Fall 2013

A Jackknife Empirical Likelihood Approach To Goodness Of Fit U-Statistic Testing With Side Information

Qun Lin
Purdue University

Graduate School ETD Form 9
(Revised 12/07)

PURDUE UNIVERSITY
GRADUATE SCHOOL
Thesis/Dissertation Acceptance

This is to certify that the thesis/dissertation prepared

By Qun Lin

Entitled
A JACKKNIFE EMPIRICAL LIKELIHOOD APPROACH TO GOODNESS OF FIT U-STATISTIC TESTING WITH SIDE INFORMATION

For the degree of Doctor of Philosophy

Is approved by the final examining committee:

Hanxiang Peng
Chair

Wei Zheng

Benzion Boukai

Guang Cheng

To the best of my knowledge and as understood by the student in the *Research Integrity and Copyright Disclaimer (Graduate School Form 20)*, this thesis/dissertation adheres to the provisions of Purdue University's "Policy on Integrity in Research" and the use of copyrighted material.

Approved by Major Professor(s): Hanxiang Peng

Approved by: Evgeny Mukhin 11/18/2013
Head of the Graduate Program Date

A JACKKNIFE EMPIRICAL LIKELIHOOD APPROACH TO GOODNESS OF FIT U-STATISTIC TESTING WITH SIDE INFORMATION

A Dissertation

Submitted to the Faculty

of

Purdue University

by

Qun Lin

In Partial Fulfillment of the

Requirements for the Degree

of

Doctor of Philosophy

December 2013

Purdue University

Indianapolis, Indiana

ACKNOWLEDGMENTS

I want to acknowledge the effort from Professor Hanxiang Peng, who graciously became to be my thesis advisor three years ago and inspired me to start the research in the area of empirical likelihood.

I would like to thank Professor Wei Zheng, who helped me on the matrix theory that makes the work more elegant; to Professor Benzion Boukai, who admitted me to the statistics program long time ago and supported me through the time.

I would also like to acknowledge the effort from Professor Zhongmin Shen, who helped me to become Dr. Peng's student.

Finally, I would like to thank the Department of Mathematical Sciences of IUPUI, without whom this thesis would not have been possible at all.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Lin, Qun Ph.D., Purdue University, December 2013. A Jackknife Empirical Likelihood Approach To Goodness of Fit U-Statistic Testing With Side Information. Major Professor: Hanxiang Peng.

Motivated by applications to goodness of fit U-statistic testing, the jackknife empirical likelihood (JEL) of Jing, *et al.* (2009) is justified with an alternative approach, and the Wilks theorem for vector U-statistics is proved. This generalizes Owen's empirical likelihood from a vector mean to a vector U-statistic-based mean and Jing's JEL for univariate U-statistics to vector U-statistics and includes the JEL for U-statistics with side information as a special case. The results are generalized to allow for the constraints to use estimated criteria functions and for the number of constraints to grow with the sample size. The latter is needed to handle naturally occurring nuisance parameters in semiparametric models. The developed theory is applied to derive the empirical-likelihood-based goodness-of-fit tests and confidence sets for U-quantiles with finite number and growing number of constraints in the Theil-estimator-based test about the slope in a simple linear regression; for the Wilcoxon signed rank test about symmetry with an unknown center; for Kendall's tau and Goodman and Kruskal's Gamma with side information; for the test about independence of two categorical outcomes; for the joint confidence sets of variances in a balanced random effects model and for the simplicial depth function with a finite number and growing number of constraints. Some of the proposed jackknife empirical likelihood based goodness of fit tests are asymptotically distribution free. A simulation study is conducted to evaluate the behaviors of the Theil test with finite number and growing number of constraints.

1. INTRODUCTION AND JACKKNIFE EMPIRICAL LIKELIHOOD

1.1 Introduction

In a series of papers on constructing confidence intervals in a nonparametric setting, Owen (1988, 1990, 2001) introduced the empirical likelihood approach. As a likelihood approach with nonparametric properties, it does not require specification of a distribution for the data and often yields more efficient estimates of the parameters than many common estimators. It allows data to decide the shape of confidence sets and is Bartlett correctable (DiCiccio, Hall and Romano, 1991). The approach has been extended to various situations, e.g., to generalized linear models (Kolaczyk, 1994), local linear smoother (Chen and Qin, 2000), partially linear models (Shi and Lau, 2000; Wang and Jing, 2003), parametric and semiparametric models in multiresponse regression (Chen and Van Keilegom, 2009), linear regression with censored data (Zhou and Li, 2008), and plug-in estimates of nuisance parameters in estimating equations in the context of survival analysis (Qin and Jing, 2001; Wang and Jing, 2001; Li and Wang, 2003), time series (Nordman and Lahiri, 2006). Qin and Lawless (1994) linked empirical likelihood with finitely many estimating equations. These estimating equations serve as finitely many equality constraints. Maximum empirical likelihood estimators for the irregular case were studied in (Lopez, Van Keilegom and Veraverbeke, 2009). Chen, Peng and Qin (2009) obtained asymptotic normality for the number of constraints growing to infinity. Hjort, McKeague and Van Keilegom (2009) and Peng and Schick (2013) generalized the empirical likelihood approach to allow for the number of constraints to grow with the sample size and for the constraints to use estimated criteria functions. The latter is needed to handle naturally occurring nuisance parameters. Peng (2013) discovered a class of maximum empirical likelihood estimators which are tractable and studied their asymptotic properties. Algorithms, calibration and higher-order precision of the approach can be found in Hall and La Scala (1990), Emerson and Owen (2009) and Liu and Chen (2010) among others.

U-statistics are useful and many popular statistics can be expressed in U-statistics, see e.g. Kowalski and Tu (2008), Lee (1990), and Serfling (1980). By exploiting the asymptotic independence of the jackknife pseduo values of U-statistics, Jing *et al.* (2009) introduced the jackknife empirical likelihood (JEL) for U-statistics. As in the case of empirical likelihood for time series (Nordman and Lahiri, 2006), the independence or at least asymptotic independence which justifies the definition of empirical likelihood as a product of probabilities is not directly available for a U-statistic of which the summands are not independent but correlated. Moreover, the usual empirical likelihood in this case involves in the nonlinearality of π_j's in the constraint equations, which leads to the consequence that there are no explicit formulas for π_j's as there are in the usual empirical likelihood. This causes difficulty in deriving its asymptotic theory. Jing *et al.* (2009) noticed the asymptotic independence of the jackknife pseudo values of a U-statistic and introduced their jackknife empirical likelihood for U-statistics. They established the Wilks theorems for one- and two-sample U-statistics by exploiting the nice properties of the jackknife pseudo values of a U-statistic, for example, the sum of these values is equal to the U-statistic and the sample variance of them is an asymptotically unbiased estimator of the asymptotic variance of the U-statistic in its asymptotic normal distribution. These two properties are crucial in obtaining the asymptotic chi-squared distribution in the Wilks theorems. In justifying the asymptotic independence, Jing *et al.* (2009) cited a theorem from Shi (1984), who proved the asymptotic independence by applying the zero-one law for a sequence of exchangeable random variables. Shi's result is not easily available as it was published in Chinese and also for the sake of self-containedness, we present an alternative, somewhat straightforward, justification of the asymptotic independence based on the Hoeffding decomposition for U-statistics, see Section 1.2. Sometimes we shall abbreviate jackknife empirical likelihood as empirical likelihood.

Often additional information about the model is available. Information is usually expressed by usual constraint equations or even vector U-statistic-defined constraint equations, see our motivating examples in Section 2. It is common in semiparametric models that constraint functions contain unknown nuisance parameters and must be estimated. This is the case in testing symmetry of a distribution when the center of symmetry is unknown, see Example 2; in constructing confidence sets for variance components when side information is available which contains unknown nuisance parameters, see Subsection 2.2. Also, in semiparametric settings, information on the model can often be expressed by means of *infinitely* many constraints which may also depend on parameters of the model. In goodness of fit testing, many test statistics can be expressed in U-statistics, and the null hypotheses can be expressed by infinitely many such constraints. This is the case when Theil-estimator-based test used for testing a specified slope (Example 1); when testing variance components; when one marginal distribution is known in the simplicial depth-based test.

Motivated by applications to goodness of fit U-statistics testing, the jackknife empirical likelihood of Jing, *et al.* (2009) is generalized to vector U-statistics and the Wilks theorem is proved. This generalizes Owen's empirical likelihood theorem for a vector mean to a vector U-statistics-based mean and includes the jackknife empirical likelihood of U-statistics with side information as a special case. The results are generalized to allow for the constraints to use estimated criteria functions and for the number of constraints to grow with the sample size. The latter is needed to handle naturally occurring nuisance parameters in semiparametric models.

An excellent overview can be found in the review paper by Chen and Kielegom (2009). The S function *elm* can be used to calculate the empirical likelihood and the codes in S for *elm* can be found in Owen's website. A package called *emplik* which was developed by Mai Zhou can be downloaded at the public URL: *http://cran.r-project.org*.

The rest of the paper is organized as follows: In Section 1.2, the jackknife empirical likelihood for empirical likelihood is introduced with a justification based on the Hoeffding decomposition for U-statistics. In Chapter 2, we give examples which motivate our research. Jackknife empirical likelihood is developed for U-quantiles with finite many constraints and with growing number of constraints in the Theil estimator for testing the slope in a simple linear regression in Example 1; for the Wilcoxon signed rank test about symmetry with a unknown center of symmetry in Example 2; for Kendall's tau and Goodman and Kruskal's Gamma with side information in Example 3; for testing about independence about two categorical outcomes in Example 4; for joint confidence sets in a variance component model in Section 2.2; for the simplicial depth function in Section 2.3, the latter two with finitely many and growing number of constraints. In Chapter 3, we present our main results. In Section 3.1, we introduce the notation, state some results from Peng and Schick (2013c) and prove Lemma 1.2.1 and a useful general theorem. In subsection 3.1.1, we introduce the notation used throughout. In subsection 3.1.2, two general results about the asymptotic behaviors of empirical likelihood are given. In subsection 3.1.3, we prove a general asymptotic result for jackknife empirical likelihood for U-statistics with side information. In Section 3.2, we study the jackknife empirical likelihood when the number of constraints are fixed. We present the Wilks theorem for vector U-statistics. This generalizes Owen's empirical likelihood theorem for vectors. We also derive the asymptotic distributions of the empirical likelihood with side information when the constraint are estimated. In Section 3.3, we discuss the empirical likelihood with random vectors whose dimension may increase with sample size. Our results cover both known and estimated constraints. In Chapter 4, we provide the technical details for our examples. In Chapter 5, we report the simulation results.

1.2 Jackknife empirical likelihood

We will first recall some facts about one-sample U-statistics. Let $(\Omega, \mathscr{A})$ be a measurable space and P be a probability measure on this space. Let Z be a random element taking values in some measurable space $(\mathcal{Z}, \mathscr{S})$ with distribution Q. Typically $\mathcal{Z}$ is a subspace of the real space $\mathcal{R}$ or the high dimensional real space. Let $Z_1, \ldots, Z_n$ be independent and identical copies of Z. Let h be a known measurable function from $\mathcal{Z}^m$ to $\mathcal{R}$ which is argument-symmetric in its m arguments, that is, $h(z_1, \ldots, z_m) = h(z_{\pi_1}, \ldots, z_{\pi_m})$ for every $z_1, \ldots, z_m \in \mathcal{Z}$, where $\pi_1, \ldots, \pi_m$ is an arbitrary permutation of integers $1, \ldots, m$. A U-statistic with kernel h of order m is defined as

$$U_n := U_{nm}(h) = \binom{n}{m}^{-1} \sum_{1 \le i_1 < \ldots < i_m \le n} h(Z_{i_1}, \ldots, Z_{i_m}), \quad n \ge 2.$$

Throughout we assume h is Q^m-square integrable, that is, $h \in L_2(Q^m)$, where $L_2(Q^m) = \left\{ f : \int f^2 \, dQ^m < \infty \right\}$. We shall abbreviate $\theta = E(h) := E(h(Z_1, \ldots, Z_m)) = \int h \, dQ^m$, $P_n f = n^{-1} \sum_{j=1}^n f(Z_j)$ and $Pf = E(f(Z))$. Then U_n is an unbiased estimate of θ. Let $h_m = h$ and $h_c(z_1, \ldots, z_c) = E(h(z_1, \ldots, z_c, Z_{c+1}, \ldots, Z_m))$ for $c = 1, \ldots, m-1$. Then h_c is a version of the conditional expectation, that is,

$$h_c(z_1, \ldots, z_c) = E(h(Z_1, \ldots, Z_m) | Z_1 = z_1, \ldots, Z_c = z_c).$$

Let δ_z be the point mass at $z \in \mathcal{Z}$. We now define

$$h_c^*(z_1, \ldots, z_c) = (\delta_{z_1} - P) \ldots (\delta_{z_c} - P) P^{m-c} h, \quad c = 0, 1, \ldots, m.$$

Let $\tilde{f} = f - Pf$ denote the centered version of an integrable function f. Obviously $h_1^* = \tilde{h}_1$. With this notation the useful Hoeffding decomposition can be stated as

$$U_n - \theta = \sum_{c=1}^{m} \binom{m}{c} U_{nc}(h_c^*). \tag{1.2.1}$$

Let $U_{n-1}^{(-j)}$ denote the U-statistic based on the $n-1$ observations $Z_1, \ldots, Z_{j-1}, Z_{j+1}, \ldots, Z_n$. The jackknife pseudo values of the U-statistic $U_n(h)$ with kernel h are defined as

$$V_{nj}(h) = nU_n(h) - (n-1)U_{n-1}^{(-j)}(h), \quad j = 1, \ldots, n.$$

For ease of notation, we sometimes will drop h and write $V_{nj} = V_{nj}(h)$ when there is no ambiguity. From (1.2.1) it follows

$$V_{nj} = \theta + m\tilde{h}_1(Z_j) + R_{nj}, \quad j = 1, \ldots, n, \tag{1.2.2}$$

where R_{nj} is the remainder given by

$$R_{nj} = \sum_{c=2}^{m} \binom{m}{c} \Big(nU_{nc}(h_c^*) - (n-1)U_{(n-1)c}^{(-j)}(h_c^*) \Big), \quad j = 1, \ldots, n.$$

Using the Hoeffding decomposition (1.2.1) and the orthogonality property of $U_{nc}(h_c^*)$'s, we can prove the following.

Lemma 1.2.1 *The jackknife pseudo values V_{nj} of $U_n(h)$ satisfy*

$$E\big((V_{nj} - \theta - m\tilde{h}_1(Z_j))^2\big) = O(n^{-1}), \quad j = 1, \ldots, n. \tag{1.2.3}$$

For a complete proof please see (3.1.5) and thereafter. Thus from (1.2.3) it immediately follows

$$V_{nj} = \theta + m\tilde{h}_1(Z_j) + O_p(n^{-1/2}), \quad j = 1, \ldots, n. \tag{1.2.4}$$

This shows that each jackknife pseudo value V_{nj} depends asymptotically on Z_j so that $V_{nj}, j = 1, \ldots, n$ are approximately *independent* for large values of n. One of the nice properties of the jackknife pseudo values V_{nj}'s is that they satisfy

$$U_n(h) = \frac{1}{n} \sum_{j=1}^{n} V_{nj}(h). \tag{1.2.5}$$

Thus a U-statistic can be expressed as the sum of approximately independent random variables (the jackknife pseudo values). Furthermore, if π_j is a probability mass placed at Z_j, then approximately the same mass π_j is placed at the jackknife pseudo value V_{nj} for $j = 1, \ldots, n$. Therefore, the likelihood of the pseudo values V_{nj}'s is approximately the product of these π_j's. In view of $E(U_n) = \theta$ and (1.2.5), we are justified to introduce the jackknife empirical likelihood of the U-statistic $U_n(h)$ with side information expressed by $\int \mathbf{g}\, dQ = 0$ as follows:

$$\mathscr{R}_n(h, \mathbf{g}) = \sup \Big\{ \prod_{j=1}^{n} n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \ \sum_{j=1}^{n} \pi_j \big(\tilde{V}_{nj}(h), \mathbf{g}(Z_j)^\top \big)^\top = 0 \Big\}, \tag{1.2.6}$$

where $\mathbf{g}$ is a measurable function from $\mathcal{Z}$ to $\mathcal{R}^r$ and $\mathscr{P}_n$ denotes the closed probability simplex in dimension n,

$$\mathscr{P}_n = \Big\{ \boldsymbol{\pi} = (\pi_1, \dots, \pi_n)^\top \in [0,1]^n : \sum_{i=1}^n \pi_i = 1 \Big\}.$$

Here r is the number of equalities that express the side information, and these equalities are referred to as *constraints* in the literature.

It must be noted that the above definition of jackknife U-statistic with side information covers the case that the side information is expressed by several U-statistics in view of the Hoeffding decompositions for U-statistics with square-integrable kernels. This is indeed the case of vector (multivariate) U-statistics. Specifically, let $h^{(k)}$ be a kernel from $\mathcal{Z}^{m_k}$ to $\mathcal{R}$ for $k = 1, \dots, r$. Let $E\big(U_{nm_k}(h^{(k)})\big) = \theta_k$ and $\tilde{V}_{nj}(h^{(k)}) = V_{nj}(h^{(k)}) - \theta_k$ be the centered jackknife pseudo values of the U-statistic $U_{nm_k}(h^{(k)})$ of order m_k with kernel $h^{(k)}$. Let $\mathbf{h} = (h^{(1)}, \dots, h^{(m_k)})^\top$ and $\tilde{\mathbf{V}}_{nj}(\mathbf{h}) = (\tilde{V}_{nj}(h^{(1)}), \dots, \tilde{V}_{nj}(h^{(m_k)}))^\top$. Based on the above discussion, the jackknife empirical likelihood of a vector U-statistic is justified to be defined by

$$\mathscr{R}_n(\mathbf{h}) = \sup \Big\{ \prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{j=1}^n \pi_j \tilde{\mathbf{V}}_{nj}(\mathbf{h}) = 0 \Big\}. \tag{1.2.7}$$

In nonparametric and semiparametric models, side information can often be expressed by either finitely or infinitely many equalities. Examples of the latter include the commonly used symmetry and independence which are equivalent to infinitely many equalities as illustrated in the examples given in Section 3. This motivates us to allow r to depend on the sample size n, $r = r_n$, and to grow to infinity slowly with n and study the asymptotic behaviors of the empirical likelihood. We are not the first one to study this, see Hjort, McKeague and Van Keilegom (2009), Chen, Peng and Chin (2009) and Peng and Schick (2013c).

2. MOTIVATING EXAMPLES

In this chapter, we give examples that motivated the research in this paper. These include U-quantiles, Theil's test about a specified slope, Wilcoxon signed rank test about the center of symmetry, Kendall's τ and Goodman and Kruskal's Gamma tests about independence, tests about variance components and the simplicial-depth based test, with or without side information.

2.1 Empirical likelihood with side information for several tests

In this section, we shall develop the empirical likelihood theory for several popular tests.

Example 1 Empirical likelihood for U-quantiles. The theory of U-quantile provides a unified treatment of several commonly used statistics, see Arcones (1996). In this example, we shall study the empirical likelihood of U-quantiles with side information.

Recall $(\Omega, \mathscr{A}, P)$ be a probability space and $(\mathcal{Z}, \mathscr{S})$ be a measurable space. Let $Z, Z_1, \ldots, Z_n$ be independent random variables from Ω to $\mathcal{Z}$ with a common unknown distribution Q. Let $\kappa : \mathcal{Z}^m \mapsto \mathcal{R}$ be a measurable function which is argument-symmetric. Associated with κ there induces a distribution function $H(t) = P\big(\kappa(Z_1, \ldots, Z_m) \leq t\big), t \in \mathcal{R}$. A minimum variance unbiased estimator (MVUE) of $H(t)$ is the U-statistic of order m given by

$$H_n(t) := H_{nm}(t) = \binom{n}{m}^{-1} \sum_{1 \leq i_1 < \cdots < i_m \leq n} \mathbf{1}[\kappa(Z_{i_1}, \ldots, Z_{i_m}) \leq t], \quad t \in \mathcal{R},$$

where $\mathbf{1}[A]$ denotes the indicator of a set A. Following Arcones (1996), the κ should be referred to as the kernel (of the U-quantile). As $H(t)$ is a distribution function, its p-th quantile q is well defined by $q = \inf\{t : \ H(t) \geq p\}$ for $p \in [0, 1]$. We are interested in testing the null hypothesis that the p-th quantile q is equal to some specified value $q_0 \in \mathcal{R}$ for a known value p_0 of p, i.e.,

$$H_0 : q = q_0.$$

Let us now mention that the above test provides a unified treatment for several commonly used tests. Below are three examples in which $p = 1/2$. The first example is the often used alternative to the median as a center of symmetry, the Hodges-Lehmann estimator, which is defined as the median of the $\binom{n}{2}$ pairwise averages $2^{-1}(Z_i + Z_j), 1 \leq i < j \leq n$. This is a U-quantile with the kernel given by $\kappa(z_1, z_2) = 2^{-1}(z_1 + z_2)$. In this case, we are testing the hypothesis that the center of symmetry of the underlying distribution is equal to some specified value. The second example is the Gini's mean difference which can be used as a spread measure of a distribution. This corresponds to the U-quantile with the kernel given by $\kappa(z_1, z_2) = |z_1 - z_2|$. In this case, we are testing the hypothesis that the spread of the underlying distribution is equal to some specified value. The third example is the Theil estimator of the slope in a simple linear regression model, which we shall give more details later.

Case 1. *Fixed number of constraints.* Suppose there is available additional information about the underlying distribution expressed by $\int \mathbf{g}\, dQ = 0$ for some measurable function from $\mathcal{Z}$ to $\mathcal{R}^r$ which has finite second moment, i.e. $\int \|\mathbf{g}\|^2 \, dQ < \infty$, where $\|\mathbf{a}\|$ denotes the euclidean norm of a vector $\mathbf{a}$. Here r is the number of constraints which is fixed. Later we will allow it to grow with the sample size. We shall employ the empirical likelihood approach to make use of the additional information. Specifically, we shall study the jackknife empirical likelihood of the U-statistic $H_n(q_0)$ with side information expressed by the mean zero of $\mathbf{g}(Z)$ as follows:

$$\mathscr{R}_n(q_0) = \sup\Big\{ \prod_{j=1}^{n} n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{j=1}^{n} \pi_j (V_{nj}(q_0) - p_0) = 0, \sum_{j=1}^{n} \pi_j \mathbf{g}(Z_j) = 0 \Big\},$$

where $V_{nj}(q_0)$'s are the jackknife pseudo values of the U-statistic $H_n(q_0)$, i.e.,

$$V_{nj}(q_0) = nH_n(q_0) - (n-1)H_{n-1}^{(-j)}(q_0), \quad j = 1, \dots, n,$$

where $H_{n-1}^{(-j)}(q_0)$ denotes the U-statistic based on the $n-1$ observations $Z_1, \dots,$ $Z_{j-1}, Z_{j+1}, \dots, Z_n$.

Often there is available side information about the underlying distribution Q. Here let us just mention two examples: (i) Q has zero median and (ii) Q has zero mean. The former corresponds to $g(z) = \text{sign}(z)$ (assuming $P(Z=0)=0$), while the latter to $g(z) = z$. Here $\text{sign}(x) = \mathbf{1}[x > 0] - \mathbf{1}[x < 0]$ is the sign function of x.

Following the convention of U-statistics theory, let $h_1(z) = P\big(\kappa(z, Z_2, \dots, Z_m) \le q_0\big)$, $z \in \mathcal{Z}$, $p_0 = H(q_0)$ and

$$\mathbf{w} = \big(m\tilde{h}_1, \mathbf{g}^\top\big)^\top, \quad \mathbb{W}(mh_1, \mathbf{g}) = \int \mathbf{w}^{\otimes 2}\, dQ.$$

For a fixed r, it follows from Corollary 3.2.1 below that if the dispersion matrix $\mathbb{W}(mh_1, \mathbf{g})$ is non-singular then $-2\log \mathscr{R}_n(q_0)$ has asymptotically a chi-square distribution with $r+1$ degrees of freedom. In other words,

$$P(-2\log \mathscr{R}_n(q_0) > \chi^2_{1-\alpha}(r+1)) \to \alpha, \quad 0 < \alpha < 1, \tag{2.1.1}$$

where $\chi^2_\beta(r)$ denotes the β-quantile of the chi-square distribution with r degrees of freedom. Thus an asymptotic $1-\alpha$ confidence interval is

$$\left\{q \in \mathcal{R} :\ -2\log \mathscr{R}_n(q) \le \chi^2_{1-\alpha}(r+1)\right\}.$$

It is noteworthy that one of the advantages of empirical-likelihood-based confidence sets is that they enjoy the properties of data-driven shape and internal studentization – no need of estimating standard deviations, while asymptotic-normality-based confidence regions are always symmetric and must estimate standard deviations.

Case 2. *Growing number of constraints.* In many semiparametric models, side information can often be expressed via infinitely many equality constraints. Here we give an example of this by the aid of a simple linear regression model. Let $(X_1, Y_1), \dots, (X_n, Y_n)$ be independent and identically distributed copies of a random vector (X, Y) which satisfies a simple linear regression model,

$$Y = \alpha + \beta X + \epsilon, \tag{2.1.2}$$

where $\alpha, \beta \in \mathcal{R}$ are regression parameters and ϵ is a random error independent of X. We assume that the distribution functions F and G of ϵ and X are continuous.

The null hypothesis of interest is that the slope β of the postulated regression line is some specified value β_0, namely,

$$H_0 : \beta = \beta_0.$$

The test statistic based on the Theil estimator is the U-quantile $H_{n2}(\beta_0)$ with the kernel given by $\kappa((x_1, x_2), (y_1, y_2)) = (y_1 - y_2)/(x_2 - x_2)$ if $x_1 \neq x_2$ and zero otherwise and with $p_0 = H(\beta_0) = 1/2$.

Since X is independent of ϵ, it follows

$$E\big(a_k(X) b_l(\epsilon)\big) = 0, \quad k, l = 1, 2, \dots \tag{2.1.3}$$

for orthonormal bases $a_k, k = 1, 2, \dots$ of $L_{2,0}(G)$ and $b_l, l = 1, 2, \dots$ of $L_{2,0}(F)$, where

$$L_2(R) = \Big\{ f : \int f^2 \, dR < \infty \Big\}$$

and $L_{2,0}(R) = \big\{ f \in L_2(R) : \int f \, dR = 0 \big\}$ for a distribution R. Because F, G are continuous, we take $a_k = \phi_k(G)$ and $b_l = \phi_l(F)$, where

$$\phi_k(t) = \sqrt{2} \cos(k\pi t), \quad t \in [0, 1], k = 1, 2, \dots, \tag{2.1.4}$$

is the usual trigonometric basis of $L_{2,0}(\mathscr{U})$ with $\mathscr{U}$ the uniform distribution on $[0, 1]$. Note that this trigonometric basis is convenient for us to use since it is orthonormal and bounded by $\sqrt{2}$. But F, G are unknown, we estimate the latter by the empirical distribution function

$$\mathbb{G}(x) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}[X_j \le x], \quad x \in \mathcal{R},$$

while the former can be estimated by

$$\mathbb{F}_{\alpha,\beta_0}(t) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}[Y_j - \alpha - \beta_0 X_j \le t], \quad t, \alpha \in \mathcal{R}.$$

Let us point out an interesting fact that it is without loss of generality to assume the intercept $\alpha = 0$ in testing the null hypothesis about the slope using the present approach. This is due to the identity

$$\mathbb{F}_{\alpha,\beta_0}(Y_i - \alpha - \beta_0 X_i) = \mathbb{F}_{0,\beta_0}(Y_i - \beta_0 X_i), \quad i = 1, \dots, n,$$

and the fact that all we need in computing the empirical likelihood (i.e. $\mathscr{R}_n(\beta_0)$ below) are the values on the left hand side of the above identity for $i = 1, \ldots, n$. This is not surprising in view of the fact that the Theil estimator of the slope is the median of the slopes $(Y_i - Y_j)/(X_i - X_j), 1 \le i < j \le n$ which clearly does not rely on the value of the intercept. Indeed, Peng, *et al.* (2008) used this fact in their investigation of the asymptotic properties of the Theil-Sen estimator. As a result, we take $\alpha = 0$ and estimate F by the empirical distribution function $\mathbb{F}_{\beta_0} = \mathbb{F}_{0,\beta_0}$ based on the observations $\epsilon_j = Y_j - \beta_0 X_j$, i.e.,

$$\mathbb{F}_{\beta_0}(t) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}[Y_j - \beta_0 X_j \le t], \quad t \in \mathcal{R}. \tag{2.1.5}$$

The preceding consideration motivates us to use the first r_n^2 equations in (2.1.3) to construct the jackknife empirical likelihood with side information as follows:

$$\begin{aligned}\mathscr{R}_n(\beta_0) = \sup\Big\{ \prod_{j=1}^{n} n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{j=1}^{n} \pi_j V_{nj}(\beta_0) - 1/2 = 0, \\ \sum_{j=1}^{n} \pi_j \phi_k\big(\mathbb{G}(X_j)\big)\phi_l\big(\mathbb{F}_{\beta_0}(Y_j - \beta_0 X_j)\big) = 0, \quad k, l = 1, \ldots, r_n \Big\},\end{aligned}$$

where $V_{nj}(\beta_0)$'s are the jackknife pseudo values of the U-statistic $H_{n2}(\beta_0)$. We shall allow r_n to grow slowly to infinity with the sample n such that r_n^6/n tends to zero. Then under the null hypothesis one has

$$\frac{-2\log \mathscr{R}_n(\beta_0) - r_n^2 - 1}{\sqrt{2(r_n^2 + 1)}} \Longrightarrow \mathscr{N}(0, 1), \quad n \to \infty. \tag{2.1.6}$$

The details of the proof of (2.1.6) can be found in the last section. This shows that under the null hypothesis $-2\log \mathscr{R}_n(\beta_0)$ is approximately chi-square-distributed with $r_n^2 + 1$ degrees of freedom. Thus for $0 < \alpha < 1$,

$$P\big(-2\log \mathscr{R}_n(\beta_0) > \chi^2_{1-\alpha}(r_n^2 + 1)\big) \xrightarrow{P} \alpha.$$

This exhibits that the test $\mathbf{1}[-2\log \mathscr{R}_n(\beta_0) > \chi^2_{1-\alpha}(r_n^2 + 1)]$ has asymptotic size α. Also, an asymptotic $1 - \alpha$ confidence interval for β_0 is given by

$$\big\{\beta \in \mathcal{R} :\ -2\log \mathscr{R}_n(\beta) \le \chi^2_{1-\alpha}(r_n^2 + 1)\big\}.$$

Example 2 Wilcoxon signed rank test. Let X be the difference of post-treatment and pre-treatment of a subject, and $X_1, \ldots, X_n$ be independent copies of X. Let us assume

(W1) The distribution function F of X has a density f which is continuous and uniformly bounded (by B say) and bounded away from zero in a neighborhood of the center θ of symmetry of X.

