SOME QUESTIONS IN RISK MANAGEMENT AND HIGH-DIMENSIONAL DATA ANALYSIS

A Thesis
Presented to
The Academic Faculty

by

Ruodu Wang

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

Georgia Institute of Technology
August 2012

SOME QUESTIONS IN RISK MANAGEMENT AND HIGH-DIMENSIONAL DATA ANALYSIS

Approved by:

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

Professor Christian Houdré
School of Mathematics
Georgia Institute of Technology

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

Professor Shaun Wang
Department of Risk Management and Insurance
Georgia State University

Professor Ming Yuan
School of Industrial and Systems Engineering
Georgia Institute of Technology

Date Approved: April 30, 2012

To my parents.

ACKNOWLEDGEMENTS

I have discovered a truly marvelous list of the people who own my gratitude, which these pages are too narrow to contain.

— Adapted from the words of Pierre de Fermat

First and foremost, I must thank my thesis advisor, Professor Liang Peng, for his patience, kindness and guidance, with which I have joyfully lived the past three years at Georgia Tech. It is a great honor and privilege to be one of his students. Professor Peng has led me into the fabulous world of statistics and treated me like a real collaborator, despite my very limited ability and experience as a researcher. I personally feel his love for all his students including myself. Professor Peng is a role model as a prominent statistician, an erudite professor, a caring advisor and an honorable person that I would like to emulate in the future.

I would like to thank the members of my dissertation committee and oral exam committee: Professors Christian Houdré, Xiaoming Huo, Shaun Wang and Ming Yuan for their time, patience and effort to support my study. I especially want to thank Professor Houdré for his research guidance during my first year at Georgia Tech, for the subsequent insightful conversations we had, for his comments and encouragement regarding my papers and presentations, and for the French sense of humor. I would also like to thank Professor Shaun Wang and Professor Ming Yuan for nicely encouraging me to pursue my dreams in academia and for their support letters which finally made my dreams real.

My journey of pursuing an academic career was originally inspired by Professor Jingping Yang at Peking University, my former advisor and a persistent collaborator. I am deeply grateful to him for all he has done for me, both during the period at

Peking University and during that at Georgia Tech. I would like to thank Professor Yongcheng Qi at University of Minnesota for our years of productive collaborations that began in 2008. I thank Professor Giovanni Puccetti at University of Firenze and Professor Rongmao Zhang at Zhejiang University for pleasant collaborations and discussions. I thank Professor Xiaoying Han at Auburn University for kindly and consistently helping me during my graduate studies. I also want to thank my dear friend since college, Bin Wang at Peking University, for his brilliant ideas and hard work during our collaboration.

Amongst many people in the School of Mathematics who have made my study enjoyable and productive, I want first to thank Professor Vladimir Koltchinskii for his ideas which have significantly contributed to my research, and for the pleasant and inspiring conversations we had. I owe Ms. Klara Grodzinsky my deepest gratitude for her constant and patient help on my teaching and for providing support letters. Professor Luca Dieci has helped me a lot on many school issues as a graduate coordinator who always answers my requests very promptly. I am especially thankful to Ms. Cathy Jacobson for her classes and personal help all these years which greatly improved my English as well as my written papers and essays. I also thank Ms. Genola Turner, Sharon McDowell, Christy Dalton, Jan Lewis, Karen Hinds and Inetta Worthy for their dedicated service to the School.

During a dark and stormy night, I was working hard on writing this thesis when lightning struck my home and shut the electricity down. When the power was back on again, my hard drive was dead with this thesis which I had never backed up. I tried every effort to rescue the data by myself and failed miserably. Therefore, I would like to especially thank Samantha Bueno and Justin Filoseta from the IT group in the School for recovering the data. They literally saved the the life of this thesis. Of course, I would like to thank all the IT group for their fantastic job supporting everyone in the School.

It is my friends at Georgia Tech who have made my life in the States so colorful. Amongst many, Huijun Feng and Yun Gong, who shared the same advisor with me and therefore we are brothers and sisters in the Chinese tradition, have helped me so much with my study and life. I would also like to thank my other dear friends: Albert Bush, Giang Do, Liang Du, Yunlong He, Allen Hoffmeyer, Huy Huynh, Jingfang Liu, Jun Lu, Ning Tan, Ruidong Wang, Qianyi Wang, Xiaolin Wang, Yi Xiao, Linwei Xin, Ran Yin, and any whom I may have failed to mention.

My life and work was financially supported by the Teaching Assistantship in the School of Mathematics at Georgia Tech. I was also partly supported by the grants of Professor Peng, by Peking University during the summer of 2011, and by the Bob Price Fellowship in the School of Mathematics. I thank all the graduate committee members and fund providers for making the financial aid possible, so that I could concentrate on my research.

Finally, I thank my parents for bringing me into this beautiful world, full of surprises and challenges, and for loving me as I love them. My dearest thanks to my beloved girlfriend Jia, without whose sacrifices, support and love, my life would never be so meaningful. Thank the god of mathematics, if one exists, for leading me here, and elsewhere on my prospective journey.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This thesis addresses three topics in the area of statistics and probability, with applications in risk management. First, for the testing problems in the high-dimensional (HD) data analysis, we present a novel method to formulate empirical likelihood tests and jackknife empirical likelihood tests by splitting the sample into subgroups. New tests are constructed to test the equality of two HD means, the coefficient in the HD linear models and the HD covariance matrices. Second, we propose jackknife empirical likelihood methods to formulate interval estimations for important quantities in actuarial science and risk management, such as the risk-distortion measures, Spearman's rho and parametric copulas. Lastly, we introduce the theory of completely mixable (CM) distributions. We give properties of the CM distributions, show that a few classes of distributions are CM and use the new technique to find the bounds for the sum of individual risks with given marginal distributions but unspecific dependence structure. The result partially solves a problem that had been a challenge for decades, and directly leads to the bounds on quantities of interest in risk management, such as the variance, the stop-loss premium, the price of the European options and the Value-at-Risk associated with a joint portfolio.

CHAPTER I

INTRODUCTION

I would use the word *amazing* to describe what I feel about the rapid and fertile development of probability and statistics during the recent few decades. As a person who loves mathematics as well as the real world, I long for the research with both theoretical depth in mathematics and practical influence in our lives. I found those interests perfectly combined in the study of statistics and risk management, from which this dissertation is finally generated.

The dissertation addresses three topics in the area of non-parametric statistical inference, multivariate dependence structures and their applications in risk management. As such, it consists of three main chapters, each of which addresses one topic.

Chapter II is dedicated to new empirical likelihood tests in high-dimensional data analysis. Four different classic test problems in the high-dimensional framework are considered: testing the equality of the mean of two samples (Section 2.2), testing the coefficient in a linear model (Section 2.3), testing the covariance matrix and testing the banded structure of the covariance matrix (Section 2.4).

Chapter III is dedicated to the applications of the jackknife empirical likelihood interval estimation to some quantities of interest in risk management, including the risk-distortion measures (Section 3.2), Spearman's rho (Section 3.3) and parametric copulas (Section 3.4).

Chapter IV is dedicated to the theory of a new class of probability distributions, called the *completely mixable distributions*. The definition, properties and main theorems about this new class are introduced. The new technique developed with this concept can be used to solve a series of problems in the Fréchet class and answer some questions in risk management.

This chapter, Chapter I, serves as the introduction. The existing statistical methods of likelihood ratio functions are reviewed in Section 1.1. The theory of copulas is introduced in Section 1.2. The problems of the Fréchet class are introduced in Sections 1.3.

1.1 Empirical likelihood methods

1.1.1 Parametric likelihood ratio

The parametric likelihood ratio function has become a common knowledge of statistics graduate students nowadays. Let us first review the definition of the likelihood ratio function. Throughout this section, let $X = (X_1, \cdots, X_n)$ be a sample of n i.i.d. observations from a distribution F_0 on $\mathbb{R}^p$, and define the likelihood function

$$L(F|X) = \prod_{i=1}^{n} f(X_i)$$

for any distribution function F, where $f(X_i)$ is the probability mass or density function of F at the point X_i, depending on the context. Since we are interested in the likelihood ratio, the case of having a probability density and the case of having a probability mass are the treated the same, as long as both the numerator and the denominator are using the same scale.

