-b_hat^2*exp(sig2)*x2^2)*x2^2
#T2=(x0^2-exp(sig2)-b_hat^2*exp(sig2)*x2^2)
#T3=(x0*x1-b_hat*exp(sig2)*x2)*x2
#ratio=el.test(cbind(T2,T3),c(0,0))$'-2LLR'
   return(ratio)
   }
ff=function(sig2,b_hat)
   {
  # T1=2*(x0^2-exp(sig2)-b_hat^2*exp(sig2)*x2^2)*b_hat*x2^2+(x0*x1-b_hat*exp(sig2)*x2)*x2
   #T2=(x0^2-exp(sig2)-b_hat^2*exp(sig2)*x2^2)*(1+b_hat^2*x2^2)+
      (x0*x1-b_hat*exp(sig2)*x2)*b_hat*x2
#ratio=el.test(cbind(T1,T2),c(0,0))$'-2LLR'
#T1= (x0^2-exp(sig2)-b_hat^2*exp(sig2)*x2^2)*x2^2
T2=(x0^2-exp(sig2)-b_hat^2*exp(sig2)*x2^2)/(1+x2^2)
T3=(x0*x1-b_hat*exp(sig2)*x2)*x2/(1+x2^2)
ratio=el.test(cbind(T2,T3),c(0,0))$'-2LLR'
   return(ratio)
   }
fnew=function(sig2,b_hat)
   {
   #T1=2*(x0^2-exp(sig2)-b_hat^2*exp(sig2)*x2^2)*b_hat*x2^2+(x0*x1-b_hat*exp(sig2)*x2)*x2
   #T2=(x0^2-exp(sig2)-b_hat^2*exp(sig2)*x2^2)*(1+b_hat^2*x2^2)+
      (x0*x1-b_hat*exp(sig2)*x2)*b_hat*x2
#ratio=el.test(cbind(T1,T2),c(0,0))$'-2LLR'
T1= (x0^2-exp(sig2)-b_hat^2*exp(sig2)*x2^2)*x2^2/(1+x2^2)
T2=(x0^2-exp(sig2)-b_hat^2*exp(sig2)*x2^2)/(1+x2^2)
T3=(x0*x1-b_hat*exp(sig2)*x2)*x2/(1+x2^2)
ratio=el.test(cbind(T1,T2,T3),c(0,0,0))$'-2LLR'
   return(ratio)
   }
fnewnew=function(thetanew)
   { sig2=thetanew[1]
     b_hat=thetanew[2]
   #T1=2*(x0^2-exp(sig2)-b_hat^2*exp(sig2)*x2^2)*b_hat*x2^2+(x0*x1-b_hat*exp(sig2)*x2)*x2
   #T2=(x0^2-exp(sig2)-b_hat^2*exp(sig2)*x2^2)*(1+b_hat^2*x2^2)+
      (x0*x1-b_hat*exp(sig2)*x2)*b_hat*x2
#ratio=el.test(cbind(T1,T2),c(0,0))$'-2LLR'
T1= (x0^2-exp(sig2)-b_hat^2*exp(sig2)*x2^2)*x2^2/(1+x2^2)
T2=(x0^2-exp(sig2)-b_hat^2*exp(sig2)*x2^2)/(1+x2^2)
T3=(x0*x1-b_hat*exp(sig2)*x2)*x2/(1+x2^2)
ratio=el.test(cbind(T1,T2,T3),c(0,0,0))$'-2LLR'
   return(ratio)
```

```
   }
if(DIS==1){
inisig2=log(sigma^2)
}
if(DIS==2){
inisig2=log(df0/(df0-2))
}
a=c(0,0)
mini=100000
for(j in 1:10){
val=try(nlm(f,inisig2-(j-5)*0.1,b_hat=b),silent=T)
if(length(val)>1){
tmp=val$minimum
if(tmp<mini){
mini=tmp
}
}
}
a=c(as.integer(mini<=qchisq(0.9,df=1)),as.integer(mini<=qchisq(0.95,df=1)))
aa=c(0,0)
minimini=100000
for(j in 1:10){
val=try(nlm(ff,inisig2-(j-5)*0.1,b_hat=b),silent=T)
if(length(val)>1){
tmp=val$minimum
if(tmp<minimini){
minimini=tmp
}
}
}
aa=c(as.integer(minimini<=qchisq(0.9,df=1)),as.integer(minimini<=qchisq(0.95,df=1)))
anew=c(0,0)
mini1=100000
for(j in 1:10){
val=try(nlm(fnew,inisig2-(j-5)*0.1,b_hat=b),silent=T)
if(length(val)>1){
tmp=val$minimum
if(tmp<mini1){
mini1=tmp
}
}
}
mini2=100000
for(j in 1:10){
val=try(nlm(fnewnew,c(inisig2-(j-5)*0.1,b-(j-5)*0.04)),silent=T)
if(length(val)>1){
tmp=val$minimum
if(tmp<mini2){
mini2=tmp
```

```
}
}
}
anew=c(as.integer(mini1-mini2<=qchisq(0.9,df=1)),as.integer(mini1-mini2<=qchisq(0.95,df=1)))