The null hypothesis of interest asserts that the distribution F is symmetrically distributed about θ, i.e., $F(\theta + t) + F(\theta - t) = 1$ for every $t \in \mathcal{R}$. If θ were known, one popular test is the Wilcoxon signed rank statistic $W_n^+ = \sum_{j=1}^n R_j \mathbf{1}[X_j - \theta > 0]$, where R_j denotes the rank of $|X_j - \theta|, j = 1, \ldots, n$. Apparently the Wilcoxon test is asymptotically equivalent to the U-statistic

$$U_n(h_\theta) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n}^{n} \mathbf{1}[X_i + X_j - 2\theta > 0].$$

with kernel $h_\theta(x_1, x_2) = \mathbf{1}[(x_1 + x_2)/2 > \theta]$. The null hypothesis can now be expressed as

$$H_0 : P(X_1 + X_2 - 2\theta > 0) = 1/2.$$

In real-life data, the center θ of symmetry is usually unknown and must be estimated. One possible estimator is the Hodges-Lehman median $\tilde{\theta}_n$. Unfortunately, this estimator leads to $v \equiv 0$, an inappropriate v in (3.2.4) of Theorem 3.2.2. To see this, we shall use a result in the proof of Theorem 1 of Arcones (1996). Under the assumption (W1), it is not difficult to verify the conditions in his Theorem 1 are met. Thus we can apply his inequality (2.5) to get

$$U_n(\tilde{\theta}_n) - 1/2 = o_p(n^{-1}),$$

where throughout this example we denote $U_n(\theta) = U_n(h_\theta)$. This equality immediately results in $v \equiv 0$. As a matter of fact, Peng and Schick (2013b) also observed a similar issue in constructing residual-based inference about a quantile.

Let us now estimate θ by the usual median $\hat{\theta}_n$. This yields the U-statistic

$$U_n(\hat{\theta}_n) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n}^{n} \mathbf{1}[X_i + X_j - 2\hat{\theta}_n > 0].$$

The corresponding jackknife empirical likelihood is as follows:

$$\mathscr{R}_n(\hat{\theta}_n) = \sup\Big\{ \prod_{j=1}^{n} n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{j=1}^{n} \pi_j \tilde{V}_{nj}(\hat{\theta}_n) = 0 \Big\},$$

where $V_{nj}(\hat{\theta}_n)$'s are the jackknife pseudo values of the U-statistic $U_n(\hat{\theta}_n)$.

To derive the asymptotic behaviors, let us denote the survival function by $\bar{F} = 1 - F$. Then under (W1), $\bar{F}$ satisfies

$$\bar{F}(t) = \bar{F}(t_0) - f(t_0)(t - t_0) + o(t - t_0), \quad \text{as} \quad t \to t_0, \tag{2.1.7}$$

for any $t_0 \in \mathcal{R}$. Let F_2 be the distribution of $X_1 + X_2$. Then F_2 has a density f_2 given by

$$f_2(y) = \int_{-\infty}^{\infty} f(y - x) f(x)\, dx, \quad y \in \mathcal{R}. \tag{2.1.8}$$

Clearly f_2 is also continuous and uniformly bounded, hence F_2 satisfies

$$\bar{F}_2(t) = \bar{F}_2(t_0) - f_2(t_0)(t - t_0) + o(t - t_0), \quad \text{as} \quad t \to t_0, \tag{2.1.9}$$

for any $t_0 \in \mathcal{R}$. Let

$$\varsigma^2 = 4E\big((F(2\theta_0 - X) - 1/2)^2\big)$$

and

$$\sigma^2 = \varsigma^2 + \frac{f_2(\theta_0)^2}{4f(\theta_0)^2} + \frac{2f_2(\theta_0)}{f(\theta_0)}\Big(\frac{1}{4} - E\big(F(2\theta_0 - X)\mathbf{1}[X \le \theta_0]\big)\Big).$$

We further assume

(W2) The density f_2 of $X_1 + X_2$ satisfies $f_2(\theta_0) > 0$. Moreover, ς^2 is positive.

Obviously $\varsigma^2 > 0$ implies $\sigma^2 > 0$ by Cauchy inequality. Then under the null hypothesis one has

$$-2\log \mathscr{R}_n(\hat{\theta}_n) \Longrightarrow \sigma^2 \varsigma^{-2} \chi^2(1). \tag{2.1.10}$$

Observe that since $\sigma^2/\varsigma^2 > 1$ it follows that the above limiting distribution has a larger variance than the chi-square distribution with one degree of freedom (i.e. $\chi^2(1)$). This extra variation is clearly resulted from the median estimator $\hat{\theta}_n$ of θ. The details of the proof of (2.1.10) can be found in the last section.

Example 3 KENDALL'S TAU AND GOODMAN & KRUSKAL'S GAMMA. Let $\mathbf{Z}_j = (X_j, Y_j), j = 1, \ldots, n$ be independent and identically distributed copies of a random vector $\mathbf{Z} = (X, Y)$.

Kendall's τ. We are interested in testing the null hypothesis that the random variables X and Y are *independent*. One can use Kendall's τ as test statistic given by

$$U_n(h_\tau) = \binom{n}{2}^{-1} \sum_{1 \le i < j \le n} \mathbf{1}[(X_i - X_j)(Y_i - Y_j) > 0].$$

This is a U-statistic with the kernel $h_\tau(z_1, z_2) = \mathbf{1}[(x_1 - x_2)(y_1 - y_2) > 0]$. Suppose the distribution functions of X and Y are continuous. Then under the null hypothesis of independence one has

$$P(C) - 1/2 = 0, \tag{2.1.11}$$

where $C = \{(X_1 - X_2)(Y_1 - Y_2) > 0\}$ denotes the event of concordance of pairs (X_1, X_2) and (Y_1, Y_2).

Goodman and Kruskal's Gamma. Discrete or categorical data are widely used in various areas of science. Unlike continuous response, there is usually a sizable number of tied observations in such data. Goodman and Kruskal's Gamma is useful in investigating such data. It is defined by

$$\gamma = \big(P(C) + P(D)\big)^{-1}(P(C) - P(D)).$$

where $D = \{(X_1 - X_2)(Y_1 - Y_2) < 0\}$ denotes the events of discordance of pairs (X_1, X_2) and (Y_1, Y_2). Equivalently we can express the preceding display as

$$(1 - \gamma)P(C) - (1 + \gamma)P(D) = 0. \tag{2.1.12}$$

The above left-hand-side expression can be estimated by a U-statistic

$$U_n(h_\gamma) = \binom{n}{2} \sum_{1 \le i < j \le n} h_\gamma(\mathbf{Z}_i, \mathbf{Z}_j), \tag{2.1.13}$$

where h_γ is the kernel given by

$$h_\gamma(\mathbf{z}_1, \mathbf{z}_2) = (1-\gamma)\mathbf{1}[(x_1 - x_2)(y_1 - y_2) > 0] - (1+\gamma)\mathbf{1}[(x_1 - x_2)(y_1 - y_2) < 0].$$

Suppose we have available additional information expressed by $\int \mathbf{g}\, dQ = 0$ for some measurable function $\mathbf{g}$ from $\mathcal{Z}$ to $\mathcal{R}^r$ such that $\int \|\mathbf{g}\|^2\, dQ$ is finite. For example, we have the partial information that the marginal distributions of X and Y have known medians m_{10} and m_{20} respectively. Then we take $g_1(x) = \mathbf{1}[x \le m_{10}] - 1/2$ and $g_2(y) = \mathbf{1}[y \le m_{20}] - 1/2$ and $\mathbf{g}(\mathbf{z}) = (g_1(x), g_2(y))^\top$ such that $\int \mathbf{g}\, dQ = 0$.

We now employ the empirical likelihood to incorporate additional information. The above two cases can be treated in one formulation as follows: the jackknife empirical likelihood of the U-statistic with side information is given by

$$\mathscr{R}_n(h, \mathbf{g}) = \sup\Big\{ \prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{j=1}^n \pi_j \tilde{V}_{nj}(h) = 0, \sum_{j=1}^n \pi_j \mathbf{g}(\mathbf{Z}_j) = 0 \Big\},$$

where h denotes either h_τ or h_γ, $V_{nj}(h)$ is the jackknife pseudo values of the U-statistic $U_n(h)$ and $\tilde{V}_{nj}(h) = V_{nj}(h_\tau) - \mu$ with $\mu = 1/2$ or 0 according to $h = h_\tau$ or h_γ respectively.

For an $s \times s$ matrix $\mathbb{A}$, $s \times r$ matrix $\mathbb{C}$ and a $r \times r$ matrix $\mathbb{M}$, we define the matrix function $\mathscr{W} : \mathcal{R}^s \times \mathcal{R}^{s\times r} \times \mathcal{R}^{r \times r} \to \mathcal{R}^{(s+r)\times(s+r)}$ by

$$\mathscr{W}(\mathbb{A}, \mathbb{C}, \mathbb{M}) = \begin{pmatrix} \mathbb{A} & \mathbb{C} \\ \mathbb{C}^\top & \mathbb{M} \end{pmatrix}. \tag{2.1.14}$$

Let us denote $\theta = \tau$ or γ corresponding $h = h_\tau$ or h_γ respectively, and set $U_n(\theta, \mathbf{g}) = U_n(h, \mathbf{g})$. Denote θ_0 the true value of parameter. Then by Corollary 3.2.1, under the null hypothesis one has

$$P\big(-2 \log \mathscr{R}_n(\theta_0, \mathbf{g}) > \chi^2_{1-\alpha}(r+1) \big) \to \alpha, \quad 0 < \alpha < 1,$$

provided that the dispersion matrix $\mathscr{W}(\mathrm{Var}(2h_1(\mathbf{Z})), \mathbf{c}, \mathbb{W}(\mathbf{g}))$ is non-singular, where $h_1(\mathbf{z}) = E(h(\mathbf{z}, \mathbf{Z}))$ and

$$\frac{1}{n}\sum_{j=1}^n \tilde{h}_1(\mathbf{Z}_j)\mathbf{g}(\mathbf{Z}_j) \xrightarrow{P} \mathbf{c}, \quad \frac{1}{n}\sum_{j=1}^n \mathbf{g}(\mathbf{Z}_j)^{\otimes 2} \xrightarrow{P} \mathbb{W}(\mathbf{g}).$$

Example 4 TESTING INDEPENDENCE BETWEEN TWO CATEGORICAL OUTCOMES. Let (U, V) be a bivariate categorical random vector whose marginals have K, L levels indexed by r_k, s_l respectively. Let $(U_i, V_i), i = 1, \dots, n$ be independent and identical copies of (U, V). Based on a random sample we are interested in testing the independence of U and V. Chi-squares or Fisher's exact tests are often used to test such a null hypothesis. In this example, we use the jackknife empirical likelihood for multivariate U-statistics to give an asymptotic test based on the multivariate U-statistic described in Example 6, page 260 of Kowalski and Tu (2008). To this end, set $\mathbf{Z}_i = (\mathbf{X}_i^\top, \mathbf{Y}_i^\top)^\top$, where

$$\mathbf{X}_i = (\mathbf{1}[U_i = r_1], \dots, \mathbf{1}[U_i = r_K])^\top, \quad \mathbf{Y}_i = (\mathbf{1}[V_i = s_1], \dots, \mathbf{1}[V_i = s_L])^\top.$$

Independence implies the components of $\mathbf{X}_i$ and $\mathbf{Y}_j$ satisfy

$$E\left((X_{ik} - X_{jk})(Y_{il} - Y_{jl})\right) = 0, \quad k = 1, \dots, K, l = 1, \dots, L.$$

This suggests us to look at the jackknife empirical likelihood of multivariate U-statistic,

$$\mathscr{R}_n = \sup\left\{\prod_{j=1}^{n} n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{j=1}^{n} \pi_j \mathbf{V}_{nj} = 0\right\},$$

where $\mathbf{V}_{nj}$ is the jackknife pseudo values of the multivariate U-statistic $U_n(\mathbf{h})$ of order 2 with the $(K-1)(L-1)$-dimensional kernel $\mathbf{h}$ given by

$$\mathbf{h} = (h^{11}, \dots, h^{1(K-1)}, h^{21}, \dots, h^{(K-1)(L-1)})^\top,$$

where $h^{kl}(\mathbf{Z}_i, \mathbf{Z}_j) = 2^{-1}(X_{ik} - X_{jk})(Y_{il} - Y_{jl})$. Here since each of the two marginal probabilities sums up to one, there are only $(K-1)(L-1)$ cell probabilities. Let $\mathbf{h}_1(\mathbf{z}) = E(\mathbf{h}(\mathbf{z}, \mathbf{Z}_2)), \mathbf{z} \in \mathcal{R}^2$. Then by Theorem 3.2.1,

$$-2 \log \mathscr{R}_n \Longrightarrow \chi^2((K-1)(L-1)).$$

provided that the dispersion matrix $\mathrm{Var}(2\mathbf{h}_1(\mathbf{Z}_1))$ is non-singular.

We show next that a necessary and sufficient condition for the non-singularity of Σ is that all the marginal probabilities are nonzero, i.e.,

$$p_{k\cdot} \neq 0, k = 1, \dots, K \quad \text{and} \quad p_{\cdot l} \neq 0, l = 1, \dots, L. \qquad (2.1.15)$$

To show this, let us first calculate Σ. Let $\mathbf{p} = E(\mathbf{X}_1\mathbf{Y}_1^\top)$. Then $p_{kl} = P(U_1 = r_k, V_1 = s_l)$ and Then $P(U_1 = r_k) = E(X_{1k}) = p_{k\cdot}$ and $P(V_1 = s_l) = E(Y_{1l}) = p_{\cdot l}$. Hence

$$\begin{aligned} h_1^{kl}(\mathbf{z}_1) &= E(h^{kl}(\mathbf{z}_1, \mathbf{Z}_2)|\mathbf{Z}_1 = \mathbf{z}_1) = 2^{-1}E((x_{1k} - X_{2k})(y_{1l} - Y_{2l})|\mathbf{Z}_1 = \mathbf{z}_1) \\ &= 2^{-1}(x_{1k} - p_{k\cdot})(y_{1l} - p_{\cdot l}) + 2^{-1}\delta_{kl}, \quad \mathbf{z}_1 = (\mathbf{x}_1^\top, \mathbf{y}_1^\top)^\top, \end{aligned}$$

and the centered version is

$$\tilde{h}_1^{kl}(\mathbf{z}_1) = 2^{-1}\big((x_{1k} - p_{k\cdot})(y_{1l} - p_{\cdot l}) - \delta_{kl}\big).$$

Let $\boldsymbol{\alpha} = (p_{1\cdot}, \ldots, p_{(K-1)\cdot})^\top$ and $\boldsymbol{\beta} = (p_{\cdot 1}, \ldots, p_{\cdot (L-1)})^\top$. Then $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are the marginal distributions of U_1 and V_1 respectively. Under the null, $\delta_{kl} = 0$,

$$\mathrm{Cov}(X_{1k}, X_{1k'}) = -\alpha_k\alpha_{k'}, \quad \mathrm{Cov}(Y_{1l}, Y_{1l'}) = -\beta_l\beta_{l'}, \quad k \neq k', l \neq l'$$

and

$$\mathrm{Var}(X_{1k}) = \alpha_k - \alpha_k^2, \quad \mathrm{Var}(Y_{1l}) = \beta_l - \beta_l^2.$$

Let

$$\mathbf{A} = \mathrm{Diag}(\boldsymbol{\alpha}) - \boldsymbol{\alpha}^{\otimes 2}, \quad \mathbf{B} = \mathrm{Diag}(\boldsymbol{\beta}) - \boldsymbol{\beta}^{\otimes 2}.$$

Since

$$E\big(\tilde{h}_1^{kl}(\mathbf{Z}_1)\tilde{h}_1^{k'l'}(\mathbf{Z}_1)\big) = 4^{-1}\mathrm{Cov}(X_{1k}, X_{1k'})\mathrm{Cov}(Y_{1l}, Y_{1l'}),$$

it follows

$$\Sigma = \mathrm{Var}(2\mathbf{h}_1(\mathbf{Z}_1)) = 4E\big(\tilde{\mathbf{h}}_1(\mathbf{Z}_1)^{\otimes 2}\big) = \mathbf{A} \otimes \mathbf{B}.$$

It is known that for square matrices $\mathbf{C}$ and $\mathbf{D}$ of orders c and d respectively,

$$|\mathbf{C} \otimes \mathbf{D}| = |\mathbf{C}|^c|\mathbf{D}|^d.$$

Hence in view of $|\Sigma| = |\mathbf{A}|^{K-1}|\mathbf{B}|^{L-1}$ we see that Σ is non-singular if and only if both $\mathbf{A}$ and $\mathbf{B}$ are non-singular. This is equivalent to (2.1.15). To see this, it suffices to show

$$|\mathbf{A}| = \prod_{k=1}^{K} \alpha_k. \tag{2.1.16}$$

where $\alpha_K = p_{K\cdot}$. The same also holds for $\mathbf{B}$. We verify this for $K = 5$ and the general case can be proved by mathematical induction. Note first that $\alpha_K = 1 - \sum_{k=1}^{K-1} \alpha_k$ and $|\mathbf{A}| = \prod_{k=1}^{K-1} \alpha_k D$, where

$$D = \begin{vmatrix} 1-\alpha_1 & -\alpha_2 & -\alpha_3 & -\alpha_4 \\ -\alpha_1 & 1-\alpha_2 & -\alpha_3 & -\alpha_4 \\ -\alpha_1 & -\alpha_2 & 1-\alpha_3 & -\alpha_4 \\ -\alpha_1 & -\alpha_2 & -\alpha_3 & 1-\alpha_4 \end{vmatrix} = \begin{vmatrix} 1-\alpha_1 & -\alpha_2 & -\alpha_3 & -\alpha_4 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{vmatrix}.$$

Using Laplace's formula to express the determinant of a matrix in terms of its minors for the last row, we find

$$D = (-1)(-1)^{1+K-1}(-\alpha_{K-1})(-1)^{1+K-2} + 1 - \sum_{k=1}^{K-2} \alpha_k = 1 - \sum_{k=1}^{K-1} \alpha_k = \alpha_K,$$

so that we conclude (2.1.16) for $K = 5$.

2.2 Confidence sets for variance components

In a balanced one-way random effects model, the response Y_{ij}, random effect u_i and random error ϵ_{ij} satisfy the structural relationship,

$$Y_{ij} = \mu + u_i + \epsilon_{ij}, \quad i = 1, \ldots, n,\ j = 1, \ldots, J(J \geq 2), \tag{2.2.1}$$

where μ is the mean response, the ϵ_{ij}'s are independent and identically distributed with mean zero, variance $\sigma_\epsilon^2 = \mathrm{Var}(\epsilon_{ij})$ and finite fourth moment, the u_i's are independent and identically distributed with mean zero, variance $\sigma_u^2 = \mathrm{Var}(u_j)$ and finite fourth moment, and ϵ_{ij}'s and u_i's are independent. Note that in the literature σ_u^2 and σ_ϵ^2 are termed as the between- and within- treatment variance components respectively.

The commonly used tests and confidence sets for the variances heavily depend on the assumption of normality of the model. Here we employ the empirical likelihood approach to give confidence sets for the variances. It is well known that the empirical likelihood approach has the advantage in constructing confidence regions without requiring an estimate of the standard deviation and the shape of confidence sets is to be data-driven, whereas asymptotic-normality-based confidence regions need estimate the standard deviation and are always symmetric. Moreover, the empirical likelihood approach allows convenient incorporation of side information.

Suppose there is available some additional information about the model which is expressed through the equality $E(\mathbf{g}(\varepsilon)) = 0$, where ε is an i.i.d. copy of $u_i + \epsilon_{i\cdot}$ and $\mathbf{g}$ is a measurable function from $\mathcal{R}$ to $\mathcal{R}^r$. For instance, $g(\varepsilon) = \varepsilon$ and $g(\varepsilon) = \text{sign}(\varepsilon)$, the former expresses that ε has mean zero while the latter indicates that ε has median zero under the assumption that ε is a continuous random variable. Averaging out the j in (2.2.1) now yields

$$Y_{i\cdot} - \mu = u_i + \epsilon_{i\cdot} = \varepsilon_i, \quad i = 1, \dots, n.$$

Let $\hat{\mu}$ be an estimator of μ. Here we take the grand mean

$$\hat{\mu} = Y_{\cdot\cdot} = (nJ)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{J} Y_{ij}.$$

Thus we shall work with the estimated residuals $\hat{\varepsilon}_i = Y_{i\cdot} - \hat{\mu}$.

It is well known that if a distribution F of a random variable X has a finite fourth moment then the minimum variance unbiased estimator (MVUE) of the variance $\text{Var}(X)$ of the distribution F is the U-statistic of order two with the kernel given by $2^{-1}(X_1 - X_2)^2$, where X_1, X_2 are i.i.d. copies of X. For more general discussion, the reader is referred to Heffernan (1997). We shall exploit the MVUE's of variances in our forthcoming investigation.

(i) Joint confidence sets for both variances. Following Arvesen (1969), put

$$\mathbf{X}_i = \begin{pmatrix} Y_{i\cdot} \\ (J-1)^{-1} \sum_{j=1}^{J} (Y_{ij} - Y_{i\cdot})^2 \end{pmatrix}, \quad i = 1, \dots, n, \tag{2.2.2}$$

where $A_{i\cdot} = J^{-1} \sum_{j=1}^{J} A_{ij}$ denotes the average of A_{ij} over j. Clearly $\mathbf{X}_1, \dots, \mathbf{X}_n$ are independent and identically distributed. Set $\mathbf{h} = (h^{(1)}, h^{(2)})^\top$, where, with $\kappa(\mathbf{X}_i) = (J-1)^{-1} \sum_{j=1}^{J} (Y_{ij} - Y_{i\cdot})^2$,

$$h^{(1)}(\mathbf{X}_i, \mathbf{X}_{i'}) = 2^{-1}(\kappa(\mathbf{X}_i) + \kappa(\mathbf{X}_{i'})), \quad h^{(2)}(\mathbf{X}_i, \mathbf{X}_{i'}) = 2^{-1}(Y_{i\cdot} - Y_{i'\cdot})^2.$$

Then one readily calculates $E(h^{(1)}(\mathbf{X}_1, \mathbf{X}_2)) = \sigma_\epsilon^2$ and $E(h^{(2)}(\mathbf{X}_1, \mathbf{X}_2)) = \sigma^2 := \sigma_u^2 + J^{-1}\sigma_\epsilon^2$. Therefore the vector U-statistic

$$\mathbf{U}_n(\mathbf{h}) = (U_n(h^{(1)}), U_n(h^{(2)}))^\top$$

is an unbiased estimators of $\boldsymbol{\theta} = (\sigma_\epsilon^2, \sigma^2)^\top$.

Joint confidence sets. To construct a confidence set for $\boldsymbol{\theta}$, we employ the jackknife empirical likelihood for the vector U-statistic $\mathbf{U}_n(\mathbf{h})$ as follows:

$$\mathscr{R}_n(\boldsymbol{\theta}) = \sup\Big\{\prod_{i=1}^n n\pi_i : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{i=1}^n \pi_i(\mathbf{V}_{ni}(\mathbf{h}) - \boldsymbol{\theta}) = 0\Big\},\ \boldsymbol{\theta} \in \mathcal{R}^+ \times \mathcal{R}^+,$$

where $\mathbf{V}_{ni}(\mathbf{h}) = \big(V_{ni}(h^{(1)}), V_{ni}(h^{(2)})\big)^\top$ is the vector whose components are the jackknife pseudo values of the U-statistics $U_n(h^{(1)})$ and $U_n(h^{(2)})$ respectively. Here $\mathcal{R}^+ = (0, \infty)$. By Theorem 3.2.1 below, if the dispersion matrix $\mathrm{Var}(\mathbf{h}_1(\mathbf{X}))$ is non-singular then

$$-2\log\mathscr{R}_n(\boldsymbol{\theta}_0) \Longrightarrow \chi^2(2),$$

where $\boldsymbol{\theta}_0 = (\sigma_{\epsilon 0}^2, \sigma_0^2)^\top \in \mathcal{R}^+ \times \mathcal{R}^+$ denotes the true value of parameter. Thus

$$P\big(-2\log\mathscr{R}_n(\boldsymbol{\theta}_0) > \chi^2_{1-\alpha}(2)\big) \to \alpha, \quad 0 < \alpha < 1. \tag{2.2.3}$$

An asymptotic $1-\alpha$ confidence set is

$$\big\{\boldsymbol{\theta} \in \mathcal{R}^+ \times \mathcal{R}^+ :\ -2\log\mathscr{R}_n(\boldsymbol{\theta}) \le \chi^2_{1-\alpha}(2)\big\}.$$

It is noteworthy that a confidence set for $\vartheta = (\sigma_\epsilon^2, \sigma_u^2)^\top$ can be obtained by the transformation $\vartheta_1 = \theta_1, \vartheta_2 = \theta_2 - \theta_1/J$. Also, a confidence set for σ_u^2 can be obtained by $J \to \infty$.

Let us now take a close look at the dispersion matrix $\mathrm{Var}(\mathbf{h}_1(\mathbf{X}))$. Note first

$$h_1^{(1)}(\mathbf{x}_1) = E(h_1^{(1)}(\mathbf{x}_1, \mathbf{X}_2)) = 2^{-1}(\kappa(\mathbf{x}_1) + \sigma_\epsilon^2),$$
$$h_1^{(2)}(\mathbf{x}_1) = E(h_1^{(2)}(\mathbf{x}_1, \mathbf{X}_2)) = 2^{-1}((Y_{1\cdot} - \mu)^2 + \sigma^2).$$

Let $\mu_{\epsilon 4} = E(\epsilon_{11}^4)$. Then it is not difficult to calculate

$$\mathrm{Var}\left(h_1^{(1)}(\mathbf{X}_1)\right) = J^{-1}(\mu_{\epsilon 4} - \sigma_\epsilon^4) + O(J^{-2}),$$
$$\mathrm{Var}\left(h_1^{(2)}(\mathbf{X}_1)\right) = E(\varepsilon^4) - \sigma_\varepsilon^4, \quad \sigma_\varepsilon = \sigma,$$
$$\mathrm{Cov}\big(h_1^{(1)}(\mathbf{X}_1), h_1^{(2)}(\mathbf{X}_1)\big) = J^{-2}\mu_{\epsilon 4} - 6\sigma_\epsilon^4/(J^2(J-1)) = O(J^{-2}).$$

Since $\mu_{\epsilon 4} - \sigma_\epsilon^4 > 0$ and $E(\varepsilon^4) - \sigma_\varepsilon^4 > 0$, it follows that the matrix $\mathrm{Var}(\mathbf{h}_1(\mathbf{X}_1))$ is non-singular at least for large values of J.

Confidence sets with side information. When side information is available, to construct a confidence set for $\boldsymbol{\theta}$, we employ the jackknife empirical likelihood for the vector U-statistic $\mathbf{U}_n(\mathbf{h})$ with side information as follows:

$$\mathscr{R}_n(\boldsymbol{\theta}, \mathbf{g}) = \sup\Big\{ \prod_{i=1}^n n\pi_i : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{i=1}^n \pi_i(\mathbf{V}_{ni}(\mathbf{h}) - \boldsymbol{\theta}) = 0, \\ \sum_{i=1}^n \pi_i \mathbf{g}(\hat{\varepsilon}_i) = 0 \Big\}, \quad \boldsymbol{\theta} \in \mathcal{R}^+ \times \mathcal{R}^+.$$

Let us consider that the additional information about the underlying distribution is that the median of ε is zero, so that $E(g(\varepsilon)) = 0$ for $g(t) = \mathrm{sign}(t)$. Assume that the distribution F of ε has a bounded and continuous density f. Let $\mathbf{C} = 2E\big(\mathbf{h}_1(\mathbf{X}_1)\mathrm{sign}(\varepsilon)\big)$. Note that $\mathrm{Var}(g(\varepsilon)) = 1$. Suppose $\mathscr{W} = \mathscr{W}(\mathrm{Var}(2\mathbf{h}_1(\mathbf{X}_1)), \mathbf{C}, 1)$ is non-singular. Then

$$-2\log \mathscr{R}_n(\boldsymbol{\theta}_0, \mathrm{sign}) \Longrightarrow \mathbf{Z}^\top \mathscr{W}^{-1}\mathbf{Z}, \tag{2.2.4}$$

where $\mathbf{Z}$ is normally distributed with vector mean zero and variance-covariance matrix $\mathrm{Var}((2\mathbf{h}_1(\mathbf{X}_1)^\top, u(\varepsilon))^\top)$ with $u(t) = \mathrm{sign}(t) - f(0)t$. Justification of (2.2.4) is given in the last section.