When we are interested in a parametric family of distributions $\{F(\theta) : \theta \in \Theta\}$, where Θ the set of parameters *theta*, it is called a parametric model. Suppose Θ is a vector space, and let Θ_0 be a subspace of Θ. Define the likelihood ratio function

$$\Lambda(\Theta_0) = \frac{\sup\{L(F(\theta)|X) : \ \theta \in \Theta_0\}}{\sup\{L(F(\theta)|X) : \ \theta \in \Theta\}}.$$

The Wilks' Theorem, presented by Wilks [109], is considered one of the most important results in the likelihood ratio problems. The theorem states that under $H_0 : \theta \in \Theta_0$ and mild regularity conditions,

$$-2 \log \Lambda(\Theta_0) \xrightarrow{d} \chi^2_q$$

where χ_q^2 is the chi-square distribution with q degrees of freedom and $q = \text{dim}\Theta - \text{dim}\Theta_0$. In particular, if Θ_0 is the set of one point, i.e. the real value of θ, then $-2\log\Lambda(\Theta_0) \xrightarrow{d} \chi^2_{\text{dim}\Theta}$.

Likelihood methods are very effective as they can be used to find efficient estimators and to construct tests with good power properties. Since the asymptotic limit of $-2\log\Lambda(\theta)$ does not depend on the underlying model, the method has great convenience in many cases. A likelihood ratio test is a test based on the statistic $\Lambda(\theta)$, to test $H_0 : \theta \in \Theta_0$ against $H_a : \theta \in \Theta \setminus \Theta_0$. By Wilks' Theorem, a test based on $l(\theta) := -2\log\Lambda(\theta)$ can be easily constructed by rejecting H_0 when $l(\theta)$ exceeds the threshold $\chi_p^2(1-\alpha)$, where $\chi_p^2(1-\alpha)$ is the $1-\alpha$ quantile of χ_p^2.

1.1.2 Empirical likelihood (EL) methods

The non-parametric version of the likelihood ratio function was first by introduced by Owen [71, 72]. First (and throughout Chapter II and Chapter III, unless otherwise notified), let us define the empirical distribution function (EDF) of X as $F_n(x) = \frac{1}{n}\sum_{i=1}^n I(X_i \le x)$.

As a well-known fact, F_n is the nonparametric maximum likelihood estimator for the true distribution function F_0, i.e.

$$L(F|X) \le L(F_n|X) = n^{-n} \tag{1.1}$$

for any distribution F and the equality holds only if $F = F_n$. The study on the EDF has been extensive; for more information we refer to Shorack and Wellner [94] and references therein.

(1.1) gives us an opportunity to build an analog to the parametric likelihood ratio function. Let $\mathcal{F}$ be the set of all distribution functions on $\mathbb{R}^p$ (recall that X_1 takes value in $\mathbb{R}^p$), and $\mathcal{F}_0$ be a subset of $\mathcal{F}$. Then we can define the empirical likelihood ratio function

$$\begin{aligned}\Lambda(\mathcal{F}_0) &= \frac{\sup\{L(F|X):\ F\in\mathcal{F}_0\}}{\sup\{L(F|X):\ F\in\mathcal{F}\}}\\ &= \frac{\sup\{L(F|X):\ F\in\mathcal{F}_0\}}{L(F_n|X)}\\ &= n^n\sup\{L(F|X):\ F\in\mathcal{F}_0\}.\end{aligned}$$

Now suppose we are interested in a quantity $\theta = T(F)$, where T is a functional of F. Let $\mathcal{F}_0(\theta)$ be the set of distributions F satisfying $T(F) = \theta$. In this case, define the empirical likelihood ratio function

$$R(\theta) = \Lambda(\mathcal{F}_0(\theta)) = n^n \sup\{L(F|X) : T(F) = \theta\}.$$

It is obvious that $\{L(F|X) : T(F) = \theta\}$ is only maximized when F is supported on the observations $X_1, \cdots, X_n$. Then $R(\theta)$ can be written as

$$R(\theta) = \sup\{\prod_{i=1}^{n}(np_i) : p_i = f(X_i), T(F) = \theta\}.$$

It is then straightforward to investigate the limit of $R(\theta)$. As one would expect from Wilks' Theorem, $-2\log R(\theta)$ should go to a chi-square distribution, with the number of degrees of freedom depending on the difference between $\mathcal{F}_0$ and $\mathcal{F}$. This turns out to be true when T is a linear functional of F. In particular, and as a good example, for the mean problem $T(F) = \mathbb{E}(X_1)$, Owen [72] gives the following theorem:

Theorem 1.1.1. *Let $X_1, \ldots, X_n$ be independent random vectors in R^p with common distribution F_0 having mean μ_0 and finite variance covariance matrix V_0 of rank $q > 0$. Then $l(\mu_0)$ converges in distribution to a χ^2_q random variable as $n \to \infty$, where $l(\mu_0) = -2\log R(\mu_0)$.*

Remark 1.1.1. Note that in the case $\theta = T(F) = \mathbb{E}(X_1)$, we have

$$R(\theta) = \sup\{\prod_{i=1}^{n}(np_i) :\ p_1 \geq 0, \cdots, p_n \geq 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i X_i = \theta\}. \qquad (1.2)$$

It is seen that θ should lie in the convex hull of the sample $(X_1, \cdots, X_n)$ to ensure the existence of a solution to the optimization (1.2). In general, when computing $R(\theta)$, θ should always lie in a convex hull formed by the sample.

The optimization problem in (1.2) can be done by the Lagrange multiplier method. In program `R`, there is a package `emplik` with which people can easily calculate the likelihood ratio function with given sample. In this thesis, we call a technique using the empirical likelihood in statistical testing and estimation an *empirical likelihood (EL) method*. For details and more information, we refer to Owen [71, 72, 73].

As another significant contribution to the empirical likelihood methods, Qin and Lawless [82] introduced the *estimating equations* to the empirical likelihood methods, making the methods more flexible with different types of model settings. Suppose we are interested in a parameter $\theta \in \mathbb{R}^q$ associated with the underlying distribution F through *estimating equations* $\mathbb{E}[g(X_1;\theta)] = 0$, where $\mathbb{E}[g(\cdot)]$ is a d-dimensional linear functional of the underlining distribution. Here d and q are the essential dimension of the functional g and parameter θ respectively, i.e. the components in g or θ are generated by a set of d or q linearly independent components. The empirical likelihood function with estimating equations is defined as

$$L(\theta) = \sup\{\prod_{i=1}^{n} p_i : \sum_{i=1}^{n} p_i g(X_i, \theta) = 0, p_i \geq 0, \sum_{i=1}^{n} p_i = 1\}. \tag{1.3}$$

Let $\tilde{\theta}$ maximize $L(\theta)$. Qin and Lawless [82] showed that under mild conditions, $-2\log(L(\theta_0)/L(\tilde{\theta})) \xrightarrow{d} \chi^2_r$, where $r = d \vee q$ and θ_0 is the true value of θ.

As a special case, if we are interested in the mean θ, then we can choose $G(x;\theta) = x - \theta$ and we will get $R(\theta)$ defined in (1.2).

Looking into the proofs in Owen [71, 72], in order to guarantee that $R(\theta)$ converges to a chi-square distribution, one will need the following conditions for some matrix Σ:

(L1) CLT,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(X_i;\theta) \xrightarrow{d} N(0,\Sigma).$$

(L2) LLN,

$$\frac{1}{n}\sum_{i=1}^{n} g(X_i;\theta)g(X_i;\theta)^T \xrightarrow{p} \Sigma.$$

(L3) Controlled maximum,

$$\max_{1\le i\le n} g(X_i;\theta) = o_p(\sqrt{n}).$$

Fortunately, since $g(X_i;\theta),\ldots,g(X_n;\theta)$ are i.i.d., (L1)–(L3) are guaranteed by a finite covariance matrix Σ of $g(X_i;\theta)$. However, this inspired us that as long as (L1)–(L3) are satisfied, Wilks' Theorem holds. Thus, the result can be applied with the method of resampling, where the sample is no longer i.i.d., but (L1)–(L3) still hold. Based on this observation, we will introduce the jackknife empirical likelihood methods later.