###vector to combine coverage and test -- this works!! can be indexed into as well
comparison <- c(coverage, a,aa,anew)
#print(comparison)
return(comparison)
}
compare <- function(b,sigma,df0,s,n,BN,DIS) { #probabilities for n iterations
        prob <- replicate(n, bootstrapFun(b,sigma,df0,s,BN,DIS)) #this is set of n coverage values
coverageProb <- c(mean(prob[1,]),mean(prob[2,]))  #should be close to 90%- 95%
profileProb1<-c(mean(prob[3,]),mean(prob[4,]))
        profileProb2<-c(mean(prob[5,]),mean(prob[6,]))
profileProb3<-c(mean(prob[7,]),mean(prob[8,]))
probVector <- c(coverageProb,profileProb1, profileProb2,profileProb3)
return(probVector)
}

set.seed(__MQSUB1__)
b=0.4
sigma=1
df0=10
s=1000
n=1000
BN=1000
DIS=1
#DIS=2
if(DIS==1){
myname=paste("DF",DIS,"_N",s,"_b",b,"_sigma",sigma,"_",__MQSUB2__,".Rdata",sep="")
}
if(DIS==2){
myname=paste("DF",DIS,"_N",s,"_b",b,"_df",df0,"_",__MQSUB2__,".Rdata",sep="")
}
res=compare(b,sigma,df0,s,n,BN,DIS)
#print(res)
save.image(file=myname)
     }
```

REFERENCES

[1] Basrak, B., Davis, R.A. and Mikosch, T. (1999). The sample ACF of a simple bilinear process. *Stochastic Processes and their Applications, 83, 1-14.*

[2] Berk, R.H. and Jones, D.H. (1979) Goodness-of-fit test statistic that dominate the Kolmogorov statistics. *Z. Wahrs. Verw. Geb. 47, 47–59.*

[3] Chan, N.H. and Ling, S. (2006). Empirical likelihood for GARCH models. *Econometric Theory, 22, 403-428.*

[4] Chen, J. and Qin, J. (1993). Empirical likelihood estimation for finite populations and the effective usage of auxiliary information. *Biometrika, 80, 107-116.*

[5] Chen, S.X. and Van Keilegom, I. (2009). A review on empirical likelihood methods for regression (with discussions). *Test, 18, 415-474.*

[6] Chuang, C.S. and Chan, N.H. (2002). Empirical likelihood for autoregressive models, with applications to unstable time series. *Statistica Sinica, 12, 387–407.*

[7] D'Agostino, R.B. and Stephens, M.A. (1986). Goodness-of-Fit Techniques. *Marcel Dekker, Inc.*

[8] Davison, A.C. (2008). Statistical Models. *Cambridge University Press.*

[9] Einmahl, J.H.J. and MeKeague, I.W. (2003). Empirical likelihood based hypothesis testing. *Bernoulli 9, 267–290.*

[10] Feng, H. and Peng, L. (2012). Jackknife Empirical Likelihood Tests For Distribution Functions. *Journal of Statistical Planning and Inference 142, 1571-1585.*

[11] Feng, H. and Peng, L. (2012). Jackknife Empirical Likelihood Tests For Error Distributions in Regres- sion Models. *Journal of Multivariate Analysis 112, 63-75.*

[12] Feng, H., Peng, L. and Zhu, F. (2012). Empirical Likelihood Tests For Simple Bilinear Models. Submitted.