(ii) Joint confidence sets for the mean and variance component. In a balanced one-way random effects model, we are interested in constructing joint confidence set for $\boldsymbol{\theta} = (\mu, \sigma_u^2)^\top$. Let us now motivate a U-statistic as a test statistic, see also Nobre, *et al.* (2008). We shall exploit MVUE's and "averaging out" j gives

$$U_{ii'} = \binom{J}{2}^{-1} \sum_{1 \le j < j' \le J} 2^{-1}(Y_{ij} - Y_{i'j'})^2, \quad i, i' = 1, \ldots, n.$$

For the between-treatment, $i \neq i'$, so $E((u_i - u_{i'})(\epsilon_{ij} - \epsilon_{i'j'})) = 0$, hence

$$E(U_{ii'}) = E\big(2^{-1}(u_i - u_{i'})^2 + 2^{-1}(\epsilon_{ij} - \epsilon_{i'j'})^2\big) = \sigma^2 = \sigma_u^2 + \sigma_\epsilon^2, \tag{2.2.5}$$

whereas for the within-treatment, $i = i'$, thus

$$E(U_{ii}) = E(2^{-1}(\epsilon_{i1} - \epsilon_{i2})^2) = \sigma_\epsilon^2.$$

Let $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iJ})^\top$ denote the observation vector in the i-th treatment. Obviously $\mathbf{Y}_1, \dots, \mathbf{Y}_n$ are independent and identically distributed. Note that $U_{ii'} - 2^{-1}(U_{ii}+U_{i'i'})$ is a function of $\mathbf{Y}_i$ and $\mathbf{Y}_{i'}$ only, say $h_u(\mathbf{Y}_i, \mathbf{Y}_{i'})$. Clearly $E(h_u(\mathbf{Y}_i, \mathbf{Y}_{i'})) = \sigma_u^2$ for every pair (i, i') of subject indices with $i \neq i'$, so every such $h_u(\mathbf{Y}_i, \mathbf{Y}_{i'})$ is an unbiased estimator of σ_u^2. Since $h_u(\mathbf{y}_1, \mathbf{y}_2)$ is not argument-symmetric, we symmetrize it to get the argument-symmetric kernel $h(\mathbf{y}_1, \mathbf{y}_2) = 2^{-1}(h_u(\mathbf{y}_1, \mathbf{y}_2) + h_u(\mathbf{y}_2, \mathbf{y}_1)), \mathbf{y}_1, \mathbf{y}_2 \in \mathcal{R}^J$. Thus an unbiased estimator of σ_u^2 based on all the observations is the U-statistic with the kernel h given by

$$U_n(h) = \binom{n}{2}^{-1} \sum_{1 \le i < i' \le n} h(\mathbf{Y}_i, \mathbf{Y}_{i'}).$$

Suppose there is available additional information about the model, for instance, that ε as an i.i.d. copy of $\varepsilon_i = u_i + \epsilon_{i\cdot}$ is *symmetric* about zero. With this as side information we now construct an empirical-likelihood-based confidence set for $\boldsymbol{\theta}$. To this end, let F denote the distribution function of ε, and $L_{2,0}(F, \text{odd})$ be the subspace of $L_{2,0}(F)$ consisting of the odd functions. Assume F is continuous. Symmetry of ε about zero implies

$$E(a_k(\varepsilon)) = E(a_k(Y_{1\cdot} - \mu)) = 0, \quad k = 1, 2, \dots, \tag{2.2.6}$$

where a_k's is an orthonormal basis of $L_2(F, \text{odd})$ and μ denotes the true value of parameter. Since ε and $-\varepsilon$ have an identical distribution, it follows

$$-\big(2F(-t) - 1\big) = 1 - 2P(\varepsilon \le -t) = 1 - 2P(\varepsilon \ge t) = 1 - 2(1 - F(t)) = 2F(t) - 1.$$

This shows that $2F(t) - 1$ is an odd function. Note that $\psi_k(t) = \sin(k\pi t), t \in [-1, 1], k = 1, 2, \dots$ is an orthonormal basis of $L_{2,0}(\mathcal{U}, \text{odd})$ (the square-integrable odd functions with respect to the uniform measure $\mathcal{U}$ on $[-1, 1]$). Hence the composites $\psi_k(2F(t) - 1)$ is a basis of $L_{2,0}(F, \text{odd})$ since the composite of two odd functions is odd. This justifies that we can take $a_k = \psi_k(2F(t) - 1)$. But F is unknown, we estimate it using the residuals $\varepsilon_i = Y_{i\cdot} - \mu_0, i = 1, \dots, n$ by the symmetrized empirical distribution function,

$$\mathbb{F}_{\mu_0}(t) = \frac{1}{2n} \sum_{i=1}^{n} \big(\mathbf{1}[Y_{i\cdot} - \mu_0 \le t] + \mathbf{1}[-(Y_{i\cdot} - \mu_0) < t]\big), \quad t \in \mathcal{R}.$$

Again we must justify $2\mathbb{F}_{\mu_0}(t)-1$ is odd. This is easy to prove. Indeed,

$$\begin{aligned}
-\big(2\mathbb{F}_{\mu_0}(-t)-1\big) &= 1-\frac{1}{n}\sum_{i=1}^{n}\big(\mathbf{1}[\varepsilon_i\le -t]+\mathbf{1}[-\varepsilon_i<-t]\big)\\
&= 1-\frac{1}{n}\sum_{i=1}^{n}\big(2-\mathbf{1}[-\varepsilon_i<t]-\mathbf{1}[\varepsilon_i\le t]\big)\\
&= 2\mathbb{F}_{\mu_0}(t)-1.
\end{aligned}$$

This motivates us to utilize the first r_n equalities in (2.2.6) as constraints to construct the jackknife empirical likelihood with side information as follows:

$$\begin{aligned}
\mathscr{R}_n(\mu,\sigma_u^2)=\sup\Big\{&\prod_{i=1}^{n} n\pi_i : \boldsymbol{\pi}\in\mathscr{P}_n,\ \sum_{i=1}^{n}\pi_i(V_{ni}(h)-\sigma_u^2)=0,\\
&\sum_{i=1}^{n}\pi_i\psi_k\big(2\mathbb{F}_\mu(Y_{i\cdot}-\mu)-1\big)=0,\quad k=1,\dots,r_n\Big\},
\end{aligned}$$

where $V_{nj}(h)$'s are the jackknife pseudo values of the U-statistic $U_n(h)$. We shall allow r_n to grow slowly to infinity with the sample n such that r_n^4/n tends to zero. Suppose $r_n h_1$ is Lindeberg. Then under the null hypothesis one has

$$\frac{-2\log\mathscr{R}_n(\mu_0,\sigma_{u0}^2)-r_n-1}{\sqrt{2(r_n+1)}}\Longrightarrow\mathscr{N}(0,1), \tag{2.2.7}$$

where $(\mu_0,\sigma_{u0}^2)\in\mathcal{R}\times\mathcal{R}^+$ denote the true values of parameter. The proof of (2.2.7) can be found the last section. This shows that under the null hypothesis $-2\log\mathscr{R}_n(\mu_0,\sigma_{u0}^2))$ is approximately chi-square-distributed with r_n+1 degrees of freedom. Thus for $0<\alpha<1$,

$$P\big(-2\log\mathscr{R}_n(\mu_0,\sigma_{u0}^2))>\chi^2_{1-\alpha}(r_n+1)\big)\xrightarrow{P}\alpha.$$

This exhibits that the test $\mathbf{1}[-2\log\mathscr{R}_n(\mu_0,\sigma_{u0}^2))>\chi^2_{1-\alpha}(r_n+1)]$ has asymptotic size α. Also, an asymptotic $1-\alpha$ confidence set for (μ_0,σ_{u0}^2) is given by

$$\big\{(\mu,\sigma^2)\in\mathcal{R}\times\mathcal{R}^+ :\ -2\log\mathscr{R}_n(\mu,\sigma^2)\le\chi^2_{1-\alpha}(r_n+1)\big\}.$$

Let us now calculate h_1. Recall $\mathbf{Y}_i=(Y_{i1},\dots,Y_{iJ})^\top$ with $J\ge 2$. Let $\mathbf{y}=(y_1,\dots,y_J)^\top$. One finds

$$E\big(U_{11}|\mathbf{Y}_1=\mathbf{y}_1\big)=U_{11},\quad E\big(U_{22}|\mathbf{Y}_1=\mathbf{y}_1\big)=E\big(U_{22}\big)=\sigma_\epsilon^2,$$

$$E\big(U_{12}|\mathbf{Y}_1=\mathbf{y}_1\big) = \binom{J}{2}^{-1}\sum_{1\le j<j'\le n} E\big(2^{-1}(y_j - Y_{2j'})^2\big)$$
$$= \binom{J}{2}^{-1}\sum_{1\le j<j'\le n} 2^{-1}(y_j-\mu)^2+\sigma^2/2,$$

and

$$E\big(U_{21}|\mathbf{Y}_1=\mathbf{y}_1\big) = \binom{J}{2}^{-1}\sum_{1\le j<j'\le n} 2^{-1}(y_{j'}-\mu)^2+\sigma^2/2.$$

Thus

$$h_1(\mathbf{y}) = 2^{-1}\big(E\big(U_{21}+U_{12}|\mathbf{Y}_1=\mathbf{y}_1\big) - E\big(U_{11}+U_{22}|\mathbf{Y}_1=\mathbf{y}_1\big)\big)$$
$$= \binom{J}{2}^{-1}\sum_{1\le j<j'\le n} 2^{-1}(y_j-\mu)(y_{j'}-\mu)+\sigma_u^2/2.$$

It is easy to compute $E(h_1(\mathbf{Y}_1)) = \sigma_u^2$. Hence the centered version of h_1 is

$$\tilde{h}_1(\mathbf{Y}_1) = \binom{J}{2}^{-1}\sum_{1\le j<j'\le n} 2^{-1}\big((Y_{1j}-\mu)(Y_{1j'}-\mu)-\sigma_u^2\big). \tag{2.2.8}$$

Clearly if Y_{ij}'s are bounded then $r_n h_1$ is Lindeberg.

2.3 Empirical likelihood for the simplicial depth function

Let $\mathbf{X}_1,\ldots,\mathbf{X}_n$ be independent and identically distributed random vectors from a distribution Q on $\mathcal{R}^m$. Liu (1990) introduced the *simplicial depth* of a point $\mathbf{x}\in\mathcal{R}^m$ with respect to distribution Q defined by the probability that the point $\mathbf{x}$ is contained inside a random simplex whose vertices are $m+1$ independent observations from Q, that is,

$$D(\mathbf{x}) = P\big(\mathbf{x}\in\Delta(\mathbf{X}_1,\ldots,\mathbf{X}_{m+1})\big),\quad \mathbf{x}\in\mathcal{R}^m,$$

where $\Delta(\mathbf{X}_1,\ldots,\ \mathbf{X}_{m+1})$ denotes the random simplex with vertices $\mathbf{X}_1,\ldots,\ \mathbf{X}_{m+1}$, i.e., the closed simplex with vertices $\mathbf{X}_1,\ldots,\ \mathbf{X}_{m+1}$. Note that $D(\mathbf{x})$ is the population simplicial depth of point $\mathbf{x}$ and can be estimated by the sample simplicial depth $D_n(\mathbf{x})$ of point $\mathbf{x}$ given by the U-statistic

$$D_n(\mathbf{x}) = \binom{n}{m+1}^{-1}\sum_{1\le i_1<\ldots<i_{m+1}\le n}\mathbf{1}[\mathbf{x}\in\Delta(\mathbf{X}_{i_1},\ldots,\mathbf{X}_{i_{m+1}})],\quad \mathbf{x}\in\mathcal{R}^m.$$

The depth function can be used to define the (multivariate) simplicial median and to give an ordering of data points in space from center outward. As a location parameter, the simplicial median coincides with the center of angular symmetry – a more general type of symmetry than the usual central symmetry introduced by Liu (1990). Specifically, a random vector $\mathbf{X}$ is *angularly symmetric* about a point $\mathbf{c}$ in $\mathcal{R}^m$ if and only if the random direction of from $\mathbf{c}$ to X *centrally symmetric* about $\mathbf{c}$,

$$\frac{\mathbf{X}-\mathbf{c}}{\|\mathbf{X}-\mathbf{c}\|} \stackrel{D}{=} -\frac{\mathbf{X}-\mathbf{c}}{\|\mathbf{X}-\mathbf{c}\|}.$$

When additional information such as angular symmetry is available about the underlying distribution Q, tests or confidence sets based the sample depth $D_n(\mathbf{x})$ do not utilize the additional information. Multivariate data can be understood through their marginal distributions. Often there is available some information about marginal distributions. For example, the marginal medians of a multivariate distribution are known. We now use the developed empirical likelihood theory for U-statistics to construct confidence sets for the value of the simplicial depth at a given point $\mathbf{x}_0$ when side information is available.

(i) Suppose side information is expressed by equality $E(\mathbf{g}(\mathbf{X})) = 0$ for some $\mathcal{R}^r$-valued square-integrable function $\mathbf{g}$. For example, angular symmetry implies

$$E\left((\mathbf{X}-\mathbf{c})/\|\mathbf{X}-\mathbf{c}\|\right) = 0, \tag{2.3.1}$$

which corresponds to $\mathbf{g}(\mathbf{x}) = (\mathbf{x}-\mathbf{c})/\|\mathbf{x}-\mathbf{c}\|, \mathbf{x} \in \mathcal{R}^m$ for some specified constant $\mathbf{c}$. As another example, consider the case that the marginal median of X_k is known and equal to b_k for some specified value b_k where $k = 1, \ldots, m$. Assume the marginal distributions are continuous. Then

$$E\left(\text{sign}(X_k - b_k)\right) = 0, \quad k = 1, \ldots, m. \tag{2.3.2}$$

In this case, we take $\mathbf{g}(x_1, \ldots, x_m) = (\text{sign}(x_1 - b_1), \ldots, \text{sign}(x_m - b_m))^\top$. This motivates us to look at the jackknife empirical likelihood with side information as follows:

$$\mathscr{R}_n(D, \mathbf{g}) = \sup\Big\{\prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \ \sum_{j=1}^n \pi_j(V_{nj} - D) = 0, \sum_{j=1}^n \pi_j \mathbf{g}(\mathbf{X}_j) = 0\Big\},$$

where V_{nj}'s are the jackknife pseudo values of the simplicial depth $D_n(\mathbf{x}_0)$ as a U-statistic and $D \in \mathcal{R}^+$. Note that the kernel is $h(\mathbf{X}_1, \ldots, \mathbf{X}_m) = \mathbf{1}[\mathbf{x}_0 \in \Delta(\mathbf{X}_1, \ldots, \mathbf{X}_m)]$ and $h_1(\mathbf{x}) = P(\mathbf{x}_0 \in \Delta(\mathbf{x}, \mathbf{X}_2, \ldots, \mathbf{X}_m)), \mathbf{x} \in \mathbf{R}^m$. Suppose the dispersion matrix $\mathrm{Var}\left((mh_1(\mathbf{X}), \mathbf{g}^\top(\mathbf{X}))^\top\right)$ is non-singular. Let $D_0 = D(\mathbf{x}_0)$. Then by Corollary 3.2.1,

$$-2\log \mathscr{R}_n(D_0, \mathbf{g}) \Longrightarrow \chi^2(r+1).$$

(ii) In semi-parametric models side information can often be expressed by infinitely many equalities. Usually there is available some partial information about a multivariate distribution, for example, we may know that two marginal distributions are independent, or one marginal distribution is known. Let us now use the latter as an example to illustrate our approach. Let F be the distribution function of $\mathbf{X}$ and write $\mathbf{X} = (X_1, \ldots, X_m)^\top$. Suppose the distribution F_1 of X_1 is known $F_1 = F_{10}$ for some continuous distribution F_{10}. This is equivalent to

$$\int a_k \, dF = \int a_k \, dF_{10} = 0, \quad k = 1, 2, \ldots, \tag{2.3.3}$$

where a_k is an orthonormal basis of $L_{2,0}(F_{10})$. Here we shall take $a_k = \phi_k \circ F_{10}, k = 1, 2, \ldots,$ where ϕ_k is the usual trigonometric basis given in (2.1.4). The above consideration suggests us to look at the jackknife empirical likelihood with side information as follows:

$$\begin{aligned}\mathscr{R}_n(D, F_{10}) = \sup\Big\{ \prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \ \sum_{j=1}^n \pi_j (V_{nj} - D) = 0, \\ \sum_{j=1}^n \pi_j \phi_k(F_{10}(X_{1j})) = 0, \quad k = 1, \ldots, r_n \Big\}, \ D \in \mathcal{R}^+,\end{aligned}$$

where X_{1j} is the first component of $\mathbf{X}_j$. Here we use the first r_n equations in (2.3.3). We now assume $m \geq 2$ and at least one of the components $X_k : k \geq 2$ of $\mathbf{X}$ is non-degenerate, i.e. $P(X_d = c) < 1$ for some $d \geq 2$ and arbitrary constant c. Then under the null hypothesis we have

$$\frac{-2\log \mathscr{R}_n(D_0, F_{10}) - (r_n + 1)}{\sqrt{2(1 + r_n)}} \Longrightarrow \mathscr{N}(0, 1), \tag{2.3.4}$$

as both r_n and n tend to infinity such that r_n^3/n tends to zero. The details of (2.3.4) can be found in the last section. This shows that under the null hypothesis $-2\log \mathscr{R}_n(D_0, F_{10})$ is approximately chi-square-distributed with $r_n + 1$ degrees of freedom. Thus for $0 < \alpha < 1$,

$$P\big(-2\log \mathscr{R}_n(D_0, F_{10}) > \chi^2_{1-\alpha}(r_n+1)\big) \xrightarrow{P} \alpha,$$

This means that the test $\mathbf{1}[-2\log \mathscr{R}_n(D_0, F_{10}) > \chi^2_{1-\alpha}(r_n+1)]$ has asymptotic size α. Our result generalizes Example 1 of Peng and Schick (2013c) to U-statistics and has the same rate for r_n as Peng and Schick (2013c).

3. MAIN RESULTS

In this chapter, we first introduce the notation and present some general results. Then we prove the Wilks theorems for vector U-statistics, study jackknife empirical likelihood for U-statistics with finitely many constraints and with estimated constraints. We also investigate the asymptotic behaviors of the jackknife empirical likelihood with growing number of constraints. The constraints are allowed to use estimated criteria functions. In the end, we prove that the Wilks theorem still holds under suitable conditions when the number of constraints grows to infinity with sample size.

3.1 General results

In this section, We first introduce some of the notation we use throughout. We then state some results from Peng and Schick (2013c) which are tailored for our use. Based on these results, we prove Lemma 1.2.1 and a useful general theorem in the end.

3.1.1 Notation

Denote $\mathbf{A}^{\otimes 2} = \mathbf{A}\mathbf{A}^\top$ and $\mathbf{A} \otimes \mathbf{A}$ the Kronecker product for a vector or matrix $\mathbf{A}$. Recall that $\|\mathbf{A}\|$ denotes the euclidean norm of a matrix $\mathbf{A}$ and write $|\mathbf{A}|_o$ for the operator (or spectral) norm which are defined by

$$\|\mathbf{A}\|^2 = \text{trace}(\mathbf{A}^\top\mathbf{A}) = \sum_{i,j} \mathbf{A}_{ij}^2 \quad \text{and} \quad |\mathbf{A}|_o = \sup_{|\mathbf{u}|=1} |\mathbf{A}\mathbf{u}| = \sup_{|\mathbf{u}|=1} (\mathbf{u}^\top\mathbf{A}^\top\mathbf{A}\mathbf{u})^{1/2}.$$

In other words, the squared euclidean norm $\|\mathbf{A}\|^2$ equals the sum of the eigen values of $\mathbf{A}^\top\mathbf{A}$, while the squared operator norm $|\mathbf{A}|_o^2$ equals the largest eigen value of $\mathbf{A}^\top\mathbf{A}$. Consequently, the inequality $|\mathbf{A}|_o \le \|\mathbf{A}\|$ holds. Thus we have

$$|\mathbf{A}\mathbf{x}| \le |\mathbf{A}|_o\|\mathbf{x}\| \le \|\mathbf{A}\|\|\mathbf{x}\|$$

for compatible vectors $\mathbf{x}$. We should also point out that

$$|\mathbf{A}|_o = \sup_{|\mathbf{u}|=1}\sup_{|\mathbf{v}|=1}\mathbf{u}^\top\mathbf{A}\mathbf{v}$$

and that this simplifies to

$$|\mathbf{A}|_o = \sup_{|\mathbf{u}|=1}\mathbf{u}^\top\mathbf{A}\mathbf{u}$$

if $\mathbf{A}$ is a non-negative definite symmetric matrix. Using this and the Cauchy-Schwartz inequality it is easy to see that

$$\Big|\int \mathbf{f}^{\otimes 2}\,d\mu\Big|_o \le \int \|\mathbf{f}\|^2\,d\mu, \tag{3.1.1}$$

whenever μ is a measure and $\mathbf{f}$ is measurable function into $\mathcal{R}^s$ such that $\int \|\mathbf{f}\|^2\,d\mu$ is finite.

3.1.2 General results

Let $\mathcal{T}_{n1},\dots,\mathcal{T}_{nn}$ be r_n-dimensional random vectors. With them we associate the empirical likelihood

$$\mathscr{R}_n = \sup\Big\{\prod_{j=1}^n n\pi_j : \boldsymbol{\pi}\in\mathscr{P}_n,\ \sum_{j=1}^n \pi_j\mathcal{T}_{nj}=0\Big\}.$$

To study the asymptotic behavior of $\mathscr{R}_n$ we introduce

$$\mathcal{T}_n^* = \max_{1\le j\le n}\|\mathcal{T}_{nj}\|,\quad \bar{\mathcal{T}}_n = \frac{1}{n}\sum_{j=1}^n \mathcal{T}_{nj},\quad \mathbb{S}_n = \frac{1}{n}\sum_{j=1}^n \mathcal{T}_{nj}^{\otimes 2},$$

and

$$\mathcal{T}_n^{(\nu)} = \sup_{\|\mathbf{u}\|=1}\frac{1}{n}\sum_{j=1}^n(\mathbf{u}^\top\mathcal{T}_{nj})^\nu,\quad \nu=3,4,$$

and let $\lambda_n = \lambda_{\min}(\mathbb{S}_n)$ and $\Lambda_n = \lambda_{\max}(\mathbb{S}_n)$ denote the smallest and largest eigen values of $\mathbb{S}_n$, i.e.,

$$\lambda_n = \inf_{\|\mathbf{u}\|=1} \mathbf{u}^\top \mathbb{S}_n \mathbf{u} \quad \text{and} \quad \Lambda_n = \sup_{\|\mathbf{u}\|=1} \mathbf{u}^\top \mathbb{S}_n \mathbf{u}.$$

Peng and Schick (2013c) investigated the asymptotic behavior of the empirical likelihood when $\mathbb{S}_n$ can be approximated by a sequence of $r_n \times r_n$ dispersion matrices $\mathscr{W}_n$ which is *regular* in the sense that

$$0 < \inf_n \lambda_{\min}(\mathscr{W}_n) \le \sup_n \lambda_{\max}(\mathscr{W}_n) < \infty.$$

We quote their theorem 6.1 below for our purpose.

Lemma 3.1.1 *Let $r_n = r$ for all n. Suppose*

$$\mathcal{T}_n^* = o_p(n^{1/2}), \quad n^{1/2}\bar{\mathcal{T}}_n \Longrightarrow \mathcal{T}, \quad \mathbb{S}_n = \mathscr{W} + o_p(1) \tag{3.1.2}$$

for some random vector $\mathcal{T}$ and $r \times r$ positive definite matrix $\mathscr{W}$. Then

$$-2\log \mathscr{R}_n \Longrightarrow \mathcal{T}^\top \mathscr{W}^{-1} \mathcal{T}.$$

Peng and Schick (2013c) also introduced the following conditions.

(A1) $\mathcal{T}_n^* = o_p(r_n^{-1/2} n^{1/2})$.

(A2) $\|\bar{\mathcal{T}}_n\| = O_p(r_n^{1/2} n^{-1/2})$.

(A3) There is a sequence of regular $r_n \times r_n$ dispersion matrices $\mathscr{W}_n$ such that

$$|\mathbb{S}_n - \mathscr{W}_n|_o = o_p(r_n^{-1/2}).$$

(A4) $\mathcal{T}_n^{(3)} = o_p(r_n^{-1} n^{1/2})$ and $\mathcal{T}_n^{(4)} = o_p(r_n^{-3/2} n)$.

They looked at the case when r_n increases with the sample size n. The following is quoted from their Theorem 6.2.

Lemma 3.1.2 *Let (A1)–(A4) hold. Suppose that r_n increases with n to infinity and that there are $r_n \times r_n$ dispersion matrices $\mathscr{V}_n$ such that $r_n/\mathrm{trace}(\mathscr{V}_n^2) = O(1)$ and*

$$\frac{n\bar{\mathcal{T}}_n^\top \mathscr{W}_n^{-1}\bar{\mathcal{T}}_n - \mathrm{trace}(\mathscr{V}_n)}{\sqrt{2\mathrm{trace}(\mathscr{V}_n^2)}} \Longrightarrow \mathscr{N}(0,1). \tag{3.1.3}$$

Then

$$\frac{-2\log \mathscr{R}_n - \mathrm{trace}(\mathscr{V}_n)}{\sqrt{2\mathrm{trace}(\mathscr{V}_n^2)}} \Longrightarrow \mathscr{N}(0,1). \tag{3.1.4}$$

3.1.3 General results for U-statistics with side information

We now apply Lemma 3.1.1 to find the asymptotic behaviors of the jackknife empirical likelihood for U-statistics with side information.

Recall that the kernel h is square-integrable. Here we further assume throughout that h is non-degenerate, that is, $\mathrm{Var}(h_1(Z)) > 0$. Let $\mathbf{T}_{n1}, \dots, \mathbf{T}_{nn}$ be r_n-dimensional random vectors. With them we associate the jackknife empirical likelihood of U-statistic as follows:

$$\mathscr{R}_n(h) = \sup\Big\{\prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{j=1}^n \pi_j(\tilde{V}_{nj}(h), \mathbf{T}_{nj}^\top)^\top = 0\Big\}.$$

where $\tilde{V}_{nj}(h)$'s are the centered jackknife pseudo values of the U-statistic $U_n(h)$. By the Hoeffding decomposition (1.2.1), we obtain

$$\begin{aligned}\tilde{V}_{nj} &= m\tilde{h}_1(Z_j) + \sum_{c=2}^m \binom{m}{c}\big(nU_{nc}(h_c^*) - (n-1)U_{(n-1)c}^{(-j)}(h_c^*)\big)\\ &= m\tilde{h}_1(Z_j) + \sum_{c=2}^m \binom{m}{c}\big(cU_{(n-1)(c-1)}(h_{(c-1)j}^*) - (c-1)U_{(n-1)c}^{(-j)}(h_c^*)\big),\end{aligned} \tag{3.1.5}$$

where $h_{(c-1)j}^*(z_1, \dots, z_{c-1}) := h_c^*(Z_j, z_1, \dots, z_{c-1})$. Since for arbitrary i, k the joint distributions of $X_j : j \neq i, j = 1, \dots, n$ and $X_j : j \neq k, j = 1, \dots, n$ are identical, it follows

$$E\big(\big(\tilde{V}_{nj} - m\tilde{h}_1(Z_j)\big)^2\big) = E\big(\big(\tilde{V}_{n1} - mh_1^*(Z_1)\big)^2\big), \quad j = 1, \dots, n.$$

Using now the inequality $(a_1+\cdots+a_m)^2 \le m(a_1^2+\cdots+a_m^2)$ for reals a_j's, we derive

$$E\big((\tilde V_{n1} - mh_1^*(Z_1))^2\big) \le 2m \sum_{c=2}^m \binom{m}{c}^2 \big(c^2 \operatorname{Var}(U_{(n-1)(c-1)}(h^*_{(c-1)1}))$$
$$+ (c-1)^2 \operatorname{Var}(U^{(-1)}_{(n-1)c}(h^*_c))\big), \quad j=1,\dots,n.$$

It is well known that the variances of U-statistics satisfy (see e.g. page 189, Serfling (1980))

$$\operatorname{Var}\big(U_{(n-1)(c-1)}(h^*_{(c-1)1})\big) = O(n^{-c+1}), \quad \operatorname{Var}\big(U^{(-1)}_{(n-1)c}(h^*_c)\big) = O(n^{-c}),$$

we thus obtain

$$E\big((\tilde V_{nj} - m\tilde h_1(Z_j))^2\big) = E\big((\tilde V_{n1} - mh_1^*(Z_1))^2\big) = O(1/n).$$

From this it yields (1.2.3) and hence Lemma 1.2.1. Moreover,

$$\sum_{j=1}^n \big(\tilde V_{nj} - m\tilde h_1(Z_j)\big)^2 = O_p(1). \tag{3.1.6}$$

This follows from

$$P\Big(\sum_{j=1}^n \big(\tilde V_{nj} - m\tilde h_1(Z_j)\big)^2 > M\Big) \le M^{-1} n E\big((\tilde V_{n1} - m\tilde h_1(Z_1))^2\big) = O(M^{-1}),$$

which converges to zero as M tends to infinity.

We now impose the following on $\mathbf{T}_{nj}$'s and the jackknife pseudo values.

(B1) $\mathbf{T}_{n1},\dots,\mathbf{T}_{nn}$ are $\mathcal{R}^{r_n}$-valued random vectors satisfying (A1) – (A4) with $\mathbb{W}_n := \mathscr{W}_n$.

(B2) There exists r_n-dimensional vector $\mathbf{C}_n$ such that

$$\Big\|\frac{1}{n}\sum_{j=1}^n \tilde V_{nj}\mathbf{T}_{nj} - \mathbf{C}_n\Big\| = o_p(r_n^{-1/2}).$$

Suppose (B1) and (B2) are met. Let $\mathcal{T}_{nj} = (\tilde V_{nj},\, \mathbf{T}_{nj}^\top)^\top$. Since $\mathbf{T}_{nj}$'s satisfy (A2) and in view of (1.2.5) and $n\operatorname{Var}(U_n) = O(1)$, it follows $\mathcal{T}_{nj}$'s also satisfy (A2). Next, by Markov inequality and in view of (1.2.3) we conclude that for any $\epsilon > 0$,

$$P\Big(\max_{1\le j\le n} |\tilde V_{nj}| > n^{1/2}\epsilon\Big) \le \sum_{j=1}^n P\left(|\tilde V_{nj}| > n^{1/2}\epsilon\right)$$
$$= nP\left(|\tilde V_{n1}| > n^{1/2}\epsilon\right) \le \epsilon^{-2} E\left(|\tilde V_{n1}|^2 \mathbf{1}[|\tilde V_{n1}| > n^{1/2}\epsilon]\right) \to 0, \quad n\to\infty.$$

This shows

$$\max_{1\le j\le n} |\tilde{V}_{nj}| = o_p(n^{1/2}). \tag{3.1.7}$$

By Cauchy inequality, we have for real numbers a_j and b_j,

$$\Big|\frac{1}{n}\sum_{j=1}^{n}(a_j^2 - b_j^2)\Big|^2 \le \frac{1}{n}\sum_{j=1}^{n}(a_j - b_j)^2 \frac{1}{n}\sum_{j=1}^{n} 2(a_j^2 + b_j^2).$$

Applying this with $a_j = \tilde{V}_{nj}$ and $b_j = m\tilde{h}_1(Z_j)$ and in view of (3.1.6), we obtain

$$\Big|\frac{1}{n}\sum_{j=1}^{n}\big(\tilde{V}_{nj}^2 - m^2 E(\tilde{h}_1(Z)^2)\big)\Big| = O_p(n^{-1/2}). \tag{3.1.8}$$

This, (B2) and (A3) yield

$$\Big|\frac{1}{n}\sum_{j=1}^{n}\mathcal{T}_{nj}^{\otimes 2} - \mathscr{W}_n\Big|_o = o_p(r_n^{-1/2}), \tag{3.1.9}$$

provided $r_n = o(n)$, where $\mathscr{W}_n = \mathscr{W}\big(m^2 E(\tilde{h}_1(Z)^2), \mathbf{C}_n, \mathbb{W}_n\big)$.