The merits of the empirical likelihood include: the shape of confidence regions is model-free as it is automatically determined using only the data; the estimation of the asymptotic variance is avoided; one can easily incorporate information using estimating equations; it is Bartlett correctable (see DiCiccio, Hall and Romano [29]). The method of empirical likelihood has been extensively studied in the past few decades. We refer to the recent review papers Chen and Van Keilegom [17] for a review of empirical likelihood in regression, and Chen, Peng and Qin [15] and Hjort, McKeague and Van Keilegom [45] for empirical likelihood in high-dimensional data analysis.

1.1.3 Jackknife empirical likelihood (JEL) methods

One notable limitation of the empirical likelihood method is that its works poorly with a nonlinear functional T.

Example 1.1.1. Assume $p = 1$ and we are interested in $\theta = \mathbb{E}(X_1 - \mathbb{E}X_1)^3$. We cannot write i.i.d. $g(X_i;\theta)$ in this case.

In general, the Wilks' Theorem does not hold when an empirical likelihood method is applied to nonlinear functionals. To overcome this difficulty, Jing, Yuan and Zhou [47] proposed a jackknife empirical likelihood (JEL) method for U-statistics to deal with nonlinear functionals.

The method of jackknife is a resampling method to reduce the variance of a statistic. The new sample, called the jackknife sample, is constructed by taking away one of the observations at each time. The jackknife sample is no longer independent, but under some mild conditions they are asymptotically i.i.d., hence (L1)–(L3) in a empirical likelihood method can be satisfied. See, e.g., Shao and Tu [93] for an introduction to the method of jackknife.

For a U-statistic, the procedure in Jing, Yuan and Zhou [47] is to construct a jackknife sample of the statistic, and then apply the standard empirical likelihood method for the mean of i.i.d. observations to the jackknife sample:

$$R(\theta) = \sup\{\prod_{i=1}^{n}(np_i) : \ p_1 \geq 0, \cdots, p_n \geq 0, \sum_{i=1}^{n} p_i = 1, \quad \sum_{i=1}^{n} p_i Z_i(\mathbf{X};\theta) = 0\}.$$

Here the function $g(X_i;\theta)$ in (1.3) is replaced by $Z_i(\mathbf{X};\theta)$, where $(Z_1, \cdots, Z_n)$ is a d-dimensional jackknife sample, with mean 0. $Z_1, \cdots, Z_n$ are no longer independent, but they could be asymptotically i.i.d to obtain Wilks' Theorem,

$$-2\log R(\theta_0) \xrightarrow{d} \chi_r^2,$$

where $r = d \vee q$.

Example 1.1.2. For $p = 1$, $\theta = \mathbb{E}(X_1 - \mathbb{E}X_1)^3$, let

$$\hat{\theta}_n = \frac{1}{n}\sum_{i=1}^{n}(X_i - \frac{1}{n}\sum_{k=1}^{n} X_k)^3$$

and

$$\hat{\theta}_{n,j} = \frac{1}{n-1}\sum_{i \neq j}^{n}(X_i - \frac{1}{n-1}\sum_{k \neq j} X_k)^3.$$

Define the jackknife sample as $Z_i = n\hat{\theta}_n - (n-1)\hat{\theta}_{n,i}$, then

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

and $R(\theta) \xrightarrow{d} \chi_1^2$ under some mild regularity conditions.

Inspired by the conditions used in the standard empirical likelihood method, to prove that the JEL version of Wilks' Theorem holds for any statistic, not necessarily a U-statistic, one needs to verify that the jackknife sample satisfies (R1)–(R3):

(R1) CLT,

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

(R2) LLN,

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

(R3) Controlled maximum,

$$\max_{1\leq i\leq n} Z_i = o_p(\sqrt{n}).$$

Theorem 1.1.2. (Wilks' Theorem for JEL.) *Assuming (R1)–(R3), then*

$$-2\log R(\theta_0) \xrightarrow{d} \chi_r^2$$

where θ_0 is the true value of θ and $r = d \vee q$.

Proof follows from standard arguments in empirical likelihood, see e.g. Owen [72]. In this thesis, the above technique will be frequently used. In Chapter II, we will investigate the use of the empirical likelihood in high-dimensional testing problems. In Chapter III, we will discuss the applications of the jackknife empirical likelihood in risk management.

1.2 Copulas

A copula is a multivariate function which characterizes the dependence structure among random variables without the information of the marginal distributions. The technique of using copulas has been very popular in statistics and actuarial science, see Nelsen [69] for an introduction to copulas. The concept of copulas has become a common knowledge in the modern research related to dependence structures.

Over the last few decades, researchers in economics, financial mathematics and actuarial science have introduced results related to the dependence structure in their own respective fields of interest. Below we list a few examples of multivariate dependence in finance and insurance.

1. Pricing financial derivatives written on several assets.
2. Structured financial products, such as the CDOs.
3. Portfolio selection and hedging.
4. Best and worst scenarios in risk management.
5. Time series analysis and econometrics.

The dependence itself is known to be mathematically mysterious and it can be dangerous if misplaced. Many people believe that the methodology of applying the Gaussian copula to model the dependence is one of the reasons behind the global financial crisis in 2008-2009; see the well-known article by Salmon [88].

1.2.1 Definition and Sklar's Theorem

As the copulas are widely used in the study of dependence related problems, in this section we briefly review the concept of copulas.

Definition 1.2.1. An n-copula $C : [0,1]^n \to [0,1]$ is a function that satisfies the following properties:

(1) C is grounded, i.e. $C(u_1, \cdots, u_{i-1}, 0, u_{i+1}, \cdots, u_n) = 0$ for any $1 \leq i \leq n$ and $u_j \in [0,1]$, $j \neq i$.

(2) C is n-increasing, i.e. for each hyperrectangle B in $I^n = [0,1]^n$ the C-volume of B is non-negative.

(3) For all $u \in [0,1]$, $C(1, \cdots, 1, u, 1, \cdots, 1) = u$, where the i-th variate is u and all the other variates are 1, for any $1 \leq i \leq n$.

It is easily checked that n-copula C have the following properties:

(i) $C(u_1, \cdots, u_n)$ is non-decreasing with respect to u_i, $i = 1, \cdots, n,$.

(ii) For all $u_i, v_i \in [0,1]$, $i = 1, \cdots, n$,

$$C(u_1, \cdots, u_n) - C(v_1, \cdots, v_n) \leq \sum_{i=1}^{n} |u_i - v_i|.$$

(iii) For $1 \leq m \leq n$, $C(u_1, \cdots, u_m, 1, \cdots, 1)$ is an m-copula.

(iv) Let $M_n(u_1, \cdots, u_n) = \min\{u_i, i \leq n\}$, $W_n(u_1, \cdots, u_n) = \max\{u_1 + u_2 + \cdots + u_n - (n-1), 0\}$, for $u_i \in [0,1]$, $1 \leq i \leq n$, then

$$W_n(u_1, \cdots, u_n) \leq C(u_1, \cdots, u_n) \leq M_n(u_1, \cdots, u_n).$$

M_n is called the *Fréchet upper bound* and W_n is called the *Fréchet lower bound*. Note that M_n is a copula for all n, and W_n is a copula only when $n = 1, 2$.

Remark 1.2.1. The complete names of Fréchet bounds are *Fréchet–Hoeffding bounds*, attributed to both Hoeffding [46] and Fréchet [40].

The main property of the copulas was first introduced by Sklar's Theorem [95]. The theorem shows that a copula itself is a multivariate distribution function, and it is one-to-one corresponding to a joint distribution when the marginal distributions are given.