[13] Fisher, R. A.(1924). The conditions under which xhi-square measures the discrepancy between observation and hypothesis. *J. Roy. Statist. Soc. 87, 442-450.*

[14] Giordano, F. (2000). The variance of CLS estimators for a simple bilinear model. *Quaderni di Statistica, 2, 147-155.*

[15] Giordano, F. (2004). The univariate distribution function for a particular bilinear model. *Statistical Methods & Applications, 13, 167-178.*

[16] Giordano, F. and Vitale, C. (2003). CLS asymptotic variance for a particular relevant bilinear time series model. *Statistical Methods & Applications, 12, 169-185.*

[17] Gong, Y., Peng, L. and Qi, Y. (2010). Smoothed jackknife empirical likelihood method for ROC curve. *J. Multi. Anal. 101, 1520–1531.*

[18] Grahn, T. (1995). A conditional least squares approach to bilinear time series estimation. *Journal of Time Series Analysis, 16, 509-529.*

[19] Hall, P. and Heyde, C.C. (1980). Martingale Limit Theory and Its Application. *New York: Academic Press.*

[20] Jager, L. and Wellner, J.A. (2007). Goodness-of-fit tests via phi-divergences. *Ann. Statist. 35, 2018–2053.*

[21] Jing, B, Yuan, J. and Zhou, W. (2009). Jackknife empirical likelihood. *J. Amer. Statist. Assoc. 104, 1224–1232.*

[22] Kim, W.K., Billard, L. and Basawa, I.V. (1990). Estimation for the first-order diagonal bilinear time series model. *Journal of Time Series Analysis, 11, 215-229.*

[23] Khmaladze, E.V. and Koul, H.L. (2004). Martingale transformations goodness-of-fit tests in regression models. *Ann. Statist. 32, 995–1034.*

[24] Koul, H. and Sakhanenko, L. (2005). Goodness-of-fit testing in regression: a finite sample comparison of bootstrap methodology and Khmaladze transformation. *Statist. & Probab. Lett. 74, 290–302.*

[25] Li, M. and Peng, L. (2011). Empirical Likelihood Test via Estimating Equations. *Journal of Statistical Planning and Inference*

[26] Li, M, Peng, L. and Qi, Y. (2011). Reduce computation in profile empirical likelihood method. *Canadian Journal of Statistics 39, 370–384.*

[27] Li, M., Peng, L. and Qi, Y. (2011). Jackknife empirical likelihood method based on estimating equations. *Canadian Journal of Statistics 39, 370–384.*

[28] Neymam, J. (1949). Contribution to the theory of the chi-square test. *Proc. Berkeley Symp. Math. Statist. and Prob., 239-273.*

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

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

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

[32] Karl Pearson (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philosophical Magazine, Series 5 50 (302), 157-175*

[33] Peng, L. (2011). Approximate jackknife empirical likelihood method for estimating equations. *Canadian Journal of Statistics.*

[34] Peng, L. and Qi, Y. (2010). Smoothed jackknife empirical likelihood method for tail copulas. *Test 19, 514–536.*

[35] Peng, L., Qi, Y. and Van Keilegom, I. (2011). Jackknife empirical likelihood method for copulas. *Test, to appear.*

[36] Pham, D.T. and Tran, L.T. (1981). On the first-order bilinear time series model. *Journal of Applied Probability, 18, 617-627.*

[37] Qin. J. (1993). Empirical likelihood in biased sample problems. *Ann. Statist., 21, 1182-1196.*

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

[39] Rao, K. C. and robson, D. S. (1974). SA chi-square statistic for goodness-of-fit within the exponential family. *Comm. Statist. 3, 1139-1153.*

[40] Shao, J. (2003). Mathematical Statistics (second edition). *Springer, New York.*

[41] Serfling, R. (2002). Approximation Theorems of Mathematical Statistics. *John Wiley & Sons, INC.*

[42] Stephens, M.A. (1966). Statistics connected with the uniform distribution: percentage points and applications to testing for randomness of directions. *Biometrka 53, 235–239.*

[43] Subba Rao, T. (1981). On the theory of bilinear time series models. *Journal of the Royal Statistical Society Series B, 43, 244-255.*

[44] Watson G. S. (1957). The chi-squared goodness-of-fit test for normal distributions. *Biometrika 44, 336-348.*

[45] Watson G. S. (1958). On chi-square goodness-of-fit tests for continuous distributions. *J. Roy. Statist. Soc. Ser. B20, 44-61.*

[46] Watson G. S. (1959). Some recent results in chi-square goodness-of-fit tests. *Biometrics 15, 440-468.*

[47] Wilks, S. (1938). The large-sample distribution of the likelihood ratio for testing composite hypotheses. *Ann. Statist. 9, 60-62.*

VITA

Huijun Feng was born in a small quiet historical city called Tai'an in China, where Mountain Tai is the most famous landmark and every emperor in Chinese history came to pray for the country's peace, civilians' safe, and good harvest. This place is believed under God's bless, and its name, in Chinese, means peace and safe.