Assume that the sequence of matrices $\mathscr{W}_n$ is regular. Then the preceding discussion shows that $\mathcal{T}_n$ satisfies (A3). Suppose now that $V_{n1}, \ldots, V_{nn}$ satisfy (A1). Then $\mathcal{T}_{nj}$'s satisfy (A1), while (A4) follows from

$$\mathcal{T}_n^{(3)} \le \Lambda_n \mathcal{T}_n^*, \quad \mathcal{T}_n^{(4)} \le \Lambda_n (\mathcal{T}_n^*)^2, \tag{3.1.10}$$

where $\Lambda_n = \lambda_{\max}(\mathscr{W}_n)$. Thus a sufficient condition for (A1) and (A4) is

$$\mathcal{T}_n^* = o_p(r_n^{-1} n^{1/2}). \tag{3.1.11}$$

Let us now consider the case of fixed number $r_n = r$ of constraints. We have the following.

Theorem 3.1.1 *Let $r_n = r$ for all n. Suppose*

$$\mathbf{T}_n^* = o_p(n^{1/2}), \quad \frac{1}{n}\sum_{j=1}^{n} m\tilde{h}_1(Z_j)\mathbf{T}_{nj} \xrightarrow{P} \mathbf{C}, \quad \frac{1}{n}\sum_{j=1}^{n}\mathbf{T}_{nj}^{\otimes 2} \xrightarrow{P} \mathbb{W} \tag{3.1.12}$$

for some r-dimensional vector $\mathbf{C}$ and $r\times r$ matrix $\mathbb{W}$ such that $\mathscr{W} := \mathscr{W}(\mathrm{Var}(mh_1(Z)), \mathbf{C}, \mathbb{W})$ is nonsingular. Assume

$$n^{-1/2}\sum_{j=1}^{n}\Big(m\tilde{h}_1(Z_j), \mathbf{T}_{nj}^\top\Big)^\top \Longrightarrow \mathcal{T}, \tag{3.1.13}$$

for some $(r+1)$*-dimensional random vector* $\mathcal{T}$. *Then*

$$-2\log \mathscr{R}_n(h) \Longrightarrow \mathcal{T}^\top \mathscr{W}^{-1}\mathcal{T}.$$

Remark 3.1.1 Theorem 3.1.1 is a generalization of Theorem 6.1 of Peng and Schick (2013c) in the sense that if the jackknife pseudo value of the U-statistic is replaced by a random variable which is amalgamated as a component to $\mathbf{T}_{nj}$ then it recovers Theorem 6.1.

Proof We shall apply Lemma 3.1.1 to prove the result by verifying its three conditions in (3.1.2) with $\mathcal{T}_{nj} = (\tilde{V}_{nj}, \mathbf{T}_{nj}^\top)^\top$. The first condition in (3.1.2) is implied by the first equality in (3.1.12) and (3.1.7). It is well known (see e.g. page 188, Serfling (1980)) that

$$U_n(h) - \theta = \frac{m}{n}\sum_{j=1}^n \tilde{h}_1(Z_j) + O_p(n^{-1}). \tag{3.1.14}$$

This, (1.2.5) and (3.1.13) yield the second condition of (3.1.2). We now verify that $\mathbf{T}_{nj}$ satisfy (A1) – (A4) and hence (B1) is met. Note first that $\mathscr{W}$ is nonsingular hence the sub matrix $\mathbb{W}$ is also nonsingular. Thus applying the inequality (3.1.11) to $\mathbf{T}_{nj}$ and noticing this is a sufficient condition for (A1) and (A4), we derive by the first equality of (3.1.12) that $\mathbf{T}_{nj}$ satisfy (A1) and (A4). It follows now from (3.1.13) that $\mathbf{T}_{nj}$ satisfy (A2), while the third equality of (3.1.12) gives (A3). We show next that (B2) is also met with $\mathbf{C}_n$ equal to the $\mathbf{C}$ given in (3.1.12). This, in fact, follows from the second equality in (3.1.12), (3.1.6), and

$$\begin{aligned}
&\|\frac{1}{n}\sum_{j=1}^n \tilde{V}_{nj}\mathbf{T}_{nj} - m\tilde{h}_1(Z_j)\mathbf{T}_{nj}\|^2 \le \frac{1}{n}\sum_{j=1}^n (\tilde{V}_{nj} - m\tilde{h}_1(Z_j))^2 \frac{1}{n}\sum_{j=1}^n \|\mathbf{T}_{nj}\|^2 \\
&= O_p(n^{-1})\text{trace}\Big(\frac{1}{n}\sum_{j=1}^n \mathbf{T}_{nj}^{\otimes 2}\Big) = O_p(n^{-1})(\text{trace}(\mathbb{W}) + o_p(1)) = O_p(n^{-1}).
\end{aligned}$$

We now use (3.1.9) to conclude that $\mathcal{T}_{nj}, j = 1, \dots, n$ satisfy the third condition of (3.1.2) and apply Lemma 3.1.1 to complete the proof.

■

3.2 The Wilks theorems for vector U-statistics

In this section, we study the jackknife empirical likelihood when the number of constraints are fixed. We first give the Wilks theorems for vector U-statistics, followed by a corollary about empirical likelihood with side information. In the end, empirical likelihood with estimated constraints are studied.

We shall now consider the application of Theorem 3.1.1 to the case of several U-statistics and derive the asymptotic distribution of the $\mathscr{R}_n(\mathbf{h})$ defined in (1.2.7). Recall that in the Introduction, $\mathbf{h} = (h^{(1)}, \dots, h^{(r)})^\top$ is a vector of argument-symmetric and square-integrable kernels and $\tilde{\mathbf{V}}_{nj}(\mathbf{h}) = (\tilde{V}_{nj}(h^{(1)}), \dots, \tilde{V}_{nj}(h^{(r)}))^\top$. We now apply Theorem 3.1.1 with $\tilde{V}_{nj}(h) = \tilde{V}_{nj}(h^{(1)})$ and $\mathbf{T}_{nj} = (\tilde{V}_{nj}(h^{(2)}), \dots, \tilde{V}_{nj}(h^{(r)}))^\top$ so that $(\tilde{V}_{nj}(h), \mathbf{T}_{nj}^\top)^\top = \tilde{\mathbf{V}}_{nj}(\mathbf{h})$. Set $\mathbf{m} = (m_1, \dots, m_r)^\top$ and $\mathbf{h}_1 = (h_1^{(1)}, \dots, h_1^{(r)})^\top$ and define

$$\mathbf{w} = (\mathbf{m}\tilde{\mathbf{h}}_1)^\top = (m_1 \tilde{h}_1^{(1)}, \dots, m_r \tilde{h}_1^{(r)})^\top.$$

By (3.1.6) and Cauchy inequality, we derive

$$\begin{aligned}
&\Big| \frac{1}{n} \sum_{j=1}^n \Big(\tilde{V}_{nj}(h^{(k)}) \tilde{V}_{nj}(h^{(l)}) - m_j h_1^{(k)}(Z_j) m_k h_1^{(l)}(Z_j) \Big) \Big|^2 \\
&\le 2 \frac{1}{n} \sum_{j=1}^n \Big(\tilde{V}_{nj}(h^{(k)}) - m_j h_1^{(k)}(Z_j) \Big)^2 \frac{1}{n} \sum_{j=1}^n \tilde{V}_{nj}(h^{(l)})^2 \\
&\quad + 2 \frac{1}{n} \sum_{j=1}^n \big(m_j h_1^{(k)}(Z_j) \big)^2 \frac{1}{n} \sum_{j=1}^n \Big(\tilde{V}_{nj}(h^{(l)}) - m_k h_1^{(l)}(Z_j) \Big)^2 \\
&= O_p(n^{-1}) = o_p(1), \quad k \ne l, k, l = 1, \dots, r.
\end{aligned}$$

Hence by the law of large numbers we get

$$\frac{1}{n} \sum_{j=1}^n \tilde{V}_{nj}(h^{(k)}) \tilde{V}_{nj}(h^{(l)}) \xrightarrow{P} E\big(m_j h_1^{(k)}(Z_1) m_k h_1^{(l)}(Z_1) \big).$$

This and applications of (3.1.8) with $m = m_k, k = 1, \dots, r$ yield

$$\frac{1}{n} \sum_{j=1}^n \tilde{\mathbf{V}}_{nj}(\mathbf{h})^{\otimes 2} \xrightarrow{P} \mathbb{W}(\mathbf{m}\mathbf{h}_1) = E\big(\mathbf{w}(Z)^{\otimes 2} \big),$$

which establishes the second and third equalities of (3.1.2). If $\mathbb{W}(\mathbf{mh}_1)$ is non-singular, then an application of the usual central limit theorem gives

$$n^{-1/2}\sum_{j=1}^{n}\mathbf{w}(Z_j)\Longrightarrow\mathscr{N}(0,\mathbb{W}(\mathbf{mh}_1)),$$

which yields (3.1.13). Apparently, the last two displays express the usual central limit theorem for a vector U-statistic. Moreover, an analogous argument to (3.1.7) yields the first equality in (3.1.2). Thus by Theorem 3.1.1 we have proved the following Wilks theorem for a vector U-statistic.

Theorem 3.2.1 *Let* $r_n = r$ *for all* n*. Suppose* $\mathbb{W}(m_1 h_1^{(1)}, \ldots, m_r h_1^{(r)})$ *is non-singular. Then* $\mathscr{R}_n(h^{(1)}, \ldots, h^{(r)})$ *defined in (1.2.7) satisfies*

$$-2\log\mathscr{R}_n(h^{(1)},\ldots,h^{(r)})\Longrightarrow\chi^2(r). \tag{3.2.1}$$

This theorem can be applied to obtain empirical likelihood tests for many commonly used tests that appear in the literature. One of the advantages of these tests is that additional information about the underlying distribution can be conveniently incorporated, resulting in more powerful tests than the usual tests. See the preceding section for a list of commonly used tests.

Theorem 3.2.1 generalizes Owen's vector empirical likelihood theorem for a usual vector mean and the jackknife empirical likelihood theorem for a univariate U-statistic-defined mean of Jing, *et al.* (2009) to the jackknife empirical likelihood theorem for a vector U-statistic-defined mean. The theorem holds under the same condition as required for the asymptotic normality of the U-statistic. A special case of it is when side information is expressed by the usual estimating equation $E(\mathbf{g}(Z)) = 0$, which corresponds to the empirical likelihood $\mathscr{R}_n(\mathbf{h}, \mathbf{g})$ given in (1.2.6) when the kernel $\mathbf{h}$ is a scalar function. This is a common case and we give a corollary here for the convenience of its application.

Corollary 3.2.1 *Let* $r_n = r$ *for all* n*. Let* $\mathbf{h}$ *be a square-integrable kernel taking values* $\mathcal{R}^s$ *and* $\mathbf{m}$ *be a* s*-dimensional vector of positive integers. Suppose* $\mathbf{g}$ *is a* *measurable function from $\mathcal{Z}$ to $\mathcal{R}^r$ such that* $\int \mathbf{g}\, dQ = 0$ *and* $\int \|\mathbf{g}\|^2\, dQ < \infty$. *Assume* $\mathbb{W}(\mathbf{mh}_1, \mathbf{g})$ *is nonsingular. Then*

$$-2\log \mathscr{R}_n(\mathbf{h}, \mathbf{g}) \Longrightarrow \chi^2(r+s).$$

We now investigate the case that kernels must be estimated. Let $\hat{h}^{(1)}, \ldots, \hat{h}^{(r)}$ be estimators of $h^{(1)}, \ldots, h^{(r)}$ respectively. Specifically, each $\hat{h}^{(k)}$ is a measurable function from $\mathcal{Z}^m$ to $\mathcal{R}$ such that it is argument-symmetric and square-integrable. We shall refer it to as an estimator of a kernel or a kernel estimator. Observe that the nice identity (1.2.5) still holds when the kernel h is replaced by a kernel estimator $\hat{h}$, i.e.,

$$U_n(\hat{h}) = \frac{1}{n}\sum_{j=1}^{n} V_{nj}(\hat{h}). \tag{3.2.2}$$

This is due to the fact that the identity holds based on the algebraic not probabilistic properties. Let us now look at the jackknife empirical likelihood

$$\hat{\mathscr{R}}_n = \sup\Big\{\prod_{j=1}^{n} n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{j=1}^{n} \pi_j\big(V_{nj}(\hat{h}^{(k)}) - \theta_k\big) = 0, k = 1, \ldots, r\Big\},$$

where $\theta_k = E\left(U_{nm_k}(h^{(k)})\right)$. Set $\hat{\mathbf{h}} = (\hat{h}^{(1)}, \ldots, \hat{h}^{(r)})$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)^\top$ and

$$\hat{\mathscr{R}}_n = \mathscr{R}_n(\hat{\mathbf{h}}), \quad \hat{\mathbf{V}}_{nj} = \mathbf{V}_{nj}(\hat{\mathbf{h}}) = \big(V_{nj}(\hat{h}^{(1)}), \ldots, V_{nj}(\hat{h}^{(r)})\big)^\top.$$

We have the following result.

Theorem 3.2.2 *Let $r_n = r$ for all n. Suppose $\hat{h}^{(k)}$ is an estimator of the kernel $h^{(k)}$ for $k = 1, \ldots, r$ such that*

$$\max_{1\le j\le n} \|\mathbf{V}_{nj}(\hat{\mathbf{h}})\| = o_p(n^{1/2}),\ \frac{1}{n}\sum_{j=1}^{n}(\mathbf{V}_{nj}(\hat{\mathbf{h}}) - \boldsymbol{\theta})^{\otimes 2} = \mathbb{W}(\mathbf{mh}_1) + o_p(1) \tag{3.2.3}$$

for the dispersion matrix $\mathbb{W}(\mathbf{mh}_1)$ that is non-singular, and that there exists some measurable function $\mathbf{v} = (v_1, \ldots, v_r)^\top$ from $\mathcal{Z}$ to $\mathcal{R}^r$ satisfying $\int \mathbf{v}\, dQ = 0$,

$$U_{nm_k}(\hat{h}^{(k)}) = \theta_k + \frac{1}{n}\sum_{j=1}^{n} v_k(Z_j) + o_p(n^{-1/2}), \quad k = 1, \ldots, r, \tag{3.2.4}$$

and

$$n^{-1/2}\sum_{j=1}^{n}\mathbf{v}(Z_j)\Longrightarrow\mathcal{T} \tag{3.2.5}$$

for some r-dimensional random vector $\mathcal{T}$. Then

$$-2\log\mathscr{R}_n(\hat{\mathbf{h}})\Longrightarrow\mathcal{T}^\top\mathbb{W}(\mathbf{m}\mathbf{h}_1)^{-1}\mathcal{T}.$$

Remark 3.2.1 The second equality in (3.2.3) ensures that the sample variance of the kernel estimators $\hat{h}^{(k)}, k=1,\dots,r$ centered at the true means of the (component) U-statistics correctly estimates the true dispersion matrix $\mathbb{W}(\mathbf{m}\mathbf{h}_1)$, while (3.2.4) implies the asymptotic distribution is allowed to be different from the asymptotic distribution of the U-statistic $U_{nm_k}(h^{(k)})$ (i.e. $\mathscr{N}(0,\mathbb{W}(\mathbf{m}\mathbf{h}_1))$). This leads to the limit which may not be Chi-square.

Proof The result immediately follows from an application of Theorem 3.1.1 with $\mathcal{T}_{nj}=\mathbf{V}_{nj}(\hat{\mathbf{h}})-\boldsymbol{\theta}$ in view of Remark 3.1.1. In fact, (3.2.3) implies (3.1.2), while (3.1.13) follows from (3.2.4), (3.2.5) and the corresponding identities of (3.2.2) by setting $\hat{h}=\hat{h}^{(k)}$ for $k=1,\dots,r$. This finishes the proof. ■

Analogous to Corollary 3.2.1, a special case of the above theorem is when the side information is expressed by the usual estimating equation $E(\mathbf{g}(Z))=0$ for some measurable function $\mathbf{g}$ from $\mathcal{Z}$ to $\mathcal{R}^r$, where $\mathbf{g}$ must be estimated by $\hat{\mathbf{g}}$, while the s-dimensional kernel vector $\mathbf{h}$ is known. This corresponds to the empirical likelihood $\mathscr{R}_n(\mathbf{h},\hat{\mathbf{g}})$ defined similar to (1.2.6). The proof can be carried out similar to Theorem 3.2.2 or Theorem 3.3.2. Recall that for two vectors $\mathbf{a},\mathbf{b}\in\mathcal{R}^d$, we define $\mathbf{ab}=(a_1b_1,\dots,a_db_d)^\top$. Let s be a fixed integer. Let $\mathbf{h}_1=(h_1^{(1)},\dots,h_1^{(s)})^\top$ and $\mathbf{m}=(m_1,\dots,m_s)^\top$ be the vector of orders of the U-statistics $U_{nm_1}(h^{(1)}),\dots,U_{nm_s}(h^{(s)})$.

Corollary 3.2.2 *Let $r_n=r$ for all n. Suppose $\mathbf{g}$ is a measurable function from $\mathcal{Z}$ to $\mathcal{R}^r$ with $\int\mathbf{g}\,dQ=0$ and (component-wise) finite moments*

$$\mathbb{C}=\int\mathbf{m}\tilde{\mathbf{h}}_1\otimes\mathbf{g}\,dQ,\quad \mathbb{W}=\int\mathbf{g}^{\otimes 2}\,dQ.$$

Suppose $\hat{\mathbf{g}}$ *is an estimators of* $\mathbf{g}$ *such that*

$$\max_{1\le j\le n} \|\hat{\mathbf{g}}(Z_j))\| = o_p(n^{1/2}), \tag{3.2.6}$$

$$\|\frac{1}{n}\sum_{j=1}^{n} \tilde{\mathbf{V}}_{nj}(\mathbf{h}) \otimes \hat{\mathbf{g}}(Z_j) - \mathbb{C}\| = o_p(1), \quad \frac{1}{n}\sum_{j=1}^{n} \hat{\mathbf{g}}^{\otimes 2}(Z_j) = \mathbb{W} + o_p(1) \tag{3.2.7}$$

for which $\mathscr{W} := \mathscr{W}(\mathrm{Var}(\mathbf{mh}_1(Z)), \mathbb{C}, \mathbb{W})$ *is non-singular. Assume*

$$n^{-1/2}\sum_{j=1}^{n} \Big((\mathbf{m}\tilde{\mathbf{h}}_1(Z_j))^\top, \hat{\mathbf{g}}(Z_j)^\top\Big)^\top \Longrightarrow \mathcal{T}, \tag{3.2.8}$$

for some $(s+r)$*-dimensional random vector* $\mathcal{T}$*. Then*

$$-2\log \mathscr{R}_n(\mathbf{h}, \hat{\mathbf{g}}) \Longrightarrow \mathcal{T}^\top \mathscr{W}^{-1}\mathcal{T}. \tag{3.2.9}$$

3.3 Growing number of constraints

In this section, we shall allow the number of constraints to grow with the sample size and study the asymptotic behaviors of the empirical likelihood with side information. We shall consider both known and estimated constraints.

Following Peng and Schick (2013c), a sequence of measurable functions v_n from $\mathcal{Z}$ to $\mathcal{R}$ is *Lindeberg* if for every $\epsilon > 0$,

$$\int |v_n|^2 \mathbf{1}[|v_n| > \epsilon\sqrt{n}]\, dQ \to 0.$$

Useful properties for Lindeberg sequences can be found in Peng and Schick (2013c). Here we quote three properties for our later use.

(L0) If the sequences u_n and v_n are Lindeberg, so are the sequences $\max\{|u_n|, |v_n|\}$ and $u_n + v_n$.

(L1) If the sequence v_n is Lindeberg, then $\max_{1\le j\le n} |v_n(Z_j)| = o_p(n^{1/2})$.

(L2) If $\int |\mathbf{v}_n|^r\, dQ = o(n^{r/2-1})$ for some $r > 2$, then v_n is Lindeberg.

3.3.1 Intermediate results

Let $(\mathcal{Z}, \mathscr{S})$ be a measurable space, and $Z_1, \dots, Z_n$ be independent copies of the $\mathcal{Z}$-valued random variable Z with distribution Q. Let r_n be a positive integer that tends to infinity with n. Recall that h is a kernel and $h_1(z) = E(h(z, Z_2, \ldots, Z_m)), z \in \mathcal{Z}$. Let $\mathbf{g}_n$ denote a measurable function from $\mathcal{Z}$ to $\mathcal{R}^{r_n}$ such that $\int \mathbf{g}_n \, dQ = 0$ and $\int \|\mathbf{g}_n\|^2 \, dQ$ is finite. With $\mathbf{g}_n$ as side information, we associate the empirical likelihood

$$\mathscr{R}_n(h_1, \mathbf{g}_n) = \sup\Big\{ \prod_{j=1}^{n} n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n, \ \sum_{j=1}^{n} \pi_j \big(\tilde{h}_1(Z_j), \mathbf{g}_n^\top(Z_j)\big) = 0 \Big\}.$$

We are interested in establishing

$$\frac{-2\log \mathscr{R}_n(h_1, \mathbf{g}_n) - r_n - 1}{\sqrt{2(r_n+1)}} \Longrightarrow \mathscr{N}(0,1), \tag{3.3.1}$$

under suitable conditions. Toward this end, let us denote $\mathbf{w}_n = (m\tilde{h}_1, \mathbf{g}_n^\top)^\top$ and set

$$\bar{\mathbf{w}}_n = \frac{1}{n}\sum_{j=1}^{n} \mathbf{w}_n(Z_j), \quad \bar{\mathscr{W}}_n = \frac{1}{n}\sum_{j=1}^{n} \mathbf{w}_n^{\otimes 2}(Z_j), \quad \mathscr{W}_n = \int \mathbf{w}_n^{\otimes 2} \, dQ. \tag{3.3.2}$$

We impose the following.

(C) The sequence of dispersion matrices $\mathscr{W}_n$ is regular.

The following is a quick consequence of Theorem 7.1 of Peng and Schick (2013c)in view of the Lindeberg property (L0).

Lemma 3.3.1 *Suppose (C) holds. Assume the sequences $r_n h_1$ and $r_n\|\mathbf{g}_n\|$ are Lindeberg. Then (3.3.1) holds as r_n tends to infinity with n.*

This result consists in its theoretical importance, and it cannot be used in constructing empirical likelihood unless h_1 is known.

Remark 3.3.1 If the kernel h is bounded, then h_1 is also bounded, hence $r_n h_1$ is Lindeberg.

3.3.2 Estimated kernels and constraints

Often in semiparametric models, the kernel h and constraint $\mathbf{g}_n$ must be estimated by some measurable functions $\hat{h}$ and $\hat{\mathbf{g}}_n$ respectively. Recall throughout $\hat{h}$ is argument-symmetric and square-integrable and $V_{nj}(\hat{h})$ denotes the jackknife pseudo values of the U-statistic $U_n(\hat{h})$. We now concentrate on the jackknife empirical likelihood with estimated kernel and constraints as follows:

$$\mathscr{R}_n(\hat{h}, \hat{\mathbf{g}}_n) = \sup\Big\{\prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{j=1}^n \pi_j\big(\tilde{V}_{nj}(\hat{h}), \hat{\mathbf{g}}_n^\top(Z_j)\big) = 0\Big\}.$$

To study its asymptotic behaviors, let us set

$$\hat{\mathbf{w}}_{nj} = \hat{\mathbf{w}}_{nj}(\hat{h}, \hat{\mathbf{g}}_n) = (\tilde{V}_{nj}(\hat{h}), \hat{\mathbf{g}}_n(Z_j)^\top)^\top, \quad \hat{\mathscr{W}}_n = \frac{1}{n}\sum_{j=1}^n \hat{\mathbf{w}}_{nj}^{\otimes 2}.$$

Peng and Schick (2013c) provided conditions under which the empirical likelihood based on estimated constraints is distributed approximately as chi-square as the number of constraints grows to infinity with increasing sample size. The following result is a re-statement of their Theorem 7.4 tailored for our jackknife empirical likelihood.

Theorem 3.3.1 *Suppose (C) holds. Assume*

$$\max_{1\le j\le n} \|\hat{\mathbf{w}}_{nj}\| = o_p(r_n^{-1} n^{1/2}), \tag{3.3.3}$$

$$|\hat{\mathscr{W}}_n - \mathscr{W}_n|_o = o_p(r_n^{-1/2}), \tag{3.3.4}$$

$$\frac{1}{n}\sum_{j=1}^n \hat{\mathbf{w}}_{nj} = \frac{1}{n}\sum_{j=1}^n \mathbf{v}_n(Z_j) + o_p(n^{-1/2}) \tag{3.3.5}$$

for some measurable function $\mathbf{v}_n$ from $\mathcal{Z}$ into $\mathcal{R}^{r_n}$ such that $\int \mathbf{v}_n\, dQ = 0$ and $\|\mathbf{v}_n\|$ is Lindeberg. Assume further the dispersion matrix of $\mathscr{W}_n^{-1/2}\mathbf{v}_n(Z)$,

$$\mathbb{U}_n = \mathscr{W}_n^{-1/2}\int \mathbf{v}_n\mathbf{v}_n^\top\, dQ \mathscr{W}_n^{-1/2},$$

satisfies $|\mathbb{U}_n|_o = O(1)$ and $r_n/\mathrm{trace}(\mathbb{U}_n^2) = O(1)$. Then, as r_n tends to infinity with n,

$$\frac{-2\log \mathscr{R}_n(\hat{h}, \hat{\mathbf{g}}_n) - \mathrm{trace}(\mathbb{U}_n)}{\sqrt{2\mathrm{trace}(\mathbb{U}_n^2)}} \Longrightarrow \mathscr{N}(0,1).$$

Proof Let $\xi_{nj} = \mathscr{W}_n^{-1/2}\mathbf{v}_n(Z_j)$ and set

$$\bar{\mathbf{v}}_n = \frac{1}{n}\sum_{j=1}^n \mathbf{v}_n(Z_j) \quad \text{and} \quad \bar{\mathbf{T}}_n = \frac{1}{n}\sum_{j=1}^n \hat{\mathbf{w}}_n(Z_j).$$

It follows from (C) that $|\mathscr{W}_n^{1/2}|_o + |\mathscr{W}_n^{-1/2}|_o = O(1)$. Using this and the Lindeberg property of $\|\mathbf{v}_n\|$ we derive

$$L_n(\epsilon) = E\left(\|\xi_{n1}\|^2\mathbf{1}[\|\xi_{n1}\| > \epsilon\sqrt{n}]\right) \to 0, \quad \epsilon > 0.$$

Note that $\mathrm{trace}(\mathbb{U}_n) \le r_n|\mathbb{U}_n|_o = O(r_n)$. Then we have $\mathrm{trace}(\mathbb{U}_n)/\mathrm{trace}(\mathbb{U}_n^2) \le |\mathbb{U}_n|_o r_n/\mathrm{trace}(\mathbb{U}_n^2) = O(1)$ and conclude $\mathrm{trace}(\mathbb{U}_n^2) \to \infty$. Thus Theorem 2 in Peng and Schick (2013a) yields

$$\frac{n\bar{\mathbf{v}}_n^\top \mathscr{W}_n^{-1}\bar{\mathbf{v}}_n - \mathrm{trace}(\mathbb{U}_n)}{\sqrt{2\mathrm{trace}(\mathbb{U}_n^2)}} \Longrightarrow \mathscr{N}(0,1).$$

Next we calculate

$$nE\left(\|\bar{\mathbf{v}}_n\|^2\right) = E\left(\|\mathbf{v}_n(Z)\|^2\right) \le |\mathscr{W}_n^{1/2}|_o^2 E\left(\|\mathscr{W}_n^{-1/2}\mathbf{v}_n(Z)\|^2\right) \le |\mathscr{W}_n^{1/2}|_o^2 \mathrm{trace}(\mathbb{U}_n).$$

This shows that $n\|\bar{\mathbf{v}}_n\|^2 = O_p(r_n)$. Thus we derive with the help of (3.3.5) and $r_n/\mathrm{trace}(\mathbb{U}_n^2) = O(1)$, that $n\|\bar{\mathbf{T}}_n\|^2 = O_p(r_n)$ and

$$\frac{n\bar{\mathbf{T}}_n \mathscr{W}_n^{-1}\bar{\mathbf{T}}_n - \mathrm{trace}(\mathbb{U}_n)}{\sqrt{2\mathrm{trace}(\mathbb{U}_n^2)}} \Longrightarrow \mathscr{N}(0,1).$$

Thus conditions (A1)–(A4) hold with $\mathcal{T}_{nj} = \hat{\mathbf{w}}_n(Z_j)$ in view of (3.1.11) and $\mathcal{T}_n^* = o_p(r_n^{-1}n^{1/2})$. The desired result now follows from Lemma 3.1.2. ■

3.3.3 Known kernels and estimated constraints

We now consider the case that the kernel h is known but the constraint function $\mathbf{g}_n$ must estimated by some measurable function $\hat{\mathbf{g}}_n$. We shall focus on the jackknife empirical likelihood with estimated constraints as follows:

$$\mathscr{R}_n(h, \hat{\mathbf{g}}_n) = \sup\left\{\prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{j=1}^n \pi_j\big(\tilde{V}_{nj}(h), \hat{\mathbf{g}}_n^\top(Z_j)\big) = 0\right\}.$$

Recall $\tilde{V}_{nj} = \tilde{V}_{nj}(h)$ and set $\hat{\mathbf{w}}_{nj}(\hat{\mathbf{g}}_n) = (\tilde{V}_{nj}(h), \hat{\mathbf{g}}_n(Z_j)^\top)^\top$ and

$$\mathbb{W}_n = \int \mathbf{g}_n\mathbf{g}_n^\top\, dQ, \quad \hat{\mathbb{W}}_n = \frac{1}{n}\sum_{j=1}^n \hat{\mathbf{g}}_n(Z_j)^{\otimes 2}, \quad \mathbf{C}_n = \int m\tilde{h}_1\mathbf{g}_n\, dQ.$$

As a special case of Theorem 3.3.1, we have the following.