Theorem 1.2.1. (Sklar's Theorem) *Let F be a joint distribution function with univariate marginal distributions $F_1, \cdots, F_n$. Then there exists a copula C such that*

$$F(x_1, \cdots, x_n) = C(F_1(x_1), \cdots, F_n(x_n)). \tag{1.4}$$

If $F_1, \cdots, F_n$ are continuous, then C is unique.

Conversely, let $F_1, \cdots, F_n$ be univariate distributions and C be a n-copula, then F in (1.4) *is a joint distribution function with univariate marginal distributions $F_1, \cdots, F_n$.*

For the random variables $X_1, \cdots, X_n$ with joint distribution F and marginal distributions $F_1, \cdots, F_n$, we say the copula of $X_1, \cdots, X_n$ or the vector $(X_1, \cdots, X_n)$ is C if C is defined by (1.4). From Sklar's Theorem, $X_1, \cdots, X_n$ are independent if and only if the copula C of $X_1, \cdots, X_n$ is $C(u_1, \cdots, u_n) = u_1 u_2 \cdots u_n$.

Let $F_i(x) = x$, $i = 1, \cdots, n$ we easily obtain that a copula is the joint distribution function of uniform distributions. This statement is usually regarded as a equivalent definition of copulas.

Definition 1.2.2. An n-copula is a joint distribution function of n U$[0, 1]$ random variables.

The following theorem gives the invariant property of copulas under a strictly increasing transformation of random variables.

Theorem 1.2.2. *For strictly increasing transformations H_i, $i = 1, \cdots, n$, The copula of $H_1(X_1) \cdots, H_n(X_n)$ is identical to the copula of $X_1, \cdots, X_n$.*

The above theorem allows people to transform any random variables to uniform random variables and study the copula. This technique is widely used in statistical inference of copulas, for example, using the rank statistics to estimate or test copulas.

Theorem 1.2.3. *For a 2-copula C, for fixed $v \in [0,1]$, $\frac{\partial}{\partial u}C(u,v)$ exists for almost all $u \in [0,1]$, and*

$$0 \le \frac{\partial}{\partial u}C(u,v) \le 1.$$

If we exchange the positions of u and v, the theorem still holds.

For proofs in this section and more details and applications about the copulas, the readers are referred to Nelsen [69]. Statistical inference for copulas has been studied extensively. The pseudo maximum likelihood estimator for parametric copulas, presented Genest, Ghoudi and Rivest [42], is most relevant to the content in Chapter III of this thesis. Peng, Qi and Van Keilegom [75] proposed a smoothed jackknife empirical likelihood method to construct confidence intervals for a non-parametric copula. We refer to the references in Genest, Ghoudi and Rivest [42], Embrechts, Lindskog and McNeil [33] and Nelsen [69] for more information on the theory, applications and statistical inference of copulas.

1.2.2 Spearman's rho and Kendall's tau

A copula contains all the information about a dependence structure, since the set of copulas is one-to-one corresponding to the set of joint distributions when marginal distributions are given and continuous. In the practice of actuarial science and finance, it is more convenient and clear to use quantities instead of functions to measure dependence, due to computational difficulties. Spearman's rho and Kendall's tau are two commonly used measures of dependence between two random variables.

Let $(X_1, Y_1), (X_2, Y_2)$ be independent random vectors with distribution function H and continuous marginals $F(x) = H(x, \infty)$ and $G(y) = H(\infty, y)$. Then the Kendall's tau and the Spearman's rho of (X_1, Y_1) are defined as

$$\tau = \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0]$$

and

$$\rho^s = 12\mathbb{E}[(F(X_1) - 1/2)(G(Y_1) - 1/2)],$$

respectively.

It is well-known that τ and ρ^s depends only on the copula C of X_1 and Y_1; see Nelsen [69] for instance. Moreover, τ and ρ^s has the copula representation:

$$\tau = 4 \iint_{[0,1]^2} C(x,y) dC(x,y) - 1,$$

and

$$\rho^s = 12 \iint_{[0,1]^2} C(x,y) dx dy - 3.$$

As measures of dependence, τ and ρ^s enjoy the following property.

Theorem 1.2.4. *Let C, τ and ρ^s be the copula, the Kendall's tau and the Spearman's rho of (X, Y), respectively. Then*

(a) $C = M_2 \Leftrightarrow \tau = 1 \Leftrightarrow \rho^s = 1.$

(b) $C = W_2 \Leftrightarrow \tau = -1 \Leftrightarrow \rho^s = -1.$

(c) $C(u,v) = uv \Rightarrow \tau = \rho^s = 0.$

For a proof, see Embrechts, McNeil, and Straumann [34]. Note that although the independence of X, Y implies $\tau = \rho^s = 0$, the converse is not true.

Statistical inferences on the above dependence measures can be found in Nelsen [69]. The Spearman's rho is also extended to the multivariate case by Schmid and Schmidt [91] and Nelsen and Úbeda-Flores [70].

In this thesis, we will investigate the statistical estimation problems related to copulas and Spearman's rho in Chapter III and solve Fréchet Class problems using the method of copulas in Chapter IV. As an application, we also find an lower bound for the multivariate version of Spearman's rho in Chapter IV.

1.3 Fréchet Class Problems

1.3.1 Fréchet classes

As mentioned in Section 1.2, the dependence structure plays an important role in the recent research of actuarial science, mathematical finance and risk management.

Among the topics related to the dependence, one notable setting is called the problem of *Fréchet class*. A Fréchet class is a class of random vectors with given marginal distributions, usually denoted by $\mathfrak{F}_n(F_1, F_2, \cdots, F_n)$. Let $\mathbf{X} = (X_1, \cdots, X_n)$ is a random vector in $\mathbb{R}^n$, and the Fréchet class is defined as

$$\mathfrak{F}_n(F_1, F_2, \cdots, F_n) = \{\mathbf{X} : \ X_i \sim F_i, \ i = 1, \cdots, n\},$$

where n is the number of individual risks and $F_1, \cdots, F_n$ are the n marginal distributions. As the simplest case, $\mathfrak{F}_n(F, \cdots, F)$ is the set of random vectors with identical given marginal distribution F. It is obvious that a random vector in a Fréchet class is one-to-one corresponding to a copula. No surprise that copula methods are widely used in the study of Fréchet classes.

The name of the Fréchet class comes from the result on the convex upper bound in any Fréchet class, which is usually attributed to both Hoeffding [46] and Fréchet [40] as mentioned in Section 1.2. In their seminal papers, it was provided that

$$F_{\mathbf{X}}(x_1, \cdots, x_n) \leq \min\{F_1(x_1), \cdots, F_n(x_n)\}$$

for any random vector $\mathbf{X} \in \mathfrak{F}_n(F_1, F_2, \cdots, F_n)$ with distribution function $F_{\mathbf{X}}$. This bound is exactly due to the Fréchet upper bound M_n as mentioned in Section 1.2. The result is closely related to the concepts of *comonotonicity* and *stochastic ordering*. The readers are referred to Deelstra, Dhaene and Vanmaele [24] for an overview of the comonotonicity and its applications in finance, and Shaked [92] for an introduction and summary of the stochastic ordering.

The Fréchet class problems are important in the practice of modern risk management, simply because statistically estimating the joint distribution of a random vector is usually much more difficult than estimating the marginal distributions from the accessible data in the financial market today. Therefore, using the bounds instead helps one to manage risks and uncertainty. Unfortunately, although the upper bound in the convex-ordering sense was given more than half a century ago, the attempts to find the lower bounds of $F_{\mathbf{X}}$ have never been that successful, as W_n is no longer a copula for $n \geq 3$.

As a more general class of problems, it has been asked for a long time to find the bounds on the distribution of $\psi(\mathbf{X})$,

$$m_\psi(s) = \inf\{\mathbb{P}(\psi(\mathbf{X}) < s) : \mathbf{X} \in \mathfrak{F}_n(F_1, F_2, \cdots, F_n)\}, \tag{1.5}$$

for for a function ψ. Makarov [63], in response to a question formulated earlier by A.N. Kolmogorov, provided the first result of $n = 2$ and $\psi = +$, the sum operator. An elegant and important duality result was later given by Rüschendorf [85]:

$$\begin{aligned} m_\psi(s) =& 1 - \inf\left\{\sum_{i=1}^{n} \int f_i dF_i : f_i \text{ are bounded measurable functions on } \mathbb{R} \text{ s.t.} \right. \\ & \left. \sum_{i=1}^{n} f_i(x_i) \geq 1_{[s,+\infty)}(\psi(x_1, \cdots, x_n)), \text{ for all } x_i \in \mathbb{R},\ i = 1, \cdots, n \right\}. \end{aligned} \tag{1.6}$$

However, this dual optimization is still hard to solve in general.

In the next sections, we will summarize the recent attempts made to solve the problems of bounds in Fréchet classes.

1.3.2 Bounds on the distribution of the total risk

Among different choices of ψ in (1.5), $\psi(\mathbf{X}) = +(\mathbf{X}) = X_1 + \cdots + X_n$ is extensively studied due to its nice mathematical properties and important applications in practice, as $\psi(\mathbf{X})$ is the total risk or the joint portfolio of individual risks or assets in this case.

Let $\mathbf{X} = (X_1, \cdots, X_n) \in \mathfrak{F}_n(F_1, F_2, \cdots, F_n)$ be a risk vector with known marginal distributions $F_1, \cdots, F_n$. Denote by $S = X_1 + \cdots + X_n$ the total risk. Researchers are looking for the best-possible bounds for the distribution of the total risk S, namely

$$m_+(s) = \inf\{\mathbb{P}(S < s) : \mathbf{X} \in \mathfrak{F}_n(F_1, F_2, \cdots, F_n)\}, \tag{1.7}$$

and

$$M_+(s) = \sup\{\mathbb{P}(S < s) : \mathbf{X} \in \mathfrak{F}_n(F_1, F_2, \cdots, F_n)\}. \tag{1.8}$$

The bounds $m_+(s)$ and $M_+(s)$ directly lead to the sharp bounds on quantile-based risk measures of S. In practice, the managers of investment banks are more interested in the Value-at-Risk of a joint portfolio. The Value-at-Risk (VaR) at level α is defined as

$$\mathrm{VaR}_\alpha(S) = \inf\{s \in \mathbb{R} : \mathbb{P}(S \leq s) \geq \alpha\}.$$

The bounds on the above VaR are called the worst (best) Value-at-Risk scenarios and are given by the inverse functions of the bounds $m_+(s)$ and $M_+(s)$.

Rüschendorf [85] first found $m_+(s)$ when all marginal distributions have the same uniform or binomial distribution, where the techniques of the duality (1.6) were employed. A complete analysis of this kind of problems was given in Rachev and Rüschendorf [83]. After the 1982 paper [85], no significant results were given for about fifteen years.

In the 1990s, the method of copulas has became more and more popular. As the ultimate modern tool for modeling dependence, copulas kicked in and helped with solving the Fréchet class problems (1.7) and (1.8). The papers of P. Embrechts at ETHZ and his colleagues were considered the most relevant during the last decade. Denuit, Genest and Marceau [26] and Embrechts, Höing and Juri [32] used copulas to obtain the so-called *standard bounds* and discussed some applications. The standard bounds are no longer sharp for $n \geq 3$. Embrechts and Puccetti [35] provided a better lower bound which is still not sharp, in the case when all marginal distributions are the same and continuous. Some results when partial information on the dependence structure were also given in that paper. Embrechts and Höing [31] provided a geometric interpretation to highlight the shape of the dependence structures with the worst VaR scenarios. Embrechts and Puccetti [36] extended this problem to multivariate marginal distributions and provided results similar to the univariate case. Kaas, Laeven and Nelsen [54] studied the worst VaR scenarios for the case when partial information on some measure of dependence is known.

Finally, we refer to Embrechts and Puccetti [37] for an overview on the importance and applications of problems (4.2) and (4.3) in quantitative risk management.

1.3.3 Bounds on other quantities

Related to the Fréchet class, another classic problem in simulation and variance reduction is to minimize the variance of the sum S of random variables $X_1, \cdots, X_n$ with given marginal distributions, i.e.

$$\inf\{\mathrm{Var}(S) : \mathbf{X} \in \mathfrak{F}_n(F_1, F_2, \cdots, F_n)\}. \tag{1.9}$$

Fishman [38] and Hammersley and Handscomb [43] present good introduction and references on this problem. It is well-known that for $n = 2$ the solution is given by the antithetic variates $X_1 = F_1^-(U)$ and $X_2 = F_2^-(1-U)$ where F^- is the inverse cdf of P and U is uniform on [0,1]. For $n \geq 3$ the problem is generally difficult to solve.

A more general version of the problem (1.9) is

$$\inf\{\mathbb{E}f(S) : \mathbf{X} \in \mathfrak{F}_n(F_1, F_2, \cdots, F_n)\}. \tag{1.10}$$

There are many special cases of (1.10), such as the variance minimization problem (1.9), the minimum of expected product

$$\inf\{\mathbb{E}(X_1 \cdots X_n) : \mathbf{X} \in \mathfrak{F}_n(F_1, F_2, \cdots, F_n)\}, \tag{1.11}$$

and bounds on the stop-loss premium

$$\inf\{\mathbb{E}[(X_1 + \cdots + X_n - t)_+] : \mathbf{X} \in \mathfrak{F}_n(F_1, F_2, \cdots, F_n)\}, \tag{1.12}$$

where $(\cdot)_+ = \max(\cdot, 0)$. Many of the special cases are related to various topics in statistics, risk theory, copulas and stochastic orders. (1.11) is directly linked to the lower bound of the multivariate Spearman's rho introduced by Schmid and Schmidt [91].

Studies for $n \geq 3$ have been done mostly in the homogenous (when $F_1 = \cdots = F_n = F$). Gaffke and Rüschendorf [41] proposed to find a dependence structure to concentrate S around its expectation as much as possible, since it is obvious $S = c$ is an optimal solution to (1.9) if such constant c exists. Then it follows a question: for which F, S is possibly a constant? Gaffke and Rüschendorf [41] studied the property of possible $S = c$ in the case of uniform distributions and binomial distributions. The case of distributions with symmetric and unimodal density was studied for $n = 3$ by Knott and Smith [60, 61] and for the general case $n \geq 2$ by Rüschendorf and Uckelmann [87] using a different method. The property was also extended to multivariate distributions in Rüschendorf and Uckelmann [87].

In Chapter IV, we will present a new concept called *complete mixability* distributions. The new technique developed here can be used to solve (1.7) (1.8) and (1.10) in the case of F is a completely mixable distribution, or F is a distribution with monotone density on its support. This result completes the convex ordering bounds in the Fréchet class $\mathfrak{F}_n(F, F, \cdots, F)$ for F with a monotone density.

CHAPTER II

EMPIRICAL LIKELIHOOD TESTS FOR HIGH-DIMENSIONAL DATA ANALYSIS

High-dimensional (HD) data analysis is arguably one of the most popular topics in the research of statistics nowadays. The developments on this topic have been very significant, with a wide range of applications. The phenomena of high-dimensionality appears extensively in genomics, economics, finance, linguistics and many other fields of the modern science. We refer to the book Cai and Shen [13] for a review of the recent developments and applications of the HD data analysis. In this chapter, we will investigate four testing problems within the HD framework, using the methods of the empirical likelihood. The contents in this chapter is mainly based on the following preprints.

1. Wang, R., Peng, L. and Qi, Y. (2012). Jackknife empirical likelihood test for equality of two high dimensional means. *Preprint*.

2. Peng, L., Qi, Y. and Wang, R. (2012). Empirical likelihood test for high-dimensional linear models. *Preprint*.

3. Zhang, R., Peng, L. and Wang, R. (2012). Tests for covariance matrix with fixed or divergent dimension. *Preprint*.

2.1 Introduction, Notations and Regularity Conditions

In this chapter, we investigate the testing problems associated with an array of i.i.d. p-dimensional vectors $X_i = X_i^{(n)} = (X_{i,1}^{(n)}, \cdots, X_{i,p}^{(n)})$ for $i = 1, \cdots, n$. When p is fixed and small, conventional tests such as the Hotelling T^2 test perform well both theoretically and computationally. However, if the dimension p approaches infinity as the sample size n goes to infinity, the classic methods do not work in general; see [5, 15, 16] for instance, and this phenomena will be discussed later in the following sections.