Theorem 3.3.2 *Suppose $r_n\tilde{h}_1$ is Lindeberg. Suppose $\hat{g}_n$ is an estimator of $\mathbf{g}_n$ such that*

$$r_n \max_{1\le j\le n} \|\hat{\mathbf{g}}_n(Z_j))\| = o_p(n^{1/2}), \tag{3.3.6}$$

$$\|\frac{1}{n}\sum_{j=1}^{n} \tilde{V}_{nj}\hat{\mathbf{g}}_n(Z_j) - \mathbf{C}_n\| = o_p(r_n^{-1/2}), \quad |\hat{\mathbb{W}}_n - \mathbb{W}_n|_o = o_p(r_n^{-1/2}) \tag{3.3.7}$$

for which $\mathscr{W}_n := \mathscr{W}(m^2\operatorname{Var}(h_1(Z)), \mathbf{C}_n, \mathbb{W}_n)$ satisfies (C), and that

$$\frac{1}{n}\sum_{j=1}^{n} \hat{\mathbf{g}}_n(Z_j) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{u}_n(Z_j) + o_p(n^{-1/2}) \tag{3.3.8}$$

for some measurable function $\mathbf{u}_n$ from $\mathcal{Z}$ into $\mathcal{R}^{r_n}$ satisfying that $\int \mathbf{u}_n\, dQ = 0$ and $\|\mathbf{u}_n\|$ is Lindeberg. Assume further the dispersion matrix of $\mathscr{W}_n^{-1/2}\mathbf{v}_n(Z)$ with $\mathbf{v}_n = (m\tilde{h}_1, \mathbf{u}_n^\top)^\top$,

$$\mathbb{U}_n = \mathscr{W}_n^{-1/2}\int \mathbf{v}_n\mathbf{v}_n^\top\, dQ \mathscr{W}_n^{-1/2}, \tag{3.3.9}$$

satisfies $\|\mathbb{U}_n|_o = O(1)$ and $r_n/\operatorname{trace}(\mathbb{U}_n^2) = O(1)$. Then, as r_n tends to infinity with n such that $r_n = o(n^{1/2})$,

$$\frac{-2\log \mathscr{R}_n(h, \hat{\mathbf{g}}_n) - \operatorname{trace}(\mathbb{U}_n)}{\sqrt{2\operatorname{trace}(\mathbb{U}_n^2)}} \Longrightarrow \mathscr{N}(0,1).$$

Proof We verify that $\hat{\mathbf{w}}_n(\hat{\mathbf{g}}_n)$ satisfies the conditions of Theorem 3.3.1. To begin with, if $r_n\tilde{h}_1$ is Lindeberg then

$$r_n \max_{1\le j\le n} |\tilde{V}_{nj}| = o_p(n^{1/2}) \tag{3.3.10}$$

in view of $r_n^2 = o(n)$. Indeed, for $\epsilon > 0$,

$$P(r_n \max_{1\le j\le n} |\tilde{V}_{nj}| > n^{1/2}\epsilon) \le P(r_n \max_{1\le j\le n} |\tilde{V}_{nj} - m\tilde{h}_1(Z_j)| > n^{1/2}\epsilon/2)$$
$$+ P(r_n \max_{1\le j\le n} |m\tilde{h}_1(Z_j)| > n^{1/2}\epsilon/2)$$

By the Lindeberg property (L1), the last probability converges to zero, whereas the second probability is bounded by

$$\sum_{j=1}^{n} P(r_n|\tilde{V}_{nj} - m\tilde{h}_1(Z_j)| > n^{1/2}\epsilon/2) \le \frac{4}{\epsilon^2} r_n^2 E(|\tilde{V}_{n1} - m\tilde{h}_1(Z_1)|^2) = \frac{4}{\epsilon^2}\frac{r_n^2}{n},$$

which converges to zero as n tends to infinity, where the last equality follows from (1.2.4). This proves (3.3.10) hence (3.3.3) in view of (3.3.6). With the aid of (3.1.9), we conclude (3.3.4) from (3.3.7), while (3.1.14), (1.2.5), (3.3.8) in which $\mathbf{v}_n = (m\tilde{h}_1, \mathbf{u}_n^\top)^\top$ and $r_n = o(n)$ together imply (3.3.5). We now apply Theorem 3.3.1 to complete the proof. ■

3.3.4 Known kernels and constraints

It often happens in semiparametric models the kernel h and constraint function $\mathbf{g}_n$ are known. This can be considered as a special case of the estimated kernel and constraints as discussed above. With $\mathbf{g}_n$ as side information, we associate the jackknife empirical likelihood

$$\mathscr{R}_n(h, \mathbf{g}_n) = \sup\Big\{\prod_{j=1}^n n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{j=1}^n \pi_j\big(\tilde{V}_{nj}(h), \mathbf{g}_n^\top(Z_j)\big) = 0\Big\},$$

where $\tilde{V}_{nj}(h)$'s are the jackknife pseudo values of the U-statistic $U_n(h)$. It turns out that in this case the limiting distribution of $-2\log\mathscr{R}_n(h, \mathbf{g}_n)$ is approximately a Chi-square distribution with $r_n + 1$ degrees of freedom as stated below.

Theorem 3.3.3 *Suppose $\mathbf{g}_n$ is a measurable function from $\mathcal{Z}$ to $\mathcal{R}^{r_n}$ which satisfies $\int \mathbf{g}_n\, dQ = 0$. Suppose further the sequences $r_n h_1$ and $r_n\|\mathbf{g}_n\|$ are Lindeberg such that $\mathscr{W}_n = \mathscr{W}(\mathrm{Var}(mh_1(Z_1)), \mathbf{C}_n, \mathbb{W}_n)$ satisfies (C) with $\mathbf{C}_n = \int m\tilde{h}_1\mathbf{g}_n\, dQ$ and $\mathbb{W}_n = \int \mathbf{g}_n^{\otimes 2}\, dQ$. Then*

$$\frac{-2\log\mathscr{R}_n(h, \mathbf{g}_n) - r_n - 1}{\sqrt{2(r_n+1)}} \Longrightarrow \mathscr{N}(0,1). \tag{3.3.11}$$

holds as r_n tends to infinity with n such that $r_n = o(n^{1/2})$.

This result generalizes Theorem 3.2.1 from finitely many constraints to infinitely many.

Proof We verify the conditions of Theorem 3.3.2. Since $r_n\|\mathbf{g}_n\|$ is Lindeberg, it follows that (3.3.6) holds in view of the Lindeberg property (L1). We show next that the Lindeberg property (L1) also implies (3.3.7), whereas (3.3.8) holds with $\mathbf{u}_n = \mathbf{g}_n$, whence $\mathbf{v}_n = (m\tilde{h}_1, \mathbf{g}_n^\top)\top$ which yields $\int \mathbf{v}_n^{\otimes 2}\, dQ = \mathscr{W}_n$ and hence $\mathbb{U}_n = \mathscr{W}_n^{-1/2} \int \mathbf{v}_n \mathbf{v}_n^\top\, dQ \mathscr{W}_n^{-\top/2} = I_{r_n+1}$ satisfies the required conditions. We are left to prove (3.3.7). Note first the $\mathbf{v}_n$ here is equal to the $\mathbf{w}_n$ in (3.3.2) which defines $\mathscr{W}_n = \int \mathbf{w}_n^{\otimes 2}\, dQ$. Since $r_n h_1$ and $r_n\|\mathbf{g}_n\|$ are Lindeberg, it follows from (L0) that $\mathbf{w}_n$ is also Lindeberg. Fix $\epsilon > 0$ and let $\bar{\mathscr{W}}_{n,1}$ and $\bar{\mathscr{W}}_{n,2}$ be the matrices obtained by replacing in the definition of $\bar{\mathscr{W}}_n$ in (3.3.2) the function $\mathbf{w}_n$ by $\mathbf{t}_n = \mathbf{w}_n \mathbf{1}[\|(r_n+1)\mathbf{w}_n\| \le \epsilon\sqrt{n}]$ and $\mathbf{w}_n - \mathbf{t}_n = \mathbf{w}_n \mathbf{1}[\|(r_n+1)\mathbf{w}_n\| > \epsilon\sqrt{n}]$, respectively. Since $\mathscr{W}_n$ satisfies (C), it follows $\lambda_{\max}(\mathscr{W}_n) \le B$ for some $B > 0$ and all n, so that

$$E(\|\mathbf{w}_n\|^2(Z)) = \operatorname{trace}(E(\mathbf{w}_n^{\otimes 2}(Z))) = \operatorname{trace}(\mathscr{W}_n) \le B(r_n+1).$$

Then we find

$$\begin{aligned} nE[\|\bar{\mathscr{W}}_{n,1} - E[\bar{\mathscr{W}}_{n,1}]\|^2] &= \sum_{i=1}^{r_n+1} \sum_{k=1}^{r_n+1} \operatorname{Var}(t_{n,i}(Z) t_{n,k}(Z)) \\ &\le E[\|\mathbf{t}_n\|^4(Z)] \le \frac{\epsilon^2 n}{(r_n+1)^2} E[\|\mathbf{w}_n\|^2(Z)] \le \frac{\epsilon^2 n B(r_n+1)}{(r_n+1)^2}, \end{aligned}$$

and

$$P(\bar{\mathscr{W}}_{n,2} \ne 0) \le P\big(\max_{1\le j\le n} \|(r_n+1)\mathbf{w}_n(Z_j)\| > \epsilon\sqrt{n}\big) \to 0,$$

and using (3.1.1),

$$|E[\bar{\mathscr{W}}_{n,2}]|_o \le E[\|\mathbf{w}_n\|^2(Z)\mathbf{1}[\|(r_n+1)\mathbf{w}_n(Z)\| > \epsilon\sqrt{n}]] = o(r_n^{-2}).$$

From these inequalities it follows immediately the desired (3.3.7). This completes the proof. ■

4. TECHNICAL DETAILS

In this chapter, we prove a lemma and provide the details for the examples introduced in the previous sections.

4.1 A useful lemma

In this section, we give a useful lemma which can be used to study the regularity of a dispersion matrix when its dimension tends to infinity with sample size.

For $a \in \mathcal{R}$, $\mathbf{c}_r \in \mathcal{R}^r$ and $r \times r$ identity matrix $\mathbb{I}_r$, let $\mathbb{M}_{r+1}$ be the $(r+1) \times (r+1)$ matrix defined by $\mathbb{M}_{r+1} = \mathscr{W}(a^2, \mathbf{c}_r, \mathbb{I}_r)$, where $\mathscr{W}$ is the matrix operation defined in (2.1.14). Denote the determinant of $\mathbb{M}$ by $|\mathbb{M}|$. Using Laplace's formula to express the determinant of a matrix in terms of its minors and the mathematical induction we can easily prove (4.1.1).

Lemma 4.1.1 *For $\lambda \in \mathcal{R}$ and integer $r \geq 1$, the characteristic polynomial of $\mathbb{M}_{r+1}$ is given by*

$$|\mathbb{M}_{r+1} - \lambda \mathbb{I}_{r+1}| = (1-\lambda)^{r-1}\big(\lambda^2 - (1+a^2)\lambda + a^2 - \|\mathbf{c}_r\|^2\big). \qquad (4.1.1)$$

Thus the sequence of matrices $\mathbb{M}_{r+1}$ is regular if $c^2 = \lim_{r\to\infty} \|\mathbf{c}_r\|^2 = \sum_{i=1}^{\infty} c_i^2 < \infty$ such that $b^2 = a^2 - c^2 > 0$.

Proof We only need show the regularity. Obviously $\lambda = 1$ is a root of multiplicity $r-1$. The other two roots are the two roots the quadratic expression on the right side, which are given by

$$\lambda_1 = (1 + a^2 + \sqrt{\Delta})/2, \quad \lambda_2 = (1 + a^2 - \sqrt{\Delta})/2,$$

where $\Delta = (a^2-1)^2 + 4\sum_{i=1}^r c_i^2$. Since $0 \le \Delta \le \delta := (a^2-1)^2 + 4c^2$, it follows $0 < 1 + a^2 \le 2\lambda_1 \le 1 + a^2 + \sqrt{\delta} < \infty$ and

$$0 < 2b^2/(1 + a^2 + \sqrt{(1+a^2)^2 + 4b^2}) \le \lambda_2 \le (1+a^2)/2 < \infty.$$

This shows that $\mathbb{M}_{r+1}$ has $r+1$ eigen values which bounded away from both zero and infinity uniformly in $r = 1, 2, \ldots$, hence the sequence of matrices $\mathbb{M}_{r+1}$ is regular. ■

Remark 4.1.1 Let $\mathcal{H}_1$ and $\mathcal{H}$ be two Hilbert spaces such that $\mathcal{H}_1$ is a true subspace of $\mathcal{H}$. Let $a_k : k = 1, 2, \ldots$ be an orthonormal basis of $\mathcal{H}_1$. For $\varphi \in \mathcal{H}$, the projection φ_p of φ onto $\mathcal{H}_1$ is given by the Fourier series $\varphi_p = \sum_{k=1}^\infty c_k a_k$, where c_k are the Fourier coefficients. Suppose $\varphi \notin \mathcal{H}_1$. By the Hilbert space theory (see e.g. Theorem 4.13, Conway (1985)), $\|\varphi_p\|^2 = \sum_{k=1}^\infty c_k^2 < \|\varphi\|^2$. Since a_k is orthonormal, the $r \times r$ matrix whose (i,j)-entry is the inner product of a_i, a_j is the $r \times r$ identity matrix $\mathbb{I}_r$. Consequently, it follows from Lemma 4.1.1 that the sequence of matrices $\mathscr{W}(\|\varphi\|^2, \mathbf{c}_r, \mathbb{I}_r), r = 1, 2, \ldots$ is regular.

4.2 Proofs

In this section, we collect the proofs for the examples.

Let $\boldsymbol{\phi}_n = (\phi_1, \ldots, \phi_{r_n})^\top$ where ϕ_k is the trigonometric basis given in (2.1.4). Since these basis functions are bounded by $\sqrt{2}$, we see that $\boldsymbol{\phi}(t)$ and its derivative $\boldsymbol{\phi}'(t)$ satisfy

$$\|\boldsymbol{\phi}_n(t)\|^2 \le 2r_n, \quad \|\boldsymbol{\phi}'(t)\|^2 \le 2\pi^2 r_n^3, \quad t \in [0,1]. \tag{4.2.1}$$

Denote $\mathbf{a} \otimes \mathbf{b}$ the Kronecker product of vectors $\mathbf{a}$ and $\mathbf{b}$.

Proof of (2.1.6). We shall prove this by applying Theorem 3.3.2 with $\mathbf{g}_n(x,t) = \boldsymbol{\phi}_n \circ G(x) \otimes \boldsymbol{\phi}_n \circ F(t)$ and $\hat{\mathbf{g}}_n(x,t) = \boldsymbol{\phi}_n \circ \mathbb{G}(x) \otimes \boldsymbol{\phi}_n \circ \mathbb{F}(t)$, where $F(t) = F_{\alpha_0,\beta_0}(t) := P(\epsilon \le t) = P(Y - \alpha_0 - \beta_0 X \le t)$ and $\mathbb{F}(t) = \mathbb{F}_{\alpha_0,\beta_0}(t)$ for $t \in \mathcal{R}$ with α_0 denoting the true value of parameter. As discussed in constructing the empirical likelihood (2.1.6), it is without loss of generality to assume that the intercept α_0 is zero. Thus

$F(t) = F_{0,\beta_0}(t) = P(Y - \beta_0 X \le t)$ and $\mathbb{F}(t) = \mathbb{F}_{0,\beta_0}(t) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}[Y_j - \beta_0 X_j \le t]$ for $t \in \mathcal{R}$. Since the trigonometric basis ϕ_k is bounded by $\sqrt{2}$, it follows that $\|\boldsymbol{\phi}(t)\| \le \sqrt{2r_n}$, hence $\|\hat{\mathbf{g}}_n(x,t)\| = \|\boldsymbol{\phi}_n \circ G(x)\| \|\boldsymbol{\phi}_n \circ F(t)\| \le 2r_n$, and of course also $\|\mathbf{g}_n(x,t)\| \le 2r_n$. Thus one obtains (3.3.6) as $r_n^4 = o(n)$ implied by $r_n^6 = o(n)$. Since $h_1(x,y) = P((y-Y)/(x-X) \le \beta_0) \le 1$, it follows that $r_n h_1$ is Lindeberg as $r_n^2 = o(n)$. Note that $\mathbb{W}_n = \int \mathbf{g}_n \mathbf{g}_n^\top \, dQ = I_{r_n^2}$. As proved in Example 4 of Peng and Schick (2013c), the second equality in (3.3.7) and (3.3.8) hold with $\mathbf{u}_n = \mathbf{g}_n$ as $r_n^6 = o(n)$. Clearly $\int \mathbf{u}_n \, dQ = 0$ and $\|\mathbf{u}_n\|$ is Lindeberg as it is bounded by $2r_n$ and $r_n^2 = o(n)$. We shall show below that the first equality in (3.3.7) holds. Thus $\mathbf{v}_n = (m\tilde{h}_1, \mathbf{g}_n^\top)^\top$ and $\int \mathbf{v}_n \mathbf{v}_n^\top \, dQ = \mathscr{W}_n$. This in turn implies that $\mathbb{U}_n$ in (3.3.9) satisfies $\mathbb{U}_n = I_{r_n+1}$, so that $|\mathbb{U}_n|_o = 1 = O(1)$ and $r_n/\text{trace}(\mathbb{U}_n^2) = r_n/(r_n+1) = O(1)$. Let us now prove that the sequence of matrices $\mathscr{W}_n$ satisfies (C). Recall we take $\alpha_0 = 0$ so $\epsilon = Y - \beta_0 X$. Let $\mathcal{H} = L_{2,0}(D)$, where $D = FG$ is the joint distribution of X and ϵ. Let $\mathcal{H}_1 = \{a(X)b(\epsilon) : a \in L_{2,0}(G), b \in L_{2,0}(F)\}$. Then clearly $\mathcal{H}_1$ is a true subspace of $\mathcal{H}$. Under the null hypothesis, $Y_2 = \beta_0 X_2 + \epsilon_2$. Using this, we find

$$\begin{aligned} h_1(X,Y) &= E\big(P((Y - Y_2)/(X - X_2) \le \beta_0 | X, Y)\big) \\ &= 1 - F(Y - \beta_0 X). \end{aligned}$$

and $\tilde{h}_1(X,Y) = h_1(X,Y) - 1/2 = 1/2 - F(Y - \beta_0 X) = 1/2 - F(\epsilon)$ in view of $Y = \beta_0 X + \epsilon$ under the null. Clearly $\tilde{h}_1 \in \mathcal{H}$. Since nonzero constants do not live either in $L_{2,0}(G)$ or $L_{2,0}(F)$, it follows that neither $a \in L_{2,0}(G)$ nor $b \in L_{2,0}(F)$ for which $a \not\equiv 0$ and $b \not\equiv 0$ belongs to $\mathcal{H}_1$, hence $\tilde{h}_1 \notin \mathcal{H}_1$. Consequently it follows from Remark 4.1.1 that $\mathscr{W}_n$ satisfies (C).

To complete the proof, we show the first equality in (3.3.7). To this end, we use Cauchy inequality to bound the euclidean norm on the left hand side of (3.3.7) by $A_n + B_n + C_n$, where

$$A_n = \|\frac{1}{n}\sum_{j=1}^{n} \tilde{V}_{nj}\big(\hat{\mathbf{g}}_n(X_j,\epsilon_j) - \mathbf{g}_n(X_j,\epsilon_j)\big)\|,$$

$$B_n = \|\frac{1}{n}\sum_{j=1}^{n} \big(\tilde{V}_{nj} - m\tilde{h}_1(\mathbf{Z}_j)\big)\mathbf{g}_n(X_j,\epsilon_j)\|,$$

$$C_n = \|\frac{1}{n}\sum_{j=1}^{n} m\tilde{h}_1(\mathbf{Z}_j)\mathbf{g}_n(X_j,\epsilon_j) - \int m\tilde{h}_1\mathbf{g}_n\, dQ\|,$$

where $\mathbf{Z}_j = (X_j, Y_j)^\top$, $\epsilon_j = Y_j - \beta_0 X_j$ and $m = 2$. Denote $\mathbf{Z}$ an i.i.d. copy of $\mathbf{Z}_1$. We bound the variance by the second moment to get

$$nE(C_n^2) \le m^2 E\big(\tilde{h}_1(\mathbf{Z}_1)^2\|\mathbf{g}_n(X,\epsilon)\|^2\big) \le m^2 E\big(\tilde{h}_1(\mathbf{Z}_1)^2\big)(4r_n^2),$$

so that $C_n = o_p(r_n^{-1/2})$ follows from $r_n^3 = o(n)$. Now by Cauchy inequality and in view of (1.2.3),

$$B_n^2 \le \frac{1}{n}\sum_{j=1}^{n} \big(\tilde{V}_{nj} - m\tilde{h}_1(\mathbf{Z}_j)\big)^2 \frac{1}{n}\sum_{j=1}^{n} \|\mathbf{g}_n(X_j,\epsilon_j)\|^2 = O_p(n^{-1})(2r_n^2),$$

so that $B_n = o_p(r_n^{-1/2})$. Finally, again by Cauchy inequality,

$$A_n^2 \le \frac{1}{n}\sum_{j=1}^{n} \tilde{V}_{nj}^2 \,\frac{1}{n}\sum_{j=1}^{n} \|\hat{\mathbf{g}}_n(X_j,\epsilon_j) - \mathbf{g}_n(X_j,\epsilon_j)\|^2 = O_p(1)(A_{1n} + A_{2n}),$$

where

$$A_{1n} = \frac{1}{n}\sum_{j=1}^{n} \|\boldsymbol{\phi}_n(\mathbb{G}(X_j)) - \boldsymbol{\phi}_n(G(X_j))\|^2\, \|\boldsymbol{\phi}_n(\mathbb{F}(\epsilon_j))\|^2,$$

$$A_{2n} = \frac{1}{n}\sum_{j=1}^{n} \|\boldsymbol{\phi}_n(G(X_j))\|^2\, \|\boldsymbol{\phi}_n(\mathbb{F}(\epsilon_j)) - \boldsymbol{\phi}_n(F(\epsilon_j))\|^2.$$

By (4.2.1),

$$A_{1n} \le 2\pi^2 r_n^3 \sup_{-\infty<x<\infty} |\mathbb{G}(x) - G(x)|^2(2r_n) = O_p(r_n^4/n).$$

Thus $A_{1n} = o_p(r_n^{-1})$ follows from $r_n^5 = o(n)$. In a similar fashion, $A_{2n} = o_p(r_n^{-1})$. Hence we conclude $A_n = o_p(r_n^{-1/2})$ and apply the result of Theorem 3.3.2 to complete the proof of (2.1.6). ■

Proof of (2.1.10). We shall use Theorem 3.2.2 to prove the result. Let us start with the proof of (3.2.4). Under (W1) and (W2), one can verify conditions (i) and (ii) of Theorem 1 of Arcones (1996) are met, so that his (2.4) holds, that is,

$$U_n(\hat{\theta}_n) = U_n(\theta_0) + \bar{F}_2(\hat{\theta}_n) - \bar{F}_2(\theta_0) + o_p(n^{-1/2}). \tag{4.2.2}$$

By the Hoeffding decomposition, we obtain

$$U_n(\theta_0) - 1/2 = \frac{1}{n}\sum_{j=1}^{n} 2\big(\bar{F}(2\theta_0 - X_j) - 1/2\big) + o_p(n^{-1/2}). \tag{4.2.3}$$

Since f is bounded away from zero in a neighborhood of θ_0, the usual median estimator $\hat{\theta}_n$ satisfies the stochastic expansion

$$\hat{\theta}_n = \theta_0 - \frac{1}{n}\sum_{j=1}^{n} \frac{\mathbf{1}[X_j \le \theta_0] - 1/2}{f(\theta_0)} + o_p(n^{-1/2}). \tag{4.2.4}$$

This of course implies $\hat{\theta}_n - \theta_0 = O_p(n^{-1/2})$. Moreover, in view of (2.1.9),

$$\bar{F}_2(\hat{\theta}_n) - \bar{F}_2(\theta_0) = \frac{f_2(\theta_0)}{f(\theta_0)}\frac{1}{n}\sum_{j=1}^{n}\big(\mathbf{1}[X_j \le \theta_0] - 1/2\big) + o_p(n^{-1/2}). \tag{4.2.5}$$

Thus from (4.2.2)-(4.2.5) it immediately follows

$$U_n(\hat{\theta}_n) - 1/2 = \frac{1}{n}\sum_{j=1}^{n} v(X_j) + o_p(n^{-1/2}), \tag{4.2.6}$$

where

$$v(X_j) = 2\bar{F}(2\theta_0 - X_j) - 1 + \frac{f_2(\theta_0)}{f(\theta_0)}\big(\mathbf{1}[X_j \le \theta_0] - 1/2\big), \quad j = 1, \ldots, n.$$

Clearly $\int v\, dQ = 0$. These establish (3.2.4) and hence (3.2.5) holding with $\mathcal{T}$ being normally distributed with mean zero and variance $\sigma^2 = \mathrm{Var}(v(X_1))$.

To simplify our notation, let $\hat{V}_{nj} = V_{nj}(\hat{\theta}_n)$ and $\hat{U}_n = U_n(\hat{\theta}_n)$. By the definition of jackknife pseudo values and using some algebra, we derive

$$\begin{aligned}
\hat{V}_{nj} &= nU_n(\hat{\theta}_n) - (n-1)U_{n-1}^{(-j)}(\hat{\theta}_n) \\
&= \frac{2}{n-1}\sum_{1\le i<k\le n} \mathbf{1}[X_i + X_k > 2\hat{\theta}_n] - \frac{2}{n-2}\sum_{1\le i<k\le n}^{(-j)} \mathbf{1}[X_i + X_k > 2\hat{\theta}_n] \\
&= -\frac{n}{n-2}\hat{U}_n + \frac{n-1}{n-2} 2A_{nj}(2\hat{\theta}_n - X_j) \\
&= -\hat{U}_n + 2A_{nj}(2\hat{\theta}_n - X_j) + O_p(1/n), \quad j = 1, \ldots, n, \qquad (4.2.7)
\end{aligned}$$

where

$$A_{nj}(x) = \frac{1}{n-1}\sum_{i:i\neq j}\mathbf{1}[X_i > x], \quad x \in \mathcal{R}.$$

We shall show below that for $j = 1, \ldots, n$,

$$A_{nj}(2\hat{\theta}_n - X_j) = A_{nj}(2\theta_0 - X_j) - 2f(2\theta_0 - X_j)(\hat{\theta}_n - \theta_0) + o_p(n^{-1/2}). \qquad (4.2.8)$$

This and (4.2.4) then give

$$A_{nj}(2\hat{\theta}_n - X_j) - 1/2 = \frac{1}{n-1}\sum_{i:i\neq j} w(X_i|X_j) + o_p(n^{-1/2}), \qquad (4.2.9)$$

where

$$w(X_i|X_j) = \mathbf{1}[X_i + X_j > 2\theta_0] - 1/2 + \frac{2f(2\theta_0 - X_j)}{f(\theta_0)}(\mathbf{1}[X_i \leq \theta_0] - 1/2).$$

From (4.2.6)-(4.2.9) it follows that for $j = 1, \ldots, n$,

$$\hat{V}_{nj} - 1/2 = \frac{1}{n-1}\sum_{i:i\neq j}\big(-v(X_i) + 2w(X_i|X_j)\big) + o_p(n^{-1/2}). \qquad (4.2.10)$$

We prove below that

$$\frac{1}{n}\sum_{j=1}^{n}(\hat{V}_{nj} - 1/2)^2 = \varsigma^2 + o_p(1). \qquad (4.2.11)$$

This shows that the second statement of (3.2.3) holds with $\mathscr{W} = \varsigma^2$, while the first statement of (3.2.3) follows from (4.2.7) and the inequalities $|\hat{U}_n| \leq 1$ and $|A_{nj}(2\hat{\theta}_n - X_j)| \leq 1$. We now apply the result of Theorem 3.2.2 to conclude the desired (2.1.10).