The classic testing problems are of our interest, where all quantities may depend on n and p.

(i) Suppose $(X_1, \cdots, X_{n_1})$ and $(Y_1, \cdots, Y_{n_2})$ are two independent random samples with sample sizes n_1, n_2 and unknown means μ_1, μ_2 respectively. Consider the testing problem

$$H_0 : \mu_1 = \mu_2 \text{ against } H_1 : \mu_1 \neq \mu_2. \tag{2.1}$$

(ii) Suppose $X_1, \cdots, X_n$ are independent and $Y_i = \beta^T X_i + \epsilon_i$, for $i = 1, \cdots, n$, where $\beta = (\beta_1, \cdots, \beta_p)^T$ is the vector of unknown parameters and $\epsilon_1, \cdots \epsilon_n$ are iid random errors. Consider the testing problem

$$H_0 : \beta = \beta_0 \text{ against } H_1 : \beta \neq \beta_0. \tag{2.2}$$

(iii) Suppose $X_1, \cdots, X_n$ are independent with an unknown covariance matrix $\Sigma = (\sigma_{ij})_{p\times p}$. Consider the testing problem

$$H_0 : \Sigma = \Sigma_0 \text{ against } H_1 : \Sigma \neq \Sigma_0. \tag{2.3}$$

(iv) Similar to (iii), consider the testing problem

$$H_0 : \sigma_{ij} = 0 \text{ for all } |i-j| \geq \tau \;\text{ against }\; H_1\text{: } H_0 \text{ is false.} \tag{2.4}$$

In this chapter, to apply new empirical likelihood methods to those problems, the following regularity condition will be frequently used.

(P). An estimator T with sample size n satisfies condition (P) if $\mathbb{E}T^2 > 0$ and for some $\delta > 0$,

$$\frac{\mathbb{E}|T|^{2+\delta}}{(\mathbb{E}T^2)^{1+\delta/2}} = o(n^{\frac{\delta+\min(\delta,2)}{4}}).$$

For example, if $\mathbb{E}(T^4)/(\mathbb{E}(T^2))^2 = o(n)$, then T satisfies (P) with $\delta = 2$. Note that this condition is generally satisfied by Gaussian random vectors.

Condition (P) is concise and necessary to guarantee the conditions (L1)–(L3) used in the empirical likelihood. However, (P) is sometimes inconvenient to check when the estimator T is complicated. Hence, we propose the following two models.

In the following models, let $X = (X_1, \cdots, X_n)$ be a random sample of size n, with mean μ and covariance matrix Σ, and $\lambda_1 \le \cdots \le \lambda_d$ be the p eigenvalues of the matrix Σ.

(A). A random sample X of size n satisfies condition (A) if

(A1) $0 < \liminf_{n\to\infty} \lambda_1 \le \limsup_{n\to\infty} \lambda_p < \infty$.

(A2) For some $\delta > 0$, $\frac{1}{p}\sum_{i=1}^{p} \mathbb{E}|X_{1,i} - \mu_i|^{2+\delta} = O(1)$, and

(A3) $p = o(m^{\frac{\delta+\min(\delta,2)}{2(2+\delta)}})$.

Condition (A3) is a somewhat restrictive condition for the dimension p. Note that conditions (A1) and (A2) are related only to the covariance matrices and some higher moments on the components of the random vectors. The higher moments we have, the less restriction is imposed on p. Condition (A3) can be removed for models with some special dependence structures. For comparison purpose, we will also consider the following model (B) used in Bai and Saranadasa [5], Chen, Peng and Qin [15] and Chen and Qin [16].

(B). (Factor model.) A random sample X of size n satisfies condition (B) if

$$X_i = \Gamma B_i + \mu_1$$

for $i = 1, \cdots, n_1$, where Γ is a $p\times k$ matrix with $\Gamma\Gamma^T = \Sigma$, $\{B_i = (B_{i,1}, \cdots, B_{i,k})^T\}_{i=1}^{n_1}$ is an independent random sample satisfying that $\mathbb{E}B_i = 0$, $\mathrm{Var}(B_i) = I_{k\times k}$,

$\mathbb{E}B_{i,j}^4 = 3 + \xi < \infty$, and

$$\mathbb{E}\prod_{l=1}^{k} B_{i,i_l}^{\nu_l} = \prod_{l=1}^{k} \mathbb{E}B_{i,i_l}^{\nu_l}$$

when $\nu_1 + \cdots + \nu_k = 4$ for distinct nonnegative integers i_l's.

In the problem of testing covariance matrices (Section 2.4), a stronger condition on the moment is imposed due to the effect of high-order statistics. Therefore, we list the alternative model (B') below for covariance testing problems.

(B'). A random sample X of size n satisfies condition (B') if (B) holds, and each of $B_{i,j}$ in (B) has uniformly bounded 8th moment, and

$$\mathbb{E}\prod_{l=1}^{k} B_{i,i_l}^{\nu_l} = \prod_{l=1}^{k} \mathbb{E}B_{i,i_l}^{\nu_l}$$

when $\nu_1 + \cdots + \nu_k = 8$ for distinct nonnegative integers $\imath_l$'s.

The idea of constructing tests in this chapter is as follows. In order to test H_0: a vector parameter $v = 0$ (e.g. in problem (i) $v = \mu_1 - \mu_2$), we first find an estimator T such that $\mathbb{E}(T) = 0$ is equivalent to H_0. Then we use $\mathbb{E}(T) = 0$ as the estimating equation to apply the empirical likelihood method. Such a test may not be powerful; we add one more linear functional to enhance the power of the test. The methods are new and they usually require a weaker assumption on the model compared to existing work in the literature. Most of the proofs in this chapter are justifying conditions (R1)–(R3). Lastly, it is worth mentioning that the power of the tests proposed in this chapter perform better in the case of dense model (i.e. in the alternative hypothesis, many components of $v \neq 0$), rather than the sparse model (i.e. in the alternative hypothesis, many components of $v = 0$).

The rest of this chapter is organized as follows. In Section 2.2, we present a jackknife empirical likelihood test for problem (i). An empirical likelihood test for problem (ii) is introduced in Section 2.3. Tests for problem (iii) and (iv) are discussed in Section 2.4. In each section, there are separate subsections of an introduction, main results, simulation studies and proofs.

2.2 Test for Equality of Two High-dimensional Means

It has been a long history to test the equality of two multivariate means. One popular test is the so-called Hotelling T^2 test. However, as the dimension diverges, the Hotelling T^2 test performs poorly due to the possible inconsistency of the sample covariance estimation. To overcome this issue and allow the dimension to diverge as fast as possible, Bai and Saranadasa [5] and Chen and Qin [16] proposed tests without the sample covariance involved, and derived the asymptotic limits which depend on whether the dimension is fixed or diverges under a specific multivariate model. In this section, we propose a jackknife empirical likelihood test which has a chi-square limit independent of the dimension, and the conditions are much weaker than those in the existing methods. A simulation study shows that the proposed new test has a very robust size with respect to the dimension, and is powerful too.