Proof of (4.2.11). Let $\xi_{nj} = \hat{V}_{nj} - 1/2$ and $u(x, x_1) = -v(x) + 2w(x|x_1)$, $x, x_1 \in \mathcal{R}$, where v, w are given in (4.2.10). Apparently, for each fixed j, $E(\xi_{nj}^2) = E(\xi_{n1}^2)$ at least as n tends to infinity. So let us first calculate $E(\xi_{n1}^2)$. Our approach is to find the dominating terms of $E(\xi_{n1}^2)$. Note first that

$$\begin{aligned} E(\xi_{n1}^2) &= E\Big(\Big[\frac{1}{n-1}\sum_{i:i\neq 1} u(X_i, X_1)\Big]^2\Big) + o(1) \\ &= \frac{1}{n^2}\sum_{i,k:i\neq k, i,k>1} E\big(u(X_i, X_1)u(X_k, X_1)\big) + o(1) \\ &= E\big(u(X_2, X_1)u(X_3, X_1)\big) + o(1) \end{aligned}$$

Denote $\tilde{\mathbf{1}}[X_2 + X_1 > \theta_0] = \mathbf{1}[X_2 + X_1 > \theta_0] - 1/2$. Then under the null hypothesis $E(\tilde{\mathbf{1}}[X_2 + X_1 > \theta_0]) = 0$. Note $E(v(X_2)) = 0$ and observe

$$u(X_2, X_1) = -v(X_2) + 2\tilde{\mathbf{1}}[X_2 + X_1 > 2\theta_0] + \frac{4f(2\theta_0 - X_1)}{f(\theta_0)}\tilde{\mathbf{1}}[X_2 \leq \theta_0],$$

we find that the only nonzero term in the expansion of $E\big(u(X_2, X_1)u(X_3, X_1)\big)$ is $4E(\tilde{\mathbf{1}}[X_2 + X_1 > 2\theta_0]\tilde{\mathbf{1}}[X_3 + X_1 > 2\theta_0])$, which is equal to ς^2 given in (W2). Hence

$$E\big(u(X_2, X_1)u(X_3, X_1)\big) = \varsigma^2 + o(1). \tag{4.2.12}$$

To show (4.2.11), it suffices to show the following second moment converges to zero. To this end, we proceed to calculate the nonzero terms as follows:

$$\begin{aligned}
&E\Big(|\frac{1}{n-1}\sum_{i:i\neq 1}(\xi_{ni}^2 - E(\xi_{ni}^2))|^2\Big) = E\big((\xi_{n1}^2 - E(\xi_{n1}^2))(\xi_{n2}^2 - E(\xi_{n2}^2))\big) + o(1)\\
&= E\big(\xi_{n1}^2\xi_{n2}^2\big) - \varsigma^4 + o(1)\\
&= n^{-4}E\Big(\big(\sum_{j\neq 1} u(X_j, X_1)\big)^2\big(\sum_{k\neq 2} u(X_k, X_2)\big)^2\Big) - \varsigma^4 + o(1)\\
&= n^{-4}E\Big(\big(\sum_{j\neq 1,2} u(X_j, X_1) + u(X_2, X_1)\big)^2\big(\sum_{k\neq 1,2} u(X_k, X_2) + u(X_1, X_2)\big)^2\Big) - \varsigma^4 + o(1)\\
&= n^{-4}E\Big(\big(\sum_{j\neq 1,2} u(X_j, X_1)\big)^2\big(\sum_{k\neq 1,2} u(X_k, X_2))\big)^2\Big) - \varsigma^4 + o(1)
\end{aligned}$$

Let E_{12} denote the conditional expectation given X_1, X_2. Then we continue to compute the main terms,

$$\begin{aligned}
&n^{-4}E\Big(\big(\sum_{j\neq 1,2} u(X_j, X_1)\big)^2\big(\sum_{k\neq 1,2} u(X_k, X_2)\big)^2\Big)\\
&= n^{-4}\sum_{i,j\neq 1,2}\sum_{k,l\neq 1,2} E_{12}\big(u(X_i, X_1)u(X_j, X_1)u(X_k, X_2)u(X_l, X_2)\big)\\
&= E\big(u(X_3, X_1)u(X_4, X_1)u(X_5, X_2)u(X_6, X_2)\big) + o(1)\\
&= E\Big(E_{12}\big(u(X_3, X_1)u(X_4, X_1)\big)E_{12}\big(u(X_5, X_2)u(X_6, X_2)\big)\Big) + o(1)\\
&= E\big((\bar{F}(2\theta_0 - X_1) - 1/2)^2\big)E\big((\bar{F}(2\theta_0 - X_2) - 1/2)^2\big) + o(1)\\
&= \varsigma^4 + o(1).
\end{aligned}$$

From the combination of the above statements it immediately yields the desired (4.2.11).

Proof of (4.2.8). We shall prove this for $j = 1$. To this end, let $A_n(\theta) = A_{n1}(2\theta - X_1)$, $A(\theta) = \bar{F}(2\theta - X_1)$ and $\theta_n(t) = \theta_0 + n^{-1/2}t, t \in \mathcal{R}$. Thus it suffices to prove that for every finite $C > 0$ and $\varepsilon > 0$,

$$P\Big(\sup_{|t|\le C} |A_n\big(\theta_n(t)\big) - A_n(\theta_0) - A\big(\theta_n(t)\big) + A(\theta_0)| > \varepsilon n^{-1/2}|X_1\Big) \xrightarrow{P} 0. \qquad (4.2.13)$$

We shall employ the Bernstein's inequality to reach this goal. To this end, write P_1 for the conditional probability given X_1 so that $P_1() = P(|X_1)$, and $E_1 = E(|X_1)$ for the conditional expectation and etc. Partition $[-C, C]$ into equal K subintervals $[t_0, t_1]$, $(t_{k-1}, t_k], k = 1, \ldots, K$ with $t_k = -C + 2Ck/K$. Let $\theta_{nk} = \theta_n(t_k)$ and set

$$B_n(t) = A_n\big(\theta_n(t)\big) - A_n(\theta_0) - A\big(\theta_n(t)\big) + A(\theta_0).$$

Then we have $B_n(t) = (n-1)^{-1}\sum_{j\neq 1} b_{nj}(t)$, where

$$b_{nj}(t) = \mathbf{1}[Y_j > \theta_n(t)] - \mathbf{1}[Y_j > \theta] - \bar{F}(2\theta_n(t) - X_1) + \bar{F}(2\theta_0 - X_1)$$

with $Y_j = (X_j + X_1)/2$. Clearly $E_1(b_{nj}(t_k)) = 0$ and $\sigma^2_{nk} = \mathrm{Var}_1(b_{nj}(t_k))$ satisfies

$$\sigma^2_{nk} = \bar{F}(2\theta_{nk} - X_1) + \bar{F}(2\theta_0 - X_1) - 2\bar{F}(2\max(\theta_{nk}, \theta_0) - X_1) + o_p(n^{-1}),$$

in view of (2.1.7) and the inequality

$$|\bar{F}(2\theta_{nk} - X_1) - \bar{F}(2\theta_0 - X_1)| \le 3BCn^{-1/2},$$

which is implied by

$$\bar{F}(2\theta_{nk} - X_1) = \bar{F}(2\theta_0 - X_1) - f(2\theta_0 - X_1)2(\theta_{nk} - \theta_0) + o_p(n^{-1/2}),$$

where B is the upper bound of f in (W1). Again by (2.1.7), we derive

$$\bar{F}(2\max(\theta_{nk}, \theta_0) - X_1) = \bar{F}(2\theta_0 - X_1) - f(2\theta_0 - X_1)2\delta_{nk}, + o_p(n^{-1/2}).$$

where $\delta_{nk} = \max(\theta_{nk}, \theta_0) - \theta_0$. Obviously $|\delta_{nk}| \le n^{-1/2}|t_k|$. Combining the above, we get

$$\begin{aligned}\sigma_{nk}^2 &= f(2\theta - X_1)(4\delta_{nk} - 2n^{-1/2}t_k) + o_p(n^{-1/2})\\ &\le f(2\theta - X_1)2n^{-1/2}|t_k| + o_p(n^{-1/2})\\ &\le 2BCn^{-1/2} + o_p(n^{-1/2})\\ &\le 3BCn^{-1/2}, \quad \text{say.}\end{aligned}$$

Now applying the Bernstein's inequality (see e.g. (a) (with $m = 1$) of Proposition 2.3 of Arcones and Giné (1993)) and noticing $|b_{nk}| \le 1$, we arrive at

$$\begin{aligned}P_1\big(|B_n(t_k)| > \varepsilon n^{-1/2}/2\big) &\le \exp\Big(-\frac{\varepsilon^2}{8\sigma_{nk}^2 + (8/3)\varepsilon n^{-1/2}}\Big)\\ &\le \exp\Big(-\frac{n^{1/2}\varepsilon^2}{24BC + (8/3)\varepsilon}\Big), \quad k = 1, \dots, K.\end{aligned}$$

Denote the probability in (4.2.13) by $p_n(\varepsilon, C)$. Then we have

$$\begin{aligned}p_n(\varepsilon, C) &\le \sum_{k=1}^{K} P_1\big(|B_n(t_k)| > \varepsilon n^{-1/2}/2\big)\\ &\quad + \sum_{k=1}^{K} P_1\Big(\sup_{t\in(t_{k-1},t_k]} |b_{nj}(t) - b_{nj}(t_k)| > \varepsilon n^{-1/2}/2\Big)\\ &\le K\exp\Big(-\frac{n^{1/2}\varepsilon^2}{24BC + (8/3)\varepsilon}\Big) + q_n(\varepsilon),\end{aligned} \tag{4.2.14}$$

where $q_n(\varepsilon)$ denotes the preceding last sum. Note that for $t \in (t_{k-1}, t_k]$,

$$\begin{aligned}|b_{nj}(t) - b_{nj}(t_k)| &\le |\mathbf{1}[Y_j > \theta_n(t)] - \mathbf{1}[Y_j > \theta_n(t_k)]|\\ &\quad + |\bar{F}(2\theta_n(t) - X_1) - \bar{F}(2\theta_n(t_k) - X_1)|\\ &\le 1[Y_j \in I_{nk}] + 4BCn^{-1/2}/K + o_p(n^{-1/2})\\ &\le 1[Y_j \in I_{nk}] + 5BCn^{-1/2}/K, \quad \text{say.}\end{aligned}$$

where $I_{nk} = (\theta_n(t_{k-1}), \theta_n(t_k)]$. Hence,

$$q_n(\varepsilon) \le \sum_{k=1}^{K} P_1\Big(1[Y_j \in I_{nk}] > \varepsilon n^{-1/2}/4\Big) + K\mathbf{1}[5BCn^{-1/2}/K > \varepsilon n^{-1/2}/4].$$

Denote the above last sum by $r_n(\varepsilon)$. Similarly, we derive

$$P_1(Y_j \in I_{nk}) \le 4BCn^{-1/2}/K.$$

Hence using $\text{Var}_1(1[Y_j \in I_{nk}]) \le P_1(Y_j \in I_{nk})$ and by the Bernstein's inequality, we get

$$\begin{aligned} r_n(\varepsilon) &\le \sum_{k=1}^{K} P_1\Big(1[Y_j \in I_{nk}] - P_1(Y_j \in I_{nk}) > \varepsilon n^{-1/2}/8\Big) \\ &\quad + K\mathbf{1}[4BCn^{-1/2}/K > \varepsilon n^{-1/2}/8] \\ &\le K \exp\Big(-\frac{n^{1/2}\varepsilon^2}{512BC/K + (16/3)\varepsilon}\Big) + K\mathbf{1}[32BC/K > \varepsilon]. \end{aligned}$$

Now by choosing $K = \log n$ we see that $r_n(\varepsilon) \to 0$, $q_n(\varepsilon) \to 0$, $p_n(C, \varepsilon) \to 0$ and hence the desired (4.2.13) as $n \to \infty$ for every $C > 0$ and $\varepsilon > 0$. ∎

Proof of (2.2.4). Recall that $\mathbf{X}_i$ is defined in (2.2.2), the residuals $\hat{\varepsilon}_i = Y_{i\cdot} - Y_{\cdot\cdot}$ and ε is an i.i.d. copy of $\varepsilon_i = u_i + \epsilon_{i\cdot}$ for $i = 1, \ldots, n$. Using the same notation as in subsection 2.2, we first prove the following lemma.

Lemma 4.2.1 *Suppose* $\mathbf{g}$ *is a measurable function from* $\mathcal{R}$ *to* $\mathcal{R}^r$ *satisfying* $E(\mathbf{g}(\varepsilon)) = 0$ *and* $E(\|\mathbf{g}(\varepsilon)\|^2) < \infty$ *such that*

$$\max_{1 \le j \le n} \|g(\hat{\varepsilon}_i)\| = o_p(n^{1/2}), \tag{4.2.15}$$

$$\frac{1}{n}\sum_{i=1}^{n} \|\mathbf{g}(\hat{\varepsilon}_i) - \mathbf{g}(\varepsilon_i)\|^2 = o_p(1), \tag{4.2.16}$$

and $\mathscr{W} = \mathscr{W}(\text{Var}(2\mathbf{h}_1(\mathbf{X}_1)), \mathbb{C}, \mathbb{W})$ *is non-singular for* $\mathbb{C} = E(2\mathbf{h}_1(\mathbf{X}_1) \otimes \mathbf{g}(\varepsilon))$ *and* $\mathbb{W} = E(\mathbf{g}^{\otimes 2}(\varepsilon))$, *and that*

$$n^{-1/2}\sum_{i=1}^{n} \mathbf{g}(\hat{\varepsilon}_i) = n^{-1/2}\sum_{i=1}^{n} \mathbf{u}(\varepsilon_i) + o_p(1) \tag{4.2.17}$$

for some measurable function $\mathbf{u}$ *from* $\mathcal{R}$ *to* $\mathcal{R}^r$ *satisfying* $E(\mathbf{u}(\varepsilon)) = 0$ *and* $E(\|\mathbf{u}(\varepsilon)\|^2) < \infty$. *Then*

$$-2\log \mathscr{R}_n(\mathbf{h}, \mathbf{g}) \Longrightarrow \mathbf{Z}^\top \mathscr{W}^{-1}\mathbf{Z}, \tag{4.2.18}$$

where $\mathbf{Z}$ *is normally distributed with vector mean zero and variance-covariance matrix equal to* $\text{Var}\big((2\mathbf{h}_1(\mathbf{X})^\top, \mathbf{u}(\varepsilon)^\top)^\top\big)$.

Proof We shall prove the result by an application of Corollary 3.2.2 with $\hat{g}(\mathbf{X}_i) = g(\hat{\varepsilon}_i)$ and $Z_i = \mathbf{X}_i$. Obviously (4.2.15) corresponds to (3.2.6). We now show that (4.2.16) implies (3.2.7). In fact,

$$\begin{aligned}\|\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{V}}_{ni}(\mathbf{h})\otimes\mathbf{g}(\hat{\varepsilon}_i)-\mathbb{C}\| &\le \|\frac{1}{n}\sum_{i=1}^{n}\tilde{\mathbf{V}}_{ni}(\mathbf{h})\otimes\big(\mathbf{g}(\hat{\varepsilon}_i)-\mathbf{g}(\varepsilon_i)\big)\| \\ &+\|\frac{1}{n}\sum_{i=1}^{n}\big(\tilde{\mathbf{V}}_{ni}(\mathbf{h})-2\tilde{\mathbf{h}}_1(\mathbf{X}_i)\big)\otimes\mathbf{g}(\varepsilon_i)\|+\|\frac{1}{n}\sum_{i=1}^{n}2\tilde{\mathbf{h}}_1(\mathbf{X}_i)\otimes\mathbf{g}(\varepsilon_i)-\mathbb{C}\|.\end{aligned}$$

By the law of large numbers, the last term is $o_p(1)$. We bound the square of the first term on the above right side by

$$\frac{1}{n}\sum_{i=1}^{n}\|\tilde{\mathbf{V}}_{ni}\|^2\,\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{g}(\hat{\varepsilon}_i)-\mathbf{g}(\varepsilon_i)\|^2,$$

which is $o_p(1)$ by (4.2.16) and the equality $\frac{1}{n}\sum_{i=1}^{n}\|\tilde{\mathbf{V}}_{ni}\|^2 = O_p(1)$ implied by the square-integrability of the jackknife pseudo values of U-statistics, while by Cauchy inequality the square of the middle term on the above right side is bounded by

$$\frac{1}{n}\sum_{i=1}^{n}\big\|\tilde{\mathbf{V}}_{ni}(\mathbf{h})-2\tilde{\mathbf{h}}_1(\mathbf{X}_i)\big\|^2\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{g}(\varepsilon_i)\|^2,$$

which is $o_p(1)$ as the last average is $O_p(1)$ by the square-integrability of $\mathbf{g}$ whereas the first average is $o_p(1)$ as an component wise application of (3.1.6). This shows the first equality of (3.2.7). To show the second equality of (3.2.7), we need the inequality

$$\begin{aligned}|\mathbf{S}_{x+y}-\mathbf{S}_x|_o &\le |\mathbf{S}_y|_o + 2|\mathbf{S}_x|_o^{1/2}|\mathbf{S}_y|_o^{1/2} \\ &\le \frac{1}{n}\sum_{i=1}^{n}\|\mathbf{y}_i\|^2 + 2|\mathbf{S}_x|_o^{1/2}\Big(\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{y}_i\|^2\Big)^{1/2}\end{aligned} \tag{4.2.19}$$

with

$$\mathbf{S}_{x+y} = \frac{1}{n}\sum_{j=1}^{n}(\mathbf{x}_j+\mathbf{y}_j)^{\otimes 2}, \quad \mathbf{S}_x = \frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_j^{\otimes 2}, \quad \mathbf{S}_y = \frac{1}{n}\sum_{j=1}^{n}\mathbf{y}_j^{\otimes 2}$$

for vectors $\mathbf{x}_1, \mathbf{y}_1, \ldots, \mathbf{x}_n, \mathbf{y}_n$ of the same dimension, see Peng and Schick (2013c). Thus

$$\begin{aligned}|\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}^{\otimes 2}(\hat{\varepsilon}_i)-\mathbf{g}^{\otimes 2}(\varepsilon_i)|_o &\le \frac{1}{n}\sum_{i=1}^{n}\|\mathbf{g}(\hat{\varepsilon}_i)-\mathbf{g}(\varepsilon_i)\|^2 \\ &+ 2|\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}^{\otimes 2}(\varepsilon_i)|_o^{1/2}\Big(\frac{1}{n}\sum_{i=1}^{n}\|\mathbf{g}(\hat{\varepsilon}_i)-\mathbf{g}(\varepsilon_i)\|^2\Big)^{1/2}.\end{aligned}$$

This yields the second equality of (3.2.7) in view of (4.2.16), the limit

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}^{\otimes 2}(\varepsilon_i)\xrightarrow{P}\mathbb{W}$$

by the law of large of numbers, and the inequality

$$|E\big(\mathbf{g}^{\otimes 2}(\varepsilon)\big)|_o \le E\big(\|\mathbf{g}(\varepsilon)\|^2\big) = \mathrm{trace}\big(E\big(\mathbf{g}^{\otimes 2}(\varepsilon)\big)\big) = \mathrm{trace}\big(\mathbb{W}\big) < \infty,$$

where the first inequality of the last line follows from an application of (3.1.1). By the property of jackknife pseudo values and the Hoeffding decomposition, we have

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{V}_{ni}(\mathbf{h})-\boldsymbol{\theta}=\mathbf{U}_n(\mathbf{h})-\boldsymbol{\theta}=\frac{1}{n}\sum_{i=1}^{n}2\tilde{\mathbf{h}}_1(\mathbf{X}_i)+o_p(n^{-1/2}).$$

This and (4.2.17) yield (3.2.8). We now apply the conclusion (3.2.9) of Corollary 3.2.2 to claim the desired (4.2.18). ∎

We now use Lemma 4.2.1 to prove (2.2.4). Since ε has median zero with a density and $g(t) = \mathrm{sign}(t)$ is bounded by one, it follows that g is square-integrable with $\int g\,dQ = 0$ and (4.2.15) holds. Let G be the distribution function of $|\varepsilon|$ and G_n be its empirical distribution function. Then it is well known

$$\sup_{t\ge 0}|G_n(t)-G(t)| = o_p(1).$$

Since $\hat{\mu}\xrightarrow{p}\mu$ and G is continuous, it follows

$$G(|\hat{\mu}-\mu|) = o_p(1).$$

Clearly $|\mathbf{1}[Y_{i\cdot}\le\hat{\mu}]-\mathbf{1}[Y_{i\cdot}\le\mu]|\le\mathbf{1}[|Y_{i\cdot}-\mu|\le|\hat{\mu}-\mu|]$. Thus we obtain

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}|g(\hat{\varepsilon}_i)-g(\varepsilon_i)|^2 &\le 4\frac{1}{n}\sum_{i=1}^{n}|\mathbf{1}[|Y_{i\cdot}-\mu|\le|\hat{\mu}-\mu|]| \\ &\le \sup_{t\ge 0}|G_n(t)-G(t)|+G(|\hat{\mu}-\mu|) = o_p(1).\end{aligned}$$

This proves (4.2.16). We are left to show that (4.2.17) holds with $u(t) = \mathrm{sign}(t) - 2f(0)t$. Note first that $\hat{\varepsilon}_i = Y_{i\cdot} - \hat{\mu}$ so that $\sum_{i=1}^{n}\hat{\varepsilon}_i = 0$. Thus (4.2.17) is equivalent to the statement

$$\frac{1}{n}\sum_{i=1}^{n}\big(\mathrm{sign}(\hat{\varepsilon}_i)-\mathrm{sign}(\varepsilon_i)-2f(0)(\hat{\varepsilon}_i-\varepsilon_i)\big) = o_p(n^{-1/2}).$$

Since f is continuous, it follows $\text{sign}(t) = 1 - 2\mathbf{1}[t \leq 0]$ almost surely. Hence the above statement is implied by

$$\frac{1}{n}\sum_{i=1}^{n}\big(\mathbf{1}[\hat{\varepsilon}_i \leq 0] - \mathbf{1}[\varepsilon_i \leq 0] + f(0)(\hat{\varepsilon}_i - \varepsilon_i)\big) = o_p(n^{-1/2}).$$

Under the null hypothesis, $\hat{\varepsilon}_i - \varepsilon_i = -(\hat{\mu} - \mu_0) = O_p(n^{-1/2})$. Consequently the preceding equality follows from the statement

$$P\Big(\sup_{|t|\leq M}\big|\frac{1}{n}\sum_{i=1}^{n}\eta_{ni}(t)\big| > n^{-1/2}\tau\Big) \to 0, \tag{4.2.20}$$

for arbitaray fixed $M > 0$ and small $\tau > 0$, where

$$\eta_{ni}(t) = \mathbf{1}[\varepsilon_i \leq n^{-1/2}t] - \mathbf{1}[\varepsilon_i \leq 0] - f(0)n^{-1/2}t.$$

As customary, we partition $[-M, M]$ with points $t_k = -M + 2Mk/K$ for $k = 0, 1, \ldots, K$ into K subitervals $I_1 = [t_0, t_1], I_k = (t_{k-1}, t_k], k = 2, \ldots, K$ of equal length. We bound the right side of (4.2.20) by $A_n + B_n$ where

$$A_n = \sum_{k=1}^{K} P\Big(\big|\frac{1}{n}\sum_{i=1}^{n}\eta_{ni}(t_k)\big| > n^{-1/2}\tau/2\Big),$$

$$B_n = P\Big(\max_k \sup_{t\in I_k}\big|\frac{1}{n}\sum_{i=1}^{n}\eta_{ni}(t) - \eta_{ni}(t_k)\big| > n^{-1/2}\tau/2\Big).$$

Now further bound A_n by $A_{n1} + A_{n2}$ where

$$A_{n1} = \sum_{k=1}^{K} P\Big(\big|\frac{1}{n}\sum_{i=1}^{n}\eta_{ni}(t_k) - E(\eta_{ni}(t_k))\big| > n^{-1/2}\tau/4\Big),$$

and

$$A_{n2} = \sum_{k=1}^{K}\mathbf{1}[|E(\eta_{ni}(t_k))| > n^{-1/2}\tau/4].$$

One easily derives

$$E(\eta_{ni}(t_k)) = F(n^{-1/2}t_k) - F(0) - f(0)n^{-1/2}t_k = o(n^{-1/2}), \tag{4.2.21}$$

and

$$\begin{aligned} &\text{Var}(\eta_{ni}(t_k)) \leq E(\eta_{ni}(t_k)^2) \\ &= F(n^{-1/2}t_k) + F(0) - 2F(\min(0, n^{-1/2}t_k)) + O(n^{-1/2}) \qquad (4.2.22) \\ &= f(0)n^{-1/2}t_k - 2f(0)\min(0, n^{-1/2}t_k) + o(n^{-1/2}) = O(n^{-1/2}). \end{aligned}$$

Using (4.2.21), we get

$$A_{n2} \le K\mathbf{1}[o(1) > \tau/4] \to 0.$$

Using (4.2.21)–(4.2.22) and by Berstein's inequality (e.g. Proposition 2.3 of Arcones and Giné (1993)), we derive

$$A_{1n} \le 2K \exp\left(-cn^{1/2}\tau\right) \to 0$$

for some constant $c > 0$. Here $K = K_n = \log n$. Combining the preceding two displays yields $A_n \to 0$. Thus to complete the proof of (4.2.20) we only need to show $B_n \to 0$. To this end, note first that for $t \in I_k$,

$$|\eta_{ni}(t) - \eta_{ni}(t_k)| \le \mathbf{1}[\varepsilon_i \in n^{-1/2}I_k] + f(0)n^{-1/2}2M/K.$$

Hence we bound B_n by $B_{n1} + B_{n2}$ where

$$B_{n1} = P\Big(\max_k \big|\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}[\varepsilon_i \in n^{-1/2}I_k]\big| > n^{-1/2}\tau/4\Big),$$

$$B_{n2} = \mathbf{1}[f(0)n^{-1/2}2M/K > n^{-1/2}\tau/4] = \mathbf{1}[f(0)2M/K > \tau/4].$$

Obviously $B_{n2} \to 0$ as $K \to \infty$. Thus it remains to show $B_{n1} \to 0$. As usual we center the summands by their exepcted values and bound B_{n1} by $C_n + D_n$ where

$$C_n \le P\Big(\max_k \big|\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}[\varepsilon_i \in n^{-1/2}I_k] - P(\varepsilon_i \in n^{-1/2}I_k)\big| > n^{-1/2}\tau/8\Big)$$

and

$$D_n = \sum_{k=1}^{K}\mathbf{1}[P(\varepsilon_1 \in n^{-1/2}I_k) > n^{-1/2}\tau/8].$$

Since

$$P(\varepsilon_1 \in n^{-1/2}I_k) = F(n^{-1/2}t_k) - F(n^{-1/2}t_{k-1}) = f(0)n^{-1/2}2M/K + o(n^{-1/2}),$$

it follows as $K \to \infty$,

$$D_n = K\mathbf{1}[f(0)2M/K > \tau/8] \to 0.$$

Again by Bernstein's inequality, we obtain

$$C_n \leq 2K \exp\left(-dn^{1/2}\tau\right)$$

for some constant $d > 0$. By choosing $K = K_n$ we conclude $C_n \to 0$ for arbitrary $M > 0$ and small $\tau > 0$. This finishes the proof. ∎

Proof of (2.2.7). We shall prove this by applying Theorem 3.3.2 with

$$\mathbf{g}_n(y) = \boldsymbol{\psi}_n(2F(y-\mu_0)-1) \quad \text{and} \quad \hat{\mathbf{g}}_n(y) = \boldsymbol{\psi}_n(2\mathbb{F}_{\mu_0}(y-\mu_0)-1)$$

with $\mathbb{W}_n = I_{r_n}$ and $\mathbf{C}_n = E(2\tilde{h}_1(\mathbf{Y}_1)\mathbf{g}_n(J^{-1}\mathbf{1}^\top\mathbf{Y}_1))$, where $\boldsymbol{\psi}_n = (\psi_1, \ldots, \psi_{r_n})^\top$. By assumption, $r_n h_1$ is Lindeberg. Using $\|\mathbf{g}_n\| \leq \sqrt{r_n}$, we derive (3.3.6) in view of $r_n^3 = o(n)$. We now show $\mathscr{W}_n = \mathscr{W}(4\operatorname{Var}(h_1(\mathbf{Y}_1)), \mathbf{C}_n, I_{r_n})$ satisfies (C). Let D be the common distribution function of ϵ_{1j} and G be the joint distribution $Y_{1j} - \mu, j = 1, \ldots, J$. Then by the independence between u_1 and ε_{1j} we derive that the distribution of ε_1 satisfies $F(t) = P(\varepsilon_1 \leq t) = P(u_1 + \epsilon_{1\cdot} \leq t) = E\big(D(t-u_1)\big), t \in \mathcal{R}$ and

$$G(\mathbf{t}) = P(u_1 + \epsilon_{1j} \leq t_j,\, j = 1, \ldots, J) = E\Big(\prod_{j=1}^{J} D(t_j - u_1)\Big), \quad \mathbf{t} \in \mathcal{R}^J.$$

Since F is continuous and $J \geq 2$, it follows $F \neq G$. Thus the Hilbert space $\mathcal{H}_1 = L_{2,0}(F)$ is a true subspace of $\mathcal{H} = L_{2,0}(G)$. Furthermore, by (2.2.8) $\tilde{h}_1 \in \mathcal{H}$ but $\tilde{h}_1 \notin \mathcal{H}_1$. Thus from Remark 4.1.1 it follows that the matrices $\mathscr{W}_n$ satisfies (C). Let us now prove the first equality of (3.3.8). To this end, we point out the bounds

$$\|\boldsymbol{\psi}_n\|^2 \leq r_n, \quad \|\boldsymbol{\psi}_n'\| \leq \sqrt{2}\pi r_n^{3/2}, \quad |2(\mathbf{t}^\top\boldsymbol{\psi}_n')(\mathbf{t}^\top\boldsymbol{\psi}_n)| \leq 4\pi r_n^2\|\mathbf{t}\|^2. \tag{4.2.23}$$

We break

$$
\begin{aligned}
&\|\frac{1}{n}\sum_{i=1}^{n}\tilde{V}_{ni}\boldsymbol{\psi}_n(2\mathbb{F}_{\mu_0}(\varepsilon_i)) - \mathbf{C}_n\|^2 \\
&\le \|\frac{1}{n}\sum_{i=1}^{n}\big(\tilde{V}_{ni} - 2\tilde{h}_1(\mathbf{Y}_i)\big)^2\frac{1}{n}\sum_{i=1}^{n}\|\boldsymbol{\psi}_n(2\mathbb{F}_{\mu_0}(\varepsilon_i))\|^2\| \\
&\quad + \frac{1}{n}\sum_{i=1}^{n}4\tilde{h}_1(\mathbf{Y}_i)^2\frac{1}{n}\sum_{i=1}^{n}\big(\boldsymbol{\psi}_n(2\mathbb{F}_{\mu_0}(\varepsilon_i)) - \boldsymbol{\psi}_n(2F(\varepsilon_i))\big)^2 \\
&\quad + \|\frac{1}{n}\sum_{i=1}^{n}\big(2\tilde{h}_1(\mathbf{Y}_i)\boldsymbol{\psi}_n(2F(\varepsilon_i)) - E\big[2\tilde{h}_1(\mathbf{Y}_1)\boldsymbol{\psi}_n(2F(\varepsilon_1))\big]\big)\|^2 \\
&\quad := A_n + B_n + C_n.
\end{aligned}
$$