2.2.1 Introduction

Suppose $X = \{X_i = (X_{i,1}, \cdots, X_{i,p})^T : i = 1, \ldots, n_1\}$ and $Y = \{Y_j = (Y_{j,1}, \cdots, Y_{j,p})^T : j = 1, \ldots, n_2\}$ are two independent random samples with means μ_1 and μ_2, respectively. It has been a long history to test $H_0 : \mu_1 = \mu_2$ against $H_a : \mu_1 \neq \mu_2$ for a fixed dimension p. When both X_1 and Y_1 have a multivariate normal distribution with equal covariance, the well-known test is the so-called Hotelling T^2 test defined as

$$T^2 = \eta(\bar{\mathbf{X}} - \bar{\mathbf{Y}})^T A_n^{-1}(\bar{\mathbf{X}} - \bar{\mathbf{Y}}), \tag{2.5}$$

where $\eta = \frac{(n_1+n_2-2)n_1 n_2}{n_1+n_2}$, $\bar{\mathbf{X}} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i$, $\bar{\mathbf{Y}} = \frac{1}{n_2}\sum_{i=1}^{n_2} Y_i$ and $A_n = \sum_{i=1}^{n_1}(X_i - \bar{\mathbf{X}})(X_i - \bar{\mathbf{X}})^T + \sum_{i=1}^{n_2}(Y_i - \bar{\mathbf{Y}})(Y_i - \bar{\mathbf{Y}})^T$. However, when $p = p(n_1, n_2) \to \infty$, the Hotelling T^2 test performs poorly due to the possible inconsistency of the sample covariance estimation. When $p/(n_1 + n_2) \to c \in (0, 1)$, Bai and Saranadasa [5] derived the asymptotic power of T^2. To overcome the restriction $c < 1$, Bai and

Saranadasa [5] proposed to employ

$$M_n = (\bar{\mathbf{X}} - \bar{\mathbf{Y}})^T(\bar{\mathbf{X}} - \bar{\mathbf{Y}}) - \eta^{-1}\mathrm{tr}(A_n)$$

instead of T^2 under a special multivariate model without assuming multivariate normality while keeping the condition of equal covariance, and derived the asymptotic limit when $p/(n_1+n_2) \to c > 0$. Recently Chen and Qin [16] proposed to use the following test statistic

$$CQ = \frac{\sum_{i\neq j}^{n_1} X_i^T X_j}{n_1(n_1-1)} + \frac{\sum_{i\neq j}^{n_2} Y_i^T Y_j}{n_2(n_2-1)} - 2\frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} X_i^T Y_j}{n_1 n_2} \tag{2.6}$$

in order to allow p to be a possible larger order than that in Bai and Saranadasa [5]. Again, the asymptotic limit of the proposed test statistic CQ depends on whether the dimension is fixed or diverges, which results in either a normal limit or a chi-square limit, and special models for $\{X_i\}$ and $\{Y_i\}$ are employed. Another modification of Hotelling T^2 test is proposed by Srivastava and Du [97] and Srivastava [96] with the covariance matrix replaced by a diagonal matrix. Rates of convergence for high dimensional means are studied by Kuelbs and Vidyashankar [59]). For nonasymptotic studies of high dimensional means, we refer to Arlot, Blanchard and Roquain [2, 3]. Here, we are interested in seeking a test which does not need to distinguish whether the dimension is fixed or diverges.

By noting that $\mu_1 = \mu_2$ is equivalent to $(\mu_1-\mu_2)^T(\mu_1-\mu_2) = 0$, one may think of applying an empirical likelihood test to the estimating equation $\mathbb{E}\{(X_{i_1} - Y_{j_1})^T(X_{i_2} - Y_{j_2})\} = 0$ for $i_1 \neq i_2$ and $j_1 \neq j_2$. If one directly applies the empirical likelihood method based on estimating equations proposed in Qin and Lawless [82] by using the samples $X_1, \cdots, X_{n_1}$ and $Y_1, \cdots, Y_{n_2}$, then one may define the empirical likelihood function

$$\sup\left\{\prod_{i=1}^{n_1}(n_1 p_i)\prod_{j=1}^{n_2}(n_2 q_j) : p_1 \geq 0, \cdots, p_{n_1} \geq 0, q_1 \geq 0, \cdots, q_{n_2} \geq 0, \sum_{i=1}^{n_1} p_i = 1,\right.$$
$$\left.\sum_{j=1}^{n_2} q_j = 1, \sum_{i_1=1}^{n_1}\sum_{i_2\neq i_1}\sum_{j_1=1}^{n_2}\sum_{j_2\neq j_1}(p_{i_1}X_{i_1} - q_{j_1}Y_{j_1})^T(p_{i_2}X_{i_2} - q_{j_2}Y_{j_2}) = 0\right\},$$

which makes the minimization unsolvable. The reason is that the estimating equation defines a nonlinear functional, and in general one has to linearize the nonlinear functional before applying the empirical likelihood method. For more details on empirical likelihood methods, we refer to Owen [73] and the review paper of Chen and Van Keilegom [17]. Recently, Jing, Yuan and Zhou [47] proposed a so-called jackknife empirical likelihood method to construct confidence regions for nonlinear functionals with a particular focus on U-statistics. Using this idea, one needs to construct a jackknife sample based on the following estimator $n_1^{-1}(n_1-1)^{-1}n_2^{-1}(n_2-1)^{-1}\sum_{i_1\neq i_2}^{n_1}\sum_{j_1\neq j_2}^{n_2}(X_{i_1}-Y_{j_1})^T(X_{i_2}-Y_{j_2})$, which equals the statistic CQ given in (2.6). However, in order to have the jackknife empirical likelihood method work, one has to show that $\sqrt{n_1n_2}n_1^{-1}(n_1-1)^{-1}n_2^{-1}(n_2-1)^{-1}\sum_{i_1\neq i_2}^{n_1}\sum_{j_1\neq j_2}^{n_2}(X_{i_1}-Y_{j_1})^T(X_{i_2}-Y_{j_2})$ has a normal limit when $\mu_1=\mu_2$. Consider $n_1=n_2=n, d=1, \mu_1=\mu_2$. Then it is easy to see that

$$
\begin{aligned}
& n^{-1}(n-1)^{-2}\sum_{i_1\neq i_2}^{n}\sum_{j_1\neq j_2}^{n}(X_{i_1}-Y_{j_1})^T(X_{i_2}-Y_{j_2}) \\
= \quad & \frac{1}{n-1}\{\sum_{i=1}^{n}(X_i-Y_i)\}^2-\frac{1}{n-1}\sum_{i=1}^{n}(X_i-Y_i)^2+\frac{2}{n(n-1)}\sum_{i=1}^{n}X_i\sum_{j=1}^{n}Y_j \\
& -\frac{2}{(n-1)}\sum_{i=1}^{n}X_iY_i \\
\xrightarrow{p} \quad & \{N(0,\mathbb{E}(X_1-Y_1)^2)\}^2-\mathbb{E}(X_1-Y_1)^2
\end{aligned}
$$

which does not have a normal limit as $n\to\infty$. Hence a direct application of the jackknife empirical likelihood method to the statistic CQ will not lead to a chi-square limit.

In this section, we propose a novel way to formulate a jackknife empirical likelihood test for testing $H_0:\mu_1=\mu_2$ against $H_a:\mu_1\neq\mu_2$ by dividing the samples into two parts. The proposed new test has no need to distinguish whether the dimension is fixed or goes to infinity. It turns out that the asymptotic limit of the new test under H_0 is a chi-square limit independent of the dimension, the conditions on p and random variables $\{X_i\}$ and $\{Y_j\}$ are weaker too. A simulation study shows that the size of the new test is quite stable with respect to the dimension and the proposed test is powerful as well.

We organize the whole section as follows. In Section 2.2.2, the new methodology and main results are given. Section 2.2.3 presents a simulation study and a real data analysis. All proofs are put in Section 2.2.4.

2.2.2 Methodology

As mentioned in Section 2.2.1, throughout assume $X_i = (X_{i,1}, \cdots, X_{i,p})^T$ for $i = 1, \cdots, n_1$ and $Y_j = (Y_{j,1}, \cdots, Y_{j,p})^T$ for $j = 1, \cdots, n_2$ are two independent random samples with means μ_1 and μ_2, respectively. Assume $\min\{n_1, n_2\}$ goes to infinity. The question is to test $H_0 : \mu_1 = \mu_2$ against $H_a : \mu_1 \neq \mu_2$. Since $\mu_1 = \mu_2$ is equivalent to $(\mu_1 - \mu_2)^T(\mu_1 - \mu_2) = 0$ and $\mathbb{E}(X_{i_1} - Y_{j_1})^T(X_{i_2} - Y_{j_2}) = (\mu_1 - \mu_2)^T(\mu_1 - \mu_2)$ for $i_1 \neq i_2$ and $j_1 \neq j_2$, we propose to apply the jackknife empirical likelihood method to the above estimating equation. As explained in the introduction, a direct application fails to have a chi-square limit. Here we propose to split the samples into two groups as follows.