By (3.1.6) and in view of (4.2.23) and $r_n^2 = o(n)$, we derive $A_n \le O_p(n^{-1})r_n = o_p(r_n^{-1})$, while by the square-integrability of h_1 and in view $r_n^4 = o(n)$, we obtain

$$
B_n \le O_p(1)2\pi^2 r_n^3 4 \sup_{-\infty<t<\infty}|\mathbb{F}_{\mu_0}(t) - F(t)|^2 = O_p(r_n^3/n) = o_p(r_n^{-1}).
$$

Now it is not difficult to calculate

$$
\begin{aligned}
E(C_n) &= n^{-1}E\big[\operatorname{Var}\big(2\tilde{h}_1(\mathbf{Y}_1)\|\boldsymbol{\psi}_n(2F(\varepsilon_1))\|\big)\big] \\
&\le n^{-1}E\big(4\tilde{h}_1(\mathbf{Y}_1)^2\|\boldsymbol{\psi}_n(2F(\varepsilon_1))\|^2\big) \\
&\le n^{-1}E\big(4\tilde{h}_1(\mathbf{Y}_1)^2\big)r_n = O(r_n n^{-1}) = o(r_n^{-1}).
\end{aligned}
$$

This concludes the proof of the first equality of (3.3.8). We prove below that (3.3.8) hold with $\mathbf{u}_n = \mathbf{g}_n$, that is,

$$
n^{-1/2}\sum_{i=1}^{n}\big(\boldsymbol{\psi}_n(2\mathbb{F}_{\mu_0}(\varepsilon_i) - 1) - \boldsymbol{\psi}_n(2F(\varepsilon_i) - 1)\big) = o_p(1). \tag{4.2.24}
$$

Clearly $\int \mathbf{u}_n\,dQ = 0$ and $\|\mathbf{u}_n\|$ is Lindeberg in view of $\|\mathbf{u}_n\| \le \sqrt{r_n}$. Thus $\mathbf{v}_n = (2h_1, \mathbf{u}_n^\top)^\top$, $\int \mathbf{v}_n^{\otimes 2}\,dQ = \mathscr{W}_n$, and $\mathbb{U}_n = I_{r_n+1}$, which implies $|\mathbb{U}_n|_o = 1$ and $r_n/\operatorname{trace}(\mathbb{U}_n^2) = r_n/(r_n+1) = O(1)$. We are now left to prove (4.2.24) and the second equality in (3.3.7) which is implied by

$$
\sup_{\|\mathbf{t}\|=1}\Big|\frac{1}{n}\sum_{i=1}^{n}(\mathbf{t}^\top\boldsymbol{\psi}_n(2\mathbb{F}_{\mu_0}(\varepsilon_i) - 1))^2 - 1\Big| = o_p(r_n^{-1/2}). \tag{4.2.25}
$$

Since $\boldsymbol{\psi}_n(2\mathbb{F}_{\mu_0}-1)$ and $\boldsymbol{\psi}_n(2F_{\mu_0}-1)$ are odd functions, the above two equations can be written as

$$\sup_{\|\mathbf{t}\|=1}\Big|\frac{1}{n}\sum_{i=1}^{n}(\mathbf{t}^\top\boldsymbol{\psi}_n(2\mathbb{F}_{\mu_0}(|\varepsilon_i|)-1))^2-1\Big| = o_p(r_n^{-1/2}), \tag{4.2.26}$$

$$n^{-1/2}\sum_{i=1}^{n}\operatorname{sign}(\varepsilon_i)\big(\boldsymbol{\psi}_n(2\mathbb{F}_{\mu_0}(|\varepsilon_i|)-1)-\boldsymbol{\psi}_n(2F(|\varepsilon_i|)-1)\big) = o_p(1). \tag{4.2.27}$$

Also, we have almost surely the identity

$$\begin{aligned}2\mathbb{F}_{\mu_0}(|\varepsilon_k|)-1 &= \frac{1}{n}\sum_{i=1}^{n}\big(\mathbf{1}[\varepsilon_i\le|\varepsilon_k|]+\mathbf{1}[-\varepsilon_i<|\varepsilon_k|]-1\big)\\ &= \frac{1}{n}\sum_{i=1}^{n}\big(\mathbf{1}[|\varepsilon_i|\le|\varepsilon_k|]-\frac{1}{n}\sum_{i=1}^{n}\mathbf{1}[\varepsilon_i=|\varepsilon_k|]\big)\\ &= (R_k-\mathbf{1}[\varepsilon_k\ge 0])/n,\end{aligned}$$

where $R_1,\dots,R_n$ are the ranks of $|\varepsilon_1|,\dots,|\varepsilon_n|$. Using the bounds (4.2.23) and $r_n^3 = o(n)$, it is sufficient for us to prove (4.2.26) and (4.2.27) with $2\mathbb{F}_{\mu_0}(|\varepsilon_i|)-1$ replaced by R_i/n. Let a be a Lipschitz function on $[0,1]$ with Lipschitz constant L. Then we approximate the sum by an integral as follows:

$$\frac{1}{n}\sum_{i=1}^{n}a(R_i/n) = \frac{1}{n}\sum_{i=1}^{n}a(i/n) = \int_0^1 a(x)\,dx + \sum_{i=1}^{n}\int_{(i-1)/n}^{i/n}(a(i/n)-a(x))\,dx$$

and therefore

$$\Big|\frac{1}{n}\sum_{i=1}^{n}a(R_i/n)-\int_0^1 a(x)\,dx\Big| \le L/n.$$

For $a = (\mathbf{u}^\top\boldsymbol{\psi})^2$ and noting $\int(\mathbf{u}^\top\boldsymbol{\psi}(x))^2\,dx = 1$ as $\psi_1,\psi_2,\dots$ are also orthonormal with respect to the uniform measure on $[0,1]$, we get

$$\Big|\frac{1}{n}\sum_{i=1}^{n}(\mathbf{u}^\top\boldsymbol{\psi})^2(R_i/n)-1\Big| \le 4\pi r_n^2/n = o_p(r_n^{-1}).$$

This shows (4.2.26). Let

$$T_n = n^{-1/2}\sum_{i=1}^{n}\operatorname{sign}(\varepsilon_i)\big(\boldsymbol{\psi}_n(R_i/n)-\boldsymbol{\psi}_n(2F(|\varepsilon_i|)-1)\big).$$

Since ε is symmetric, it follows that $\text{sign}(\varepsilon)$ and $|\varepsilon|$ are independent, $\text{sign}(\varepsilon)$ is uniformly distributed on $\{-1, 1\}$ and $|\varepsilon|$ has distribution given by $G(t) = 2F(t) - 1, t \in \mathcal{R}^+$. From this we immediately derive

$$\begin{aligned} E\big(|T_n|^2||\varepsilon_1|, \ldots, |\varepsilon_n|\big) &= \frac{1}{n}\sum_{i=1}^{n} |\boldsymbol{\psi}(R_i/n) - \boldsymbol{\psi}(G(|\varepsilon_i|)|^2 \\ &= 2\pi^2 r_n^3 \frac{1}{n}\sum_{i=1}^{n} |R_i/n - (G(|\varepsilon_i|)|^2 = O_p(r_n^3/n). \end{aligned}$$

This shows $T_m = o_p(1)$ and hence the desired (4.2.27). ■

Proof of (2.3.4). We shall apply Theorem 3.3.3 to prove the result. Note first that the kernel is $h(\mathbf{X}_1, \ldots, \mathbf{X}_m) = \mathbf{1}[x_0 \in \Delta(\mathbf{X}_1, \ldots, \mathbf{X}_m)]$ so that $h_1(\mathbf{x}) = P(x_0 \in \Delta(\mathbf{x}, \mathbf{X}_2, \ldots, \mathbf{X}_m)), \mathbf{x} \in \mathbf{R}^m$ which is bounded by 1. Also $\mathbf{g}_n = \boldsymbol{\phi}_n \circ F_{10}$ hence $\|\mathbf{g}_n\| \le \sqrt{2r_n}$ by (4.2.1). Since $r_n^3 = o(n)$, it follows that $r_n h_1$ and $r_n\|\mathbf{g}_n\|$ are Lindeberg. We are now left to show the regularity. Recall that F is the distribution functions of $\mathbf{X} = (X_1, \ldots, X_m)^\top$ and F_{10} is the distribution function of X_1. Since there is at least one component in $X_2, \ldots, X_m$ is non-degenerate and $m \ge 2$, it follows $\mathcal{H}_1 = L_{2,0}(F_{10})$ is a true subspace of $\mathcal{H} = L_{2,0}(F)$. Clearly $\tilde{h}_1$ lives in $\mathcal{H}$ but not in $\mathcal{H}_1$. It follows from Remark 4.1.1 that (C) is satisfied. This completes the proof. ■

5. SIMULATION RESULTS

In this chapter, we report some simulation results about the Theil test.

5.1 Simulation on the Theil test

We have performed the following simulations to investigate the behavior of the jackknife empirical likelihood for U-statistics with side information when the number of constraints are finite. The distributions of X and error ϵ have been chosen from both symmetrical and asymmetrical distributions. Symmetrical distributions include normal and t-distribution. Asymmetrical distributions include log-normal distribution and chi-squared distribution. Let us first recall some facts. Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent and identically distributed copied of a random vector (X, Y) satisfying a simple linear regression model $Y = \alpha + \beta X + \epsilon$, where X is a covariate which has a continuous distribution function G, and ϵ is a random error which has a continuous distribution F, and X and ϵ are independent. The null hypothesis of interest is that the slope β is equal to some specified value β_0, namely, $H_0 : \beta = \beta_0$. Theil's test statistics can be expressed as the U-statistics given by

$$U_n(h_{ts}) = \binom{n}{2}^{-1} \sum_{i<j} \mathbf{1}[(Y_i - Y_j)/(X_i - X_j) - \beta_0 > 0]$$

Recall that under the null hypothesis ϵ and X are independent, so that $E(a(X)\epsilon) = 0$ for every $a \in L_{2,0}(G)$. Since G is continuous but unknown, we estimate it by $\mathbb{G}$. We look at the empirical likelihood

$$\mathscr{R}_n = \sup\Big\{ \prod_{j=1}^{n} n\pi_j : \boldsymbol{\pi} \in \mathscr{P}_n,\ \sum_{j=1}^{n} \pi_j V_{nj} - 1/2 = 0,$$
$$\sum_{j=1}^{n} \pi_j \phi_k(\mathbb{G}(X_j))(Y_j - \beta_0 X_j) = 0,\ k = 1, \ldots, r\Big\},$$

where V_{nj}'s are the jackknife pseudo values of the U-statistics $U_n(h_{ts})$.

As can be seen from the summary tables, when the correct side information is used, the gain in power is obtained, especially at the values near the null hypothesis, where the power of JEL test only is small. On the other hand, the nominal type 1 error rates are still maintained for using the correct side information. In all the simulation scenarios, we first add the zero median constraint, followed by adding additional the constraints based on independence of X and error. One observes that as the sample increased from $n = 50$ to $n = 80$, the power increases, while the nominal levels are maintained.

Table 5.1
Simulated power for Theil test, $X \sim \mathcal{N}(10, 1)$, $\epsilon \sim \mathcal{N}(0, 1)$, $\beta_0 = 5$, $n = 50$, $M = 2000$, Nc=# of constraints.

	β						
Nc	4.5	4.8	4.9	5.0	5.1	5.2	5.5
1	0.852	0.223	0.092	0.046	0.091	0.226	0.858
2	1	1	1	0.042	0.999	1	1
3	1	1	0.997	0.040	0.999	1	1
4	1	1	0.997	0.051	0.996	1	1
5	1	1	0.997	0.063	0.997	1	1

Table 5.2
Simulated power for Theil test, $X \sim \mathcal{N}(10,1)$, $\epsilon \sim \chi^2(4)$ with zero median and $sd = 1$, $\beta_0 = 5$, $n = 50$, $M = 2000$, Nc=# of constraints.

	β						
Nc	4.5	4.8	4.9	5.0	5.1	5.2	5.5
1	0.938	0.338	0.114	0.043	0.120	0.355	0.937
2	1	1	1	0.044	0.996	1	1
3	1	1	1	0.049	0.991	1	1
4	1	1	1	0.052	0.980	1	1
5	1	1	1	0.069	0.988	1	1

Table 5.3
Simulated power for Theil test, $X \sim \mathcal{N}(10,1)$, $\epsilon \sim$ Cauchy, $\beta_0 = 5$, $n = 50$, $M = 2000$, Nc=# of constraints.

	β						
Nc	4.5	4.8	4.9	5.0	5.1	5.2	5.5
1	0.35	0.119	0.048	0.046	0.060	0.089	0.361
2	1	1	0.926	0.045	0.931	1	1
3	1	0.999	0.888	0.042	0.900	1	1
4	1	0.999	0.882	0.060	0.876	0.999	1
5	1	0.997	0.855	0.062	0.861	0.999	1

Table 5.4
Simulated power for Theil test, $X \sim$ Log-normal(mean=10, sd=1), $\epsilon \sim \mathcal{N}(0,1)$, $\beta_0 = 5$, $n = 50$, $M = 2000$, Nc=# of constraints.

	β						
Nc	4.5	4.8	4.9	5.0	5.1	5.2	5.5
1	0.847	0.255	0.085	0.034	0.089	0.241	0.858
2	1	1	1	0.049	1	1	1
3	1	1	0.998	0.038	0.998	1	1
4	1	1	0.999	0.062	0.997	1	1
5	1	1	0.997	0.065	0.996	1	1

Table 5.5
Simulated power for Theil test, $X \sim$ Log-normal(mean=10, sd=1), $\epsilon \sim \chi(4)$ with zero median and sd=1, $\beta_0 = 5$, $n = 50$, $M = 2000$, Nc=# of constraints.

	β						
Nc	4.5	4.8	4.9	5.0	5.1	5.2	5.5
1	0.934	0.324	0.127	0.044	0.121	0.317	0.935
2	1	1	1	0.044	0.994	1	1
3	1	1	1	0.045	0.988	1	1
4	1	1	1	0.060	0.985	1	1
5	1	1	1	0.068	0.977	1	1

Table 5.6
Simulated Power for Theil test, $X \sim$ log-normal(mean=10, sd=1), $\epsilon \sim$ Cauchy, $\beta_0 = 5$, $n = 50$, $M = 2000$, Nc=# of constraints.

	β						
Nc	4.5	4.8	4.9	5.0	5.1	5.2	5.5
1	0.356	0.102	0.060	0.040	0.062	0.089	0.345
2	1	1	0.921	0.056	0.929	1	1
3	1	1	0.904	0.044	0.890	1	1
4	1	0.998	0.878	0.060	0.865	0.999	1
5	1	0.999	0.851	0.062	0.857	0.996	1

Table 5.7
Simulated power for Theil test, $X \sim \mathcal{N}(10, 1)$, $\epsilon \sim \mathcal{N}(0, 1)$, $\beta_0 = 5$, $n = 80$, $M = 2000$, Nc=# of constraints.

	β						
Nc	4.5	4.8	4.9	5.0	5.1	5.2	5.5
1	0.980	0.347	0.107	0.036	0.128	0.381	0.981
2	1	1	1	0.060	1	1	1
3	1	1	1	0.038	1	1	1
4	1	1	1	0.042	1	1	1
5	1	1	1	0.055	1	1	1

Table 5.8
Simulated power for Theil test, $X \sim \mathcal{N}(10,1)$, $\epsilon \sim \chi^2(4)$ with zero median and sd=1, $\beta_0 = 5$, $n = 80$, $M = 2000$, Nc=# of constraints.

	β						
Nc	4.5	4.8	4.9	5.0	5.1	5.2	5.5
1	0.992	0.498	0.181	0.041	0.173	0.530	0.994
2	1	1	1	0.052	1	1	1
3	1	1	1	0.04	1	1	1
4	1	1	1	0.041	1	1	1
5	1	1	1	0.055	1	1	1

Table 5.9
Simulated power for Theil test, $X \sim \mathcal{N}(10,1)$, $\epsilon \sim$ Cauchy, $\beta_0 = 5$, $n = 80$, $M = 2000$, Nc=# of constraints.

	β						
Nc	4.5	4.8	4.9	5.0	5.1	5.2	5.5
1	0.534	0.146	0.059	0.056	0.061	0.147	0.536
2	1	1	0.992	0.050	0.995	1	1
3	1	1	0.986	0.034	0.984	1	1
4	1	1	0.977	0.046	0.982	1	1
5	1	1	0.974	0.06	0.978	1	1

Table 5.10
Simulated power for Theil test, $X \sim$ Log-normal(mean=10, sd=1), $\epsilon \sim \mathcal{N}(0,1)$, $\beta_0 = 5$, $n = 80$, $M = 2000$, Nc=# of constraints.

	β						
Nc	4.5	4.8	4.9	5.0	5.1	5.2	5.5
1	0.974	0.348	0.123	0.049	0.132	0.343	0.979
2	1	1	1	0.049	1	1	1
3	1	1	1	0.039	1	1	1
4	1	1	1	0.042	1	1	1
5	1	1	1	0.046	1	1	1

Table 5.11
Simulated power for Theil test, $X \sim$ Log-normal(mean=10, sd=1), $\epsilon \sim \chi^2(4)$ with zero median and sd=1, $\beta_0 = 5$, $n = 80$, $M = 2000$, Nc=# of constraints.

	β						
Nc	4.5	4.8	4.9	5.0	5.1	5.2	5.5
1	0.991	0.525	0.175	0.049	0.163	0.508	0.992
2	1	1	1	0.048	1	1	1
3	1	1	1	0.040	1	1	1
4	1	1	1	0.048	1	1	1
5	1	1	1	0.048	1	1	1

Table 5.12
Simulated power for Theil test, $X \sim$ Log-normal(mean=10, sd=1), $\epsilon \sim$ Cauchy, $\beta_0 = 5$, $n = 80$, $M = 2000$, Nc=# of constraints.

	β						
Nc	4.5	4.8	4.9	5.0	5.1	5.2	5.5
1	0.554	0.141	0.063	0.041	0.068	0.138	0.555
2	1	1	0.992	0.051	0.991	1	1
3	1	1	0.990	0.048	0.988	1	1
4	1	1	0.981	0.049	0.980	1	1
5	1	1	0.977	0.050	0.971	1	1

5.2 The Theil test: infinitely many constraints

In this section, we report some simulation results about the asymptotic normality of the standardized negative twice logarithm of the empirical likelihood, that is, (2.1.6).

Below we report the q-q plots of the standardized negative twice logarithm of the empirical likelihood versus the standard normal for the sample sizes $100, 150$ and the number of constraint $1, 5, 10, 15$.

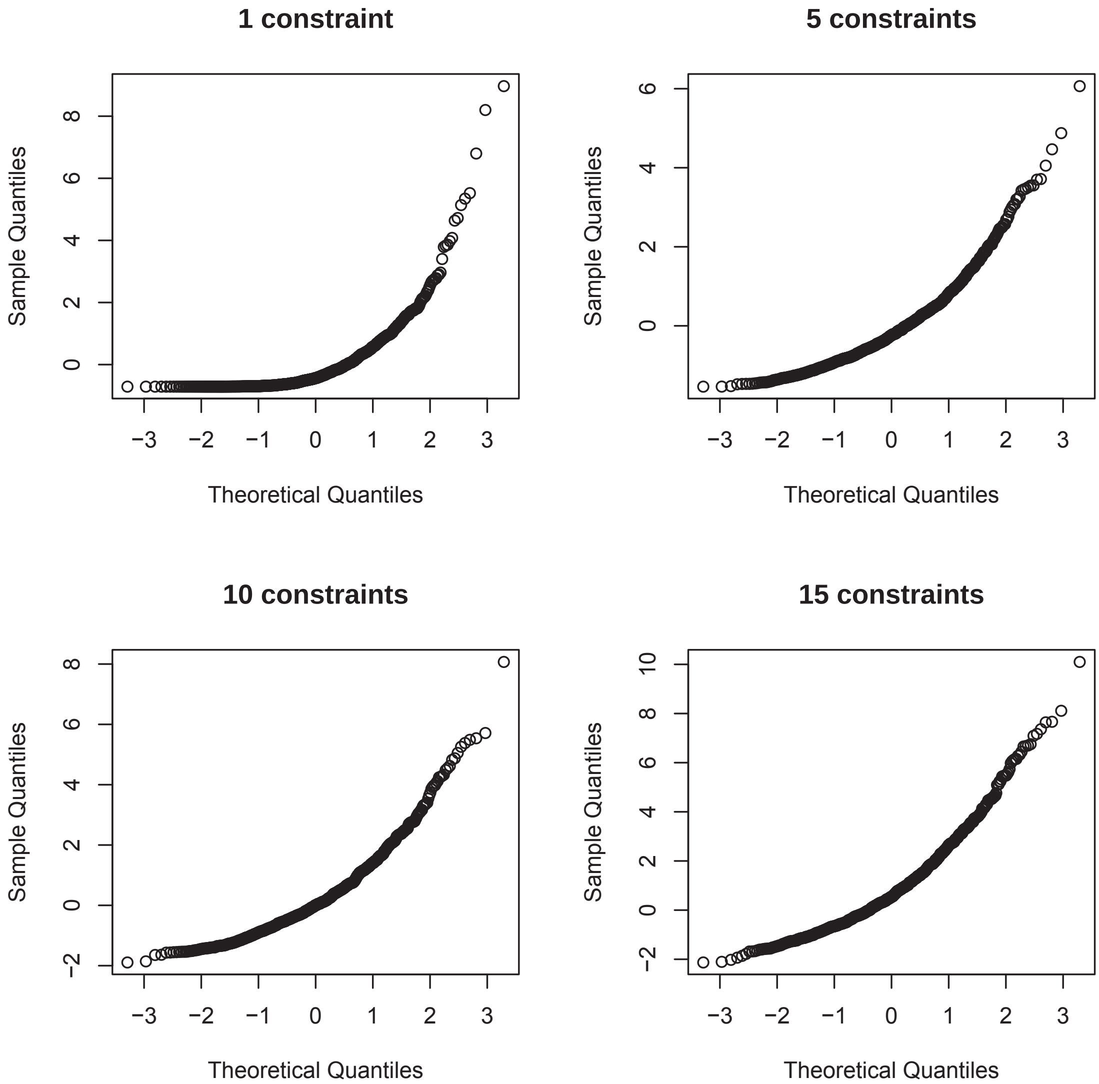

Figure 5.1. The Q-Q plot for normal X, normal ϵ and $n = 100$.

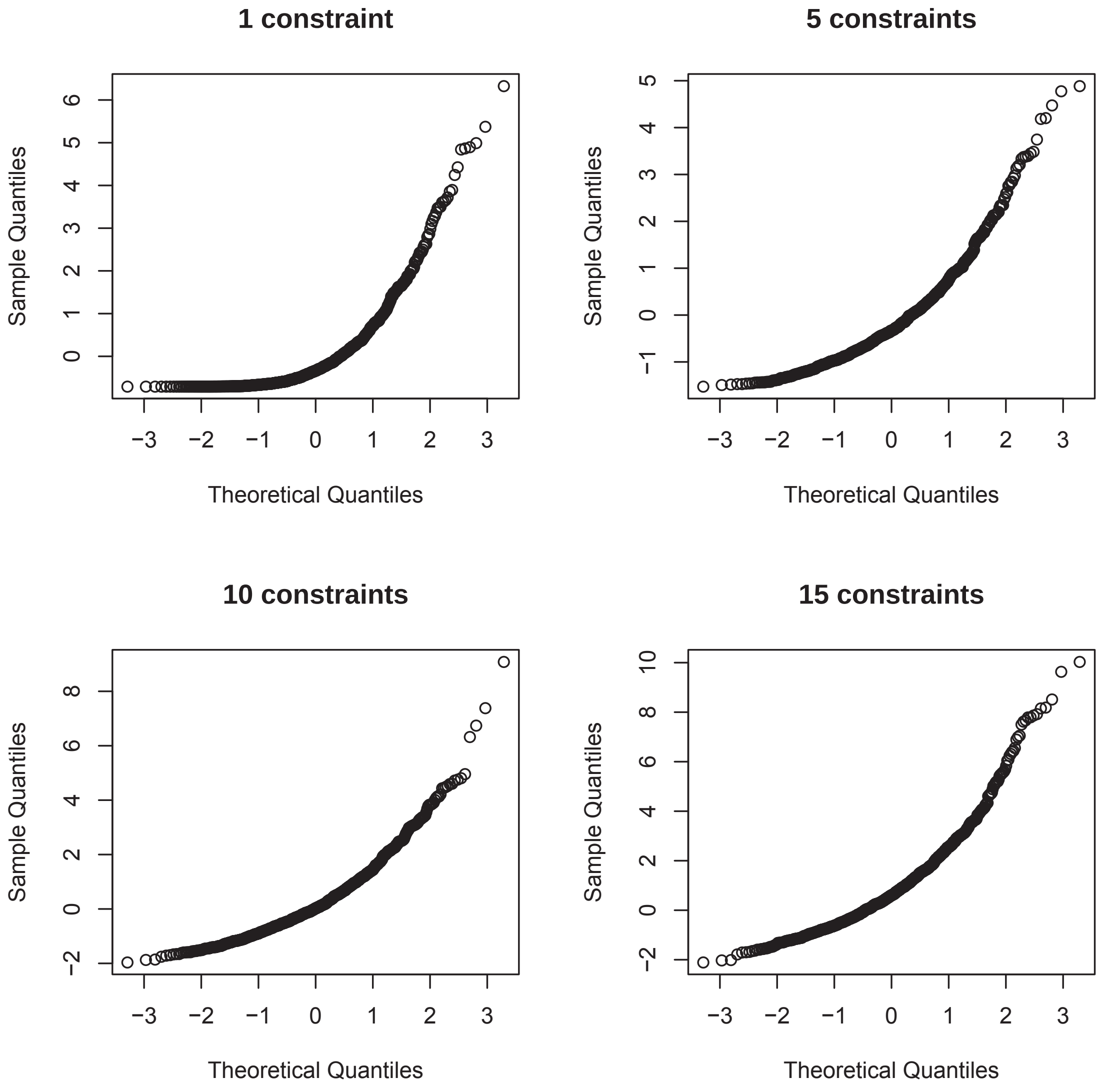

Figure 5.2. The Q-Q plot for normal X, Cauchy ϵ and $n = 100$.

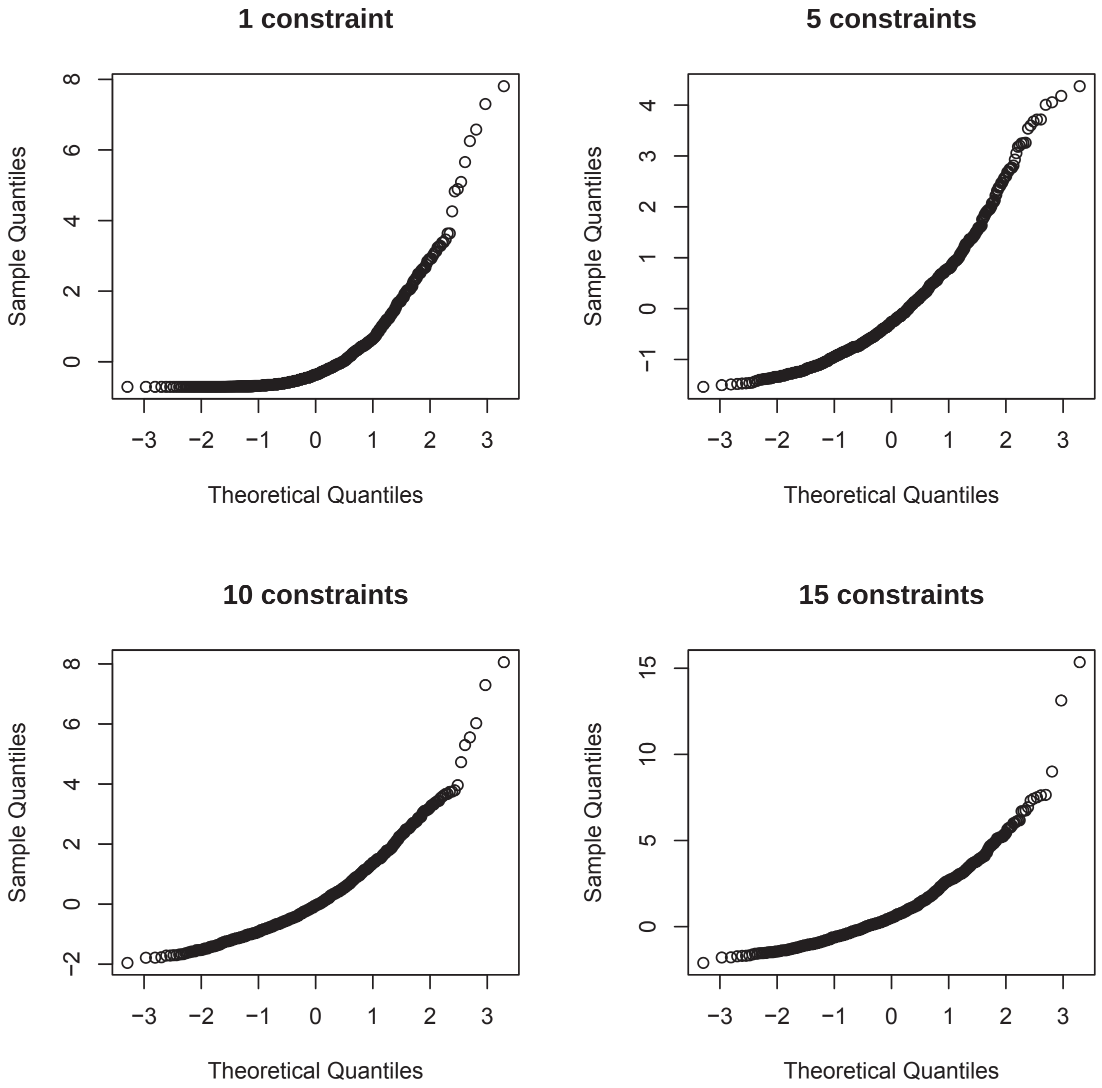

Figure 5.3. The Q-Q plot for Cauchy X, normal ϵ and $n = 100$.

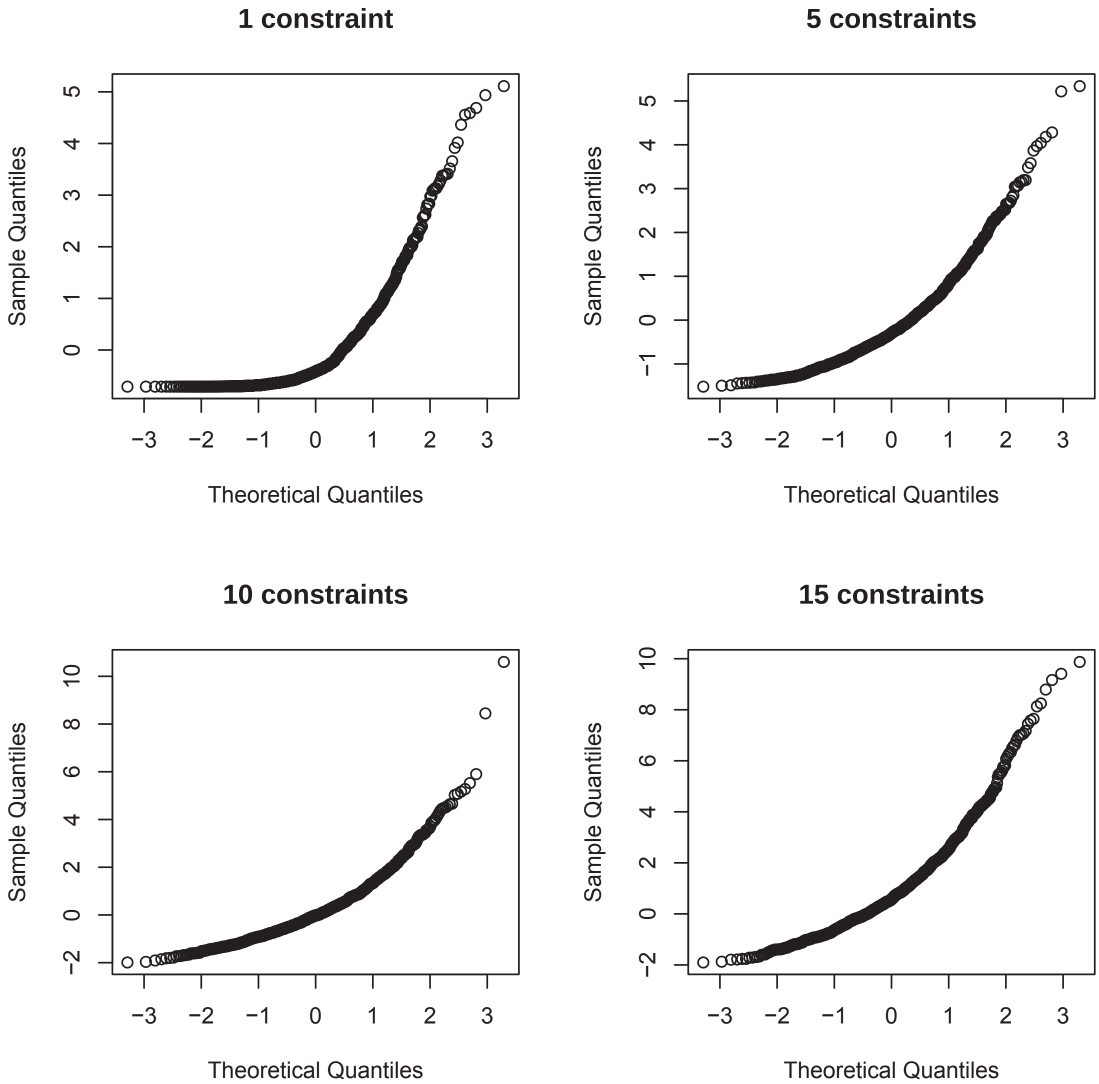

Figure 5.4. The Q-Q plot for Cauchy X, Cauchy ϵ and $n = 100$.

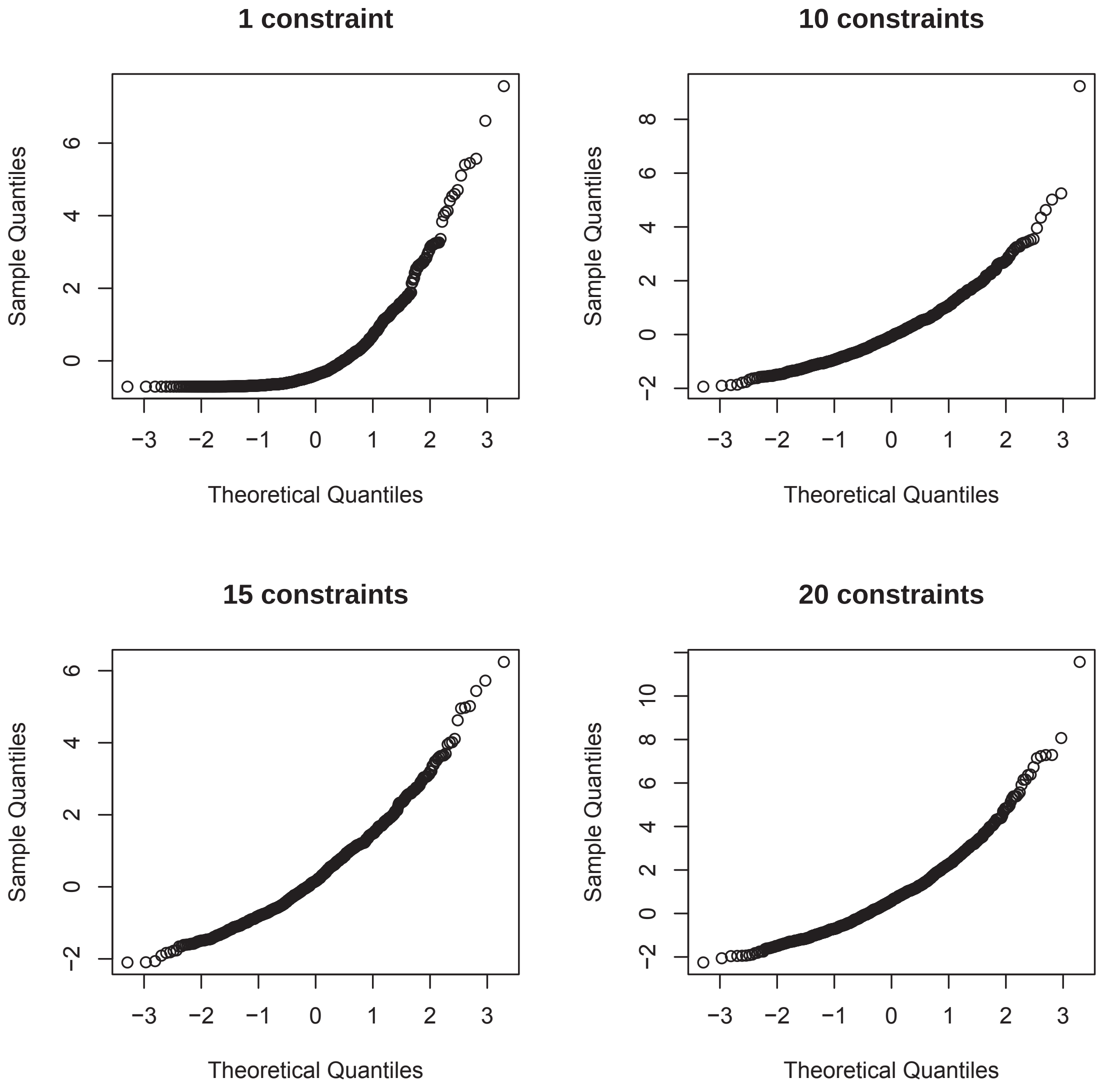

Figure 5.5. The Q-Q plot for normal X, normal ϵ and $n = 150$.

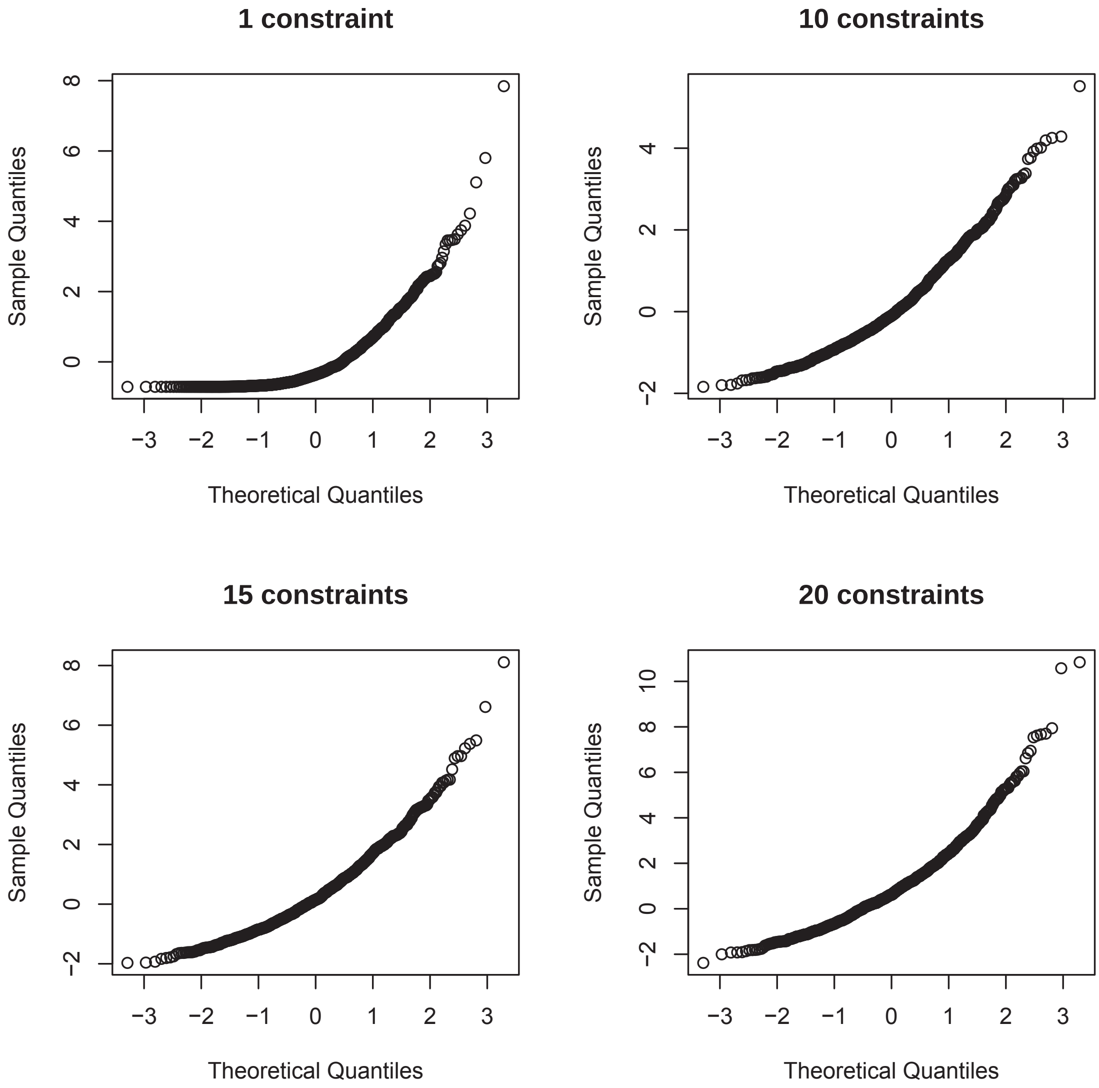

Figure 5.6. The Q-Q plot for normal X, Cauchy ϵ and $n = 150$.

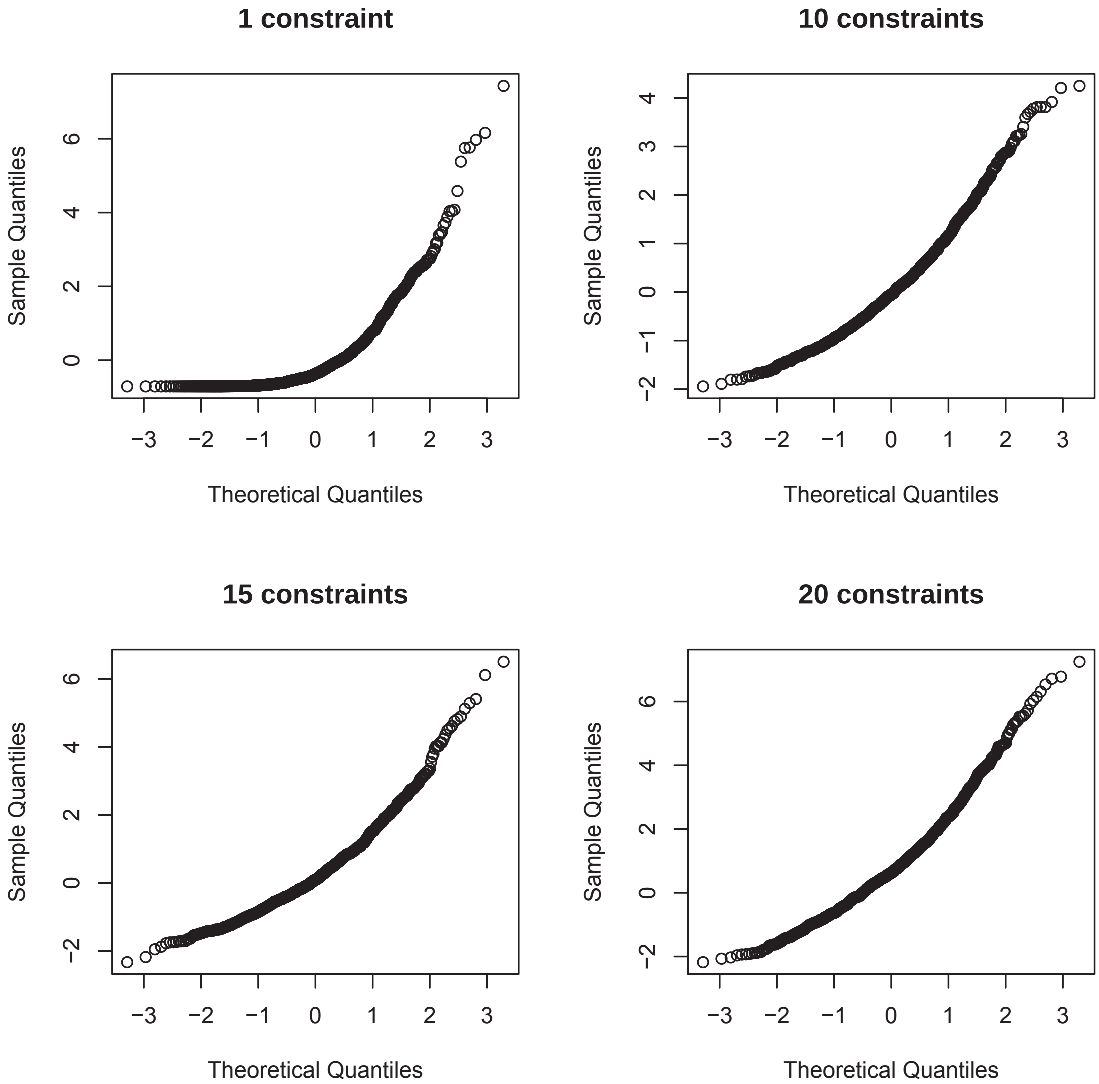

Figure 5.7. The Q-Q plot for Cauchy X, normal ϵ and $n = 150$.

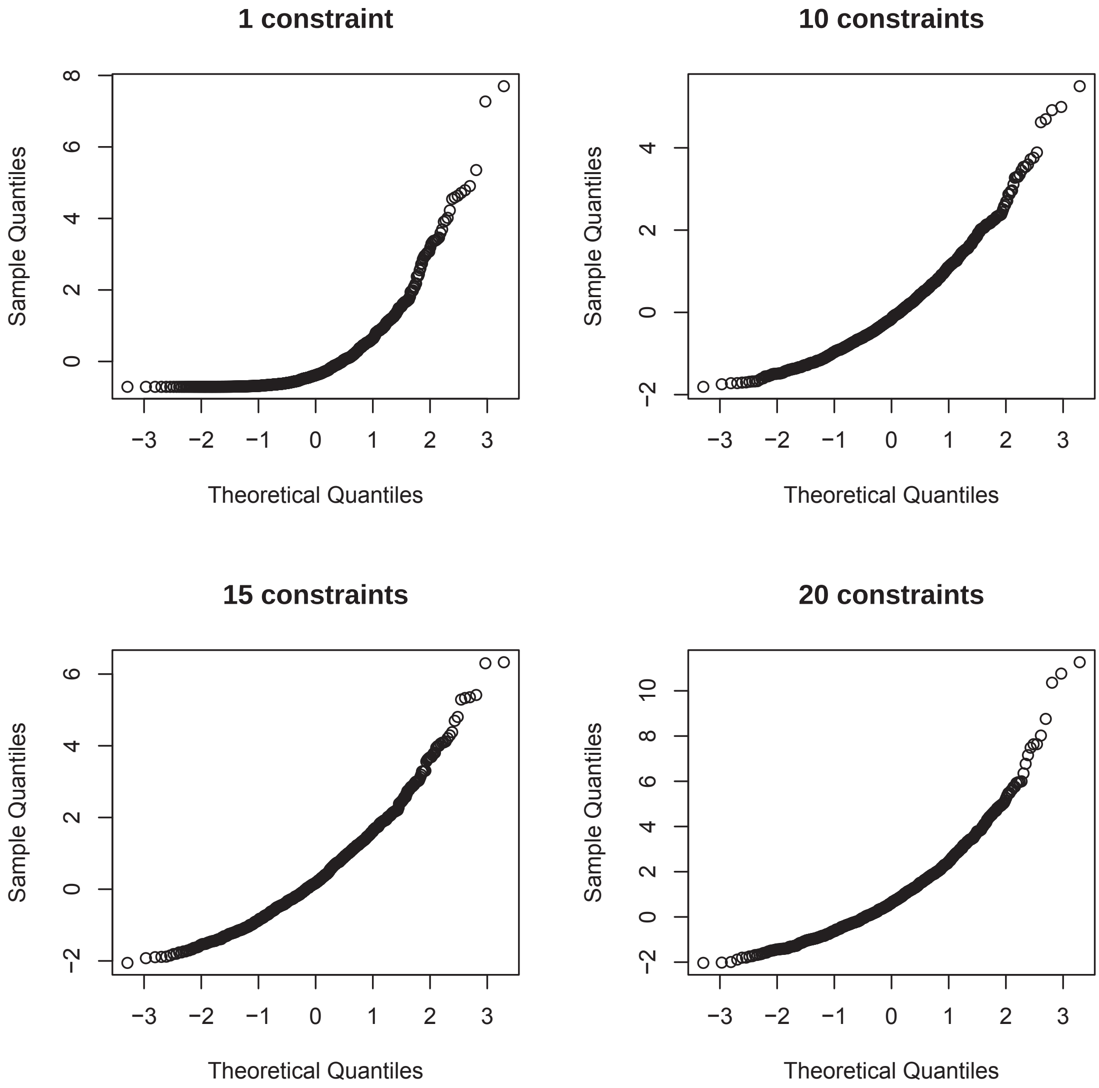

Figure 5.8. The Q-Q plot for Cauchy X, Cauchy ϵ and $n = 150$.

LIST OF REFERENCES

LIST OF REFERENCES

[1] Arcones M.A. (1996). The Bahadur-Kiefer Representation for U-Quantiles. *Ann. Statist.* **24**: 1400 – 1422.

[2] Arcones, M.A. and Giné, E. (1993). Limit Theorems for U-Processes. *Ann. Statist.* **21**: 1494 – 1542.

[3] Arvesen, J.N. (1969). Jackknifing U-statistics. *Ann. Statist.* **40**: 2076 – 2100.

[4] Bassett, G. W. and Koenker, R. (1986). Strong Consistency of Regression Quantiles and Related Empirical Processes. *Econometric Theory* **2**, 191-201.

[5] Bertail, P. (2006). Empirical likelihood in some semiparametric models. *Bernoulli* **12**, 299–331.

[6] Bickel, P.J., Klaassen, C.A.J., Ritov, Y. and Wellner, J.A. (1993). *Efficient and Adaptive Estimation in Semiparametric Models.* Johns Hopkins Univ. Press, Baltimore.

[7] Chen, S.X. and Cui, H.J. (2006). On Bartlett correction of empirical likelihood in the presence of nuisance parameters. *Biometrika* **93**, 215–220.

[8] Chen, S.X., Peng, L. and Qin, Y.-L. (2009). Effects on data dimension on empirical likelihood. *Biometrika* **96**, 711–722.

[9] Chen, S.X. and Qin, Y.S. (2000). Empirical likelihood confidence intervals for local linear smoothers. *Biometrika* **87**, 946–953.

[10] Chen, S.X. and Van Keilegom, I. (2009). A goodness-of-fit test for parametric and semiparametric models in multiresponse regression. *Bernoulli*, **15**, 955–976.

[11] Chen, S. X. and Van Keilegom, I. (2009). A review on empirical likelihood methods for regression. *Test* **18**, 415–447.

[12] Conway, J.B. (1985). *A Course in Functional Analysis.* Springer-Verlag, New York Inc.

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

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

[15] Fang, F., Hong, Q. and Shao, J. (2009). A pseudo empirical likelihood approach for stratified samples with nonresponse. *Ann. Statist.* **37**, 371–393.

[16] Fleming, T.R. and Harrington, D.P. (2005). *Counting processes and survival analysis*. John Wiley & Sons, Inc.

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

[18] Heffernan, P. M. (1997). Unbiased Estimation of Central Moments by using U-statistics. *J. R. Statist. Soc. B* **59** 861 – 863.

[19] Helland, I. (1982). Central limit theorems for martingales with discrete or continuous time. *Scand. J. Statist.* **9**, 79–94.

[20] Hjort, N. L. and Pollard, D. (1993). Asymptotics for minimisers of convex processes. *Statistical Research Report.*

[21] Hjort, N.L., McKeague, I.W. and Van Keilegom, I. (2009). Extending the scope of empirical likelihood. *Ann. Statist.* **37**, 1079–1111.

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

[23] Jurecskva, J. Picek, J. (2006). *ROBUST STATISTICAL METHODS with R*. Chapman & Hall/CRC.

[24] Kolaczyk, E.D. (1994). Empirical likelihood for generalized linear models. *Statist. Sinica* **4**, 199–218.

[25] Koltchinskii, V.I. (1997). M-estimation, convexity and quantiles. *Ann. Statist.* **25**, 435–477.

[26] Kowalski, J. and Tu, X.M. (2008). *Modern Applied U-Statistics*. John Wiley & Sons, Inc.

[27] Lee, A. J. (1990). *U-Statistics: Theory and Practice*. Marcel Dekker, Inc.

[28] Li, G. and Wang, Q.-H. (2003). Empirical likelihood regression analysis for right censored data. *Statist. Sinica* **13**, 51–68.

[29] Liang, K-Y., Zeger, S.L. and Qaqish, B. (1992). Multivariate regression analyses with categorical data. *J. R. Statist. Soc. B* **54**, 3–40.

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

[31] Lopez, E.M.M., Van Keilegom, I. and Veraverbeke, N. (2009). Empirical likelihood for non-smooth criterion functions. *Scand. J. Statist.*, **36**, 413–432.

[32] Lu, X. and Tsiatis, A.A. (2008). Improving the efficiency of the log-rank test using auxiliary covariates. *Biometrika* **95**, 679–694.

[33] Lu, X. (2009). Empirical likelihood for heteroscedastic partially linear models. *J. Multivariate Anal.* **100**, 387–396.

[34] MILLER, R.G., JR. (1968). Jackknifing variances. *Ann. Math. Statist.* **39**, 567–582

[35] Müller, U.R. and Keilegom, V.I. (2012). Efficient quantile regression with auxiliary information. *Electron. J. Statist.* **6**, 1200–1219.

[36] Nobre, J. S., Singer, J. M. and Silvapulle, M. J. (2008). U-tests for variance components in one-way random effects models. *Beyond Parametric in Interdisciplinary Research: Festschrift in Honor of Professor Pranab K. Sen* Vol. **1** 197 – 210.

[37] Nordman, D.J. and Lahiri, S.N. (2006). A frequency domain empirical likelihood for short- and long-range dependence. *Ann. Statist.* **34**, 3019–3050.

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

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

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

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

[42] Pan, X.-R. and Zhou, M. (2002). Empirical likelihood ratio in terms of cumulative hazard function for censored data. *J. Multiv. Analy.* **80**, 166–188.

[43] Parente, P.M.D.C. and Smith, R.J. (2011). GEL methods for nonsmooth moment indicators. *Econometric Theory* **27**, 74–113.

[44] Peng, H. and Schick, A. (2004). Estimation of linear functionals of bivariate distributions with parametric marginals. *Statist. Decisions* **22**, 61–77.

[45] Peng, H. and Schick, A. (2005). Efficient estimation of linear functionals of a bivariate distribution with equal, but unknown marginals: the least-squares approach. *J. Multivariate Anal.* **95**, 385–409.

[46] Peng, H. and Schick, A. (2012). Asymptotic Normality of Quadratic Forms with Random Vectors of Increasing Dimension. Submitted to the Statistica Sinica, accepted with minor revision. The *Manuscript* can be found at the URL: `https://math.iupui.edu/~hpeng/preprints_hp.html.`

[47] Peng, H. and Schick, A. (2013). Maximum empirical likelihood estimation and related topics. Submitted to the Annals of Statistics. The *Manuscript* can be found at the URL: `https://math.iupui.edu/~hpeng/preprints_hp.html.`

[48] Peng, H. (2013). On a class of maximum empirical likelihood estimators. Submitted to the Annals of Statistics. The *Manuscript* can be found at the URL: `https://math.iupui.edu/~hpeng/preprints_hp.html.`

[49] Peng, H. and Schick, A. (2013a). Asymptotic normality of quadratic forms with random vectors of increasing dimension The *Manuscript* can be found at the URL: http://www.math.binghamton.edu/anton/preprint.html.

[50] Peng, H. and Schick, A. (2013b). Maximum empirical likelihood estimation and related topics. The *Manuscript* can be found at the URL: http://www.math.binghamton.edu/anton/preprint.html.

[51] Peng, H. and Schick, A. (2013c). An empirical likelihood approach of goodness of fit testing. *Bernoulli* **19**, 954–981

[52] Peng, H., Wang, S. and Wang, X. (2008) Consistency and Asymptotic Distribution of the Theil-Sen Estimator. *J. Statist. Plann. Infer.* **138**(6), 1836-1850.

[53] Portnoy, S. (1988). Behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.* **16**, 356–366.

[54] Portnoy, S. (1988). Behavior of likelihood methods for exponential families when the number of parameters tends to infinity. *Ann. Statist.* **16**, 356–366.

[55] Qin, G. and Jing, B.-Y. (2001). Empirical likelihood for censored linear regression. *Scand. J. Statist.* **28**, 661–673.

[56] Qin, G. and Jing, B.-Y. (2001b). Censored partial linear models and empirical likelihood. *J. Multivariate Anal.* **78**, 37–61.

[57] Qin, J. and Lawless, J. (1994). Empirical likelihood and general estimating equations. *Ann. Statist.* **22**, 30–325.

[58] Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons, Inc.

[59] Shi, X. (1984). Approximate Indepedence of Jackknife Pseudo Values and the Bootstrap Mehtod. *Journal of Wuhan Institute of Hydra-Electric Engineering* (Chinese) **2**, 83 – 90.

[60] Shi, J. and Lau, T.S. (2000). Empirical likelihood for partially linear models. *J. Multivariate Anal.* **72**, 132–148.

[61] Tang, C. Y. and Leng, C. (2012). An empirical likelihood approach to quantile regression with auxiliary information. *Statist. & Probabil. Lett.* **82**, 29–36.

[62] Tripathi, G. and Kitamura, Y. (2003). Testing conditional moment restrictions *Ann. Statist.* **31**, 2059–2095.

[63] Vaart, van der A.W. (2000). *Asymptotic Statistics.* Cambridge University Press.

[64] Vaart, van der A.W. and Wellner, J.A. (1996). *Weak convergence and empirical processes.* Springer-Verlag New York, Inc.

[65] Wang, Q.-H. and Jing, B.-Y. (2001). Empirical likelihood for a class of functionals of survival distribution with censored data. *Ann. Inst. Statist. Math.* **53**, 517–527.

[66] Wang, Q.-H. and Jing, B.-Y. (2003). Empirical likelihood for partially linear models. *Ann. Inst. Statist. Meth.* **55**, 585–595.

[67] Yuan, A., He, W., Wang, B. and Qin, G. (2012). U-statistics with side information. *J. Multiv. Analy.* **111**, 20 – 38.

[68] Zhang, B. (1995). M-estimation and quantile estimation in the presence of auxiliary information. *J. Statist. Plann. Infer.* **44**, 77 – 94.

[69] Zhang, B. (1997). Quantile processes in the presence of auxiliary information. *Ann. Inst. Statist. Math.* **49**, 35 – 55.

[70] Zhou, M. and Li, G. (2008). Empirical likelihood analysis of the Buckley-James estimator. *J. Multiv. Analy.* **99**, 649–664.

[71] Zhu, L. and Xue, L. (2006). Empirical likelihood confidence regions in a partially linear single-index models. *J. Roy. Statist. Soc. Ser. B* **68**, 549–570.

VITA

VITA

My name is Qun Lin. I came from China to the United States in 1995. I obtained my Bachelor's degree in Scicencs from Jilin University, China in 1991. I studied in the Department of Mathematical Sciences of IUPUI and obtained my Master degree in Applied Statistics in 2000. Since then, I work as a statistical analyst and statistician in the pharmaceutical industry.