Put $m_1 = [n_1/2]$, $m_2 = [n_2/2]$, $m = m_1 + m_2$, $\bar{X}_i = X_{i+m_1}$ for $i = 1, \cdots, m_1$, and $\bar{Y}_j = Y_{j+m_2}$ for $j = 1, \cdots, m_2$. First we propose to estimate $(\mu_1 - \mu_2)^T(\mu_1 - \mu_2)$ by

$$\frac{1}{m_1 m_2} \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} (X_i - Y_j)^T(\bar{X}_i - \bar{Y}_j), \tag{2.7}$$

which is less efficient than the statistic CQ. However, it allows us to add more estimating equations and to employ the empirical likelihood method without estimating the asymptotic covariance. By noting that $\mathbb{E}\{(X_i - Y_j)^T(\bar{X}_i - \bar{Y}_j)\} = (\mu_1 - \mu_2)^T(\mu_1 - \mu_2) = ||\mu_1 - \mu_2||^2$ instead of $O(||\mu_1 - \mu_2||)$, one may expect that a test based on (2.7) will not be powerful for a small value of $||\mu_1 - \mu_2||$, confirmed by a brief simulation study. In order to improve the power, we propose to apply the jackknife empirical likelihood method in Jing, Yuan and Zhou [47] to both (2.7) and a linear functional such as

$$\frac{1}{m_1 m_2}\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\{(X_i - Y_j)^T\mathbf{1}_p + (\bar{X}_i - \bar{Y}_j)^T\mathbf{1}_p\} \tag{2.8}$$

rather than only (2.7), where $\mathbf{1}_p = (1, \cdots, 1)^T$. Note that equation (2.8) can be replaced by another linear functional or several functionals with at least one linear functional to further improve the power. With prior information on the model or more specific alternative hypothesis, some linear functionals can be chosen to replace (2.8) so as to improve the power of the test. With no additional information, any linear functional is a possible choice theoretically. Simulation study suggests that applying the jackknife empirical likelihood to (2.7) and (2.8) results in a test with good power and quite robust size with respect to the dimension.

As in Jing, Yuan and Zhou [47], based on (2.7) and (2.8), we formulate the jackknife sample as $Z_k = (Z_{k,1}, Z_{k,2})^T$ for $k = 1, \cdots, m$, where

$$\begin{cases} Z_{k,1} = & \frac{m_1+m_2}{m_1 m_2}\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}(X_i - Y_j)^T(\bar{X}_i - \bar{Y}_j) \\ & -\frac{m_1+m_2-1}{(m_1-1)m_2}\sum_{i\neq k, i=1}^{m_1}\sum_{j=1}^{m_2}(X_i - Y_j)^T(\bar{X}_i - \bar{Y}_j) \\ Z_{k,2} = & \frac{m_1+m_2}{m_1 m_2}\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\{(X_i - Y_j)^T\mathbf{1}_p + (\bar{X}_i - \bar{Y}_j)^T\mathbf{1}_p\} \\ & -\frac{m_1+m_2-1}{(m_1-1)m_2}\sum_{i\neq k, i=1}^{m_1}\sum_{j=1}^{m_2}\{(X_i - Y_j)^T\mathbf{1}_p + (\bar{X}_i - \bar{Y}_j)^T\mathbf{1}_p\} \end{cases}$$

for $k = 1, \cdots, m_1$, and

$$\begin{cases} Z_{k,1} = & \frac{m_1+m_2}{m_1 m_2}\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}(X_i - Y_j)^T(\bar{X}_i - \bar{Y}_j) \\ & -\frac{m_1+m_2-1}{m_1(m_2-1)}\sum_{i=1}^{m_1}\sum_{j\neq k-m_1, j=1}^{m_2}(X_i - Y_j)^T(\bar{X}_i - \bar{Y}_j) \\ Z_{k,2} = & \frac{m_1+m_2}{m_1 m_2}\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\{(X_i - Y_j)^T\mathbf{1}_p + (\bar{X}_i - \bar{Y}_j)^T\mathbf{1}_p\} \\ & -\frac{m_1+m_2-1}{m_1(m_2-1)}\sum_{i=1}^{m_1}\sum_{j\neq k-m_1, j=1}^{m_2}\{(X_i - Y_j)^T\mathbf{1}_p + (\bar{X}_i - \bar{Y}_j)^T\mathbf{1}_p\} \end{cases}$$

for $k = m_1 + 1, \cdots, m$. Based on this jackknife sample, the jackknife empirical likelihood function for testing $H_0 : \mu_1 = \mu_2$ is defined as

$$L_m = \sup\{\prod_{i=1}^{m}(mp_i) : p_1 \geq 0, \cdots, p_m \geq 0, \sum_{i=1}^{m} p_i = 1, \sum_{i=1}^{m} p_i Z_i = (0,0)^T\}.$$

By the Lagrange multiplier technique, we have $p_k = m^{-1}\{1+\beta^T Z_k\}^{-1}$ for $k = 1, \cdots, m$ and $l_m = -2\log L_m = 2\sum_{i=1}^m \log\{1+\beta^T Z_k\}$, where β satisfies

$$\frac{1}{m}\sum_{i=1}^{m} \frac{Z_k}{1+\beta^T Z_k} = (0,0)^T. \tag{2.9}$$

Write $\Sigma = (\sigma_{ij})_{1\le i\le p, 1\le j\le p} = \mathbb{E}\{(X_1-\mu_1)(X_1-\mu_1)^T\}$, the covariance matrix of X_1, and use $\lambda_1 \le \cdots \le \lambda_p$ to denote the p eigenvalues of the matrix Σ. Similarly, write $\bar{\Sigma} = (\bar{\sigma}_{ij})_{1\le i\le p, 1\le j\le p} = \mathbb{E}\{(Y_1-\mu_2)(Y_1-\mu_2)^T\}$ and use $\bar{\lambda}_1 \le \cdots \le \bar{\lambda}_p$ to denote the p eigenvalues of the matrix $\bar{\Sigma}$. Also write

$$\rho_1 = \sum_{i,j=1}^{p} \sigma_{i,j}^2, \;\; \rho_2 = \sum_{i,j=1}^{p} \bar{\sigma}_{i,j}^2, \;\; \tau_1 = 2\sum_{i,j=1}^{p} \sigma_{i,j}, \;\; \tau_2 = 2\sum_{i,j=1}^{p} \bar{\sigma}_{i,j}. \tag{2.10}$$

Note that $\rho_1 = \mathbb{E}[(X_1-\mu_1)^T(X_1-\mu_1)]^2$, $\rho_2 = \mathbb{E}[(Y_1-\mu_2)^T(Y_1-\mu_2)]^2$, $\tau_1 = 2\mathbb{E}[\mathbf{1}_p^T(X_1-\mu_1)]^2$ and $\tau_2 = 2\mathbb{E}[\mathbf{1}_p^T(Y_1-\mu_2)]^2$, and these quantities may depend on n_1, n_2 since p may depend on n_1, n_2.

Theorem 2.2.1. *Assume* $\min\{n_1, n_2\} \to \infty$, *$\tau_1$ and τ_2 in (2.10) are positive, and for some $\delta > 0$,*

$$\frac{\mathbb{E}|(X_1-\mu_1)^T(\bar{X}_1-\mu_1)|^{2+\delta}}{\rho_1^{(2+\delta)/2}} = o(m_1^{\frac{\delta+\min(\delta,2)}{4}}), \tag{2.11}$$

$$\frac{\mathbb{E}|(Y_1-\mu_2)^T(\bar{Y}_1-\mu_2)|^{2+\delta}}{\rho_2^{(2+\delta)/2}} = o(m_2^{\frac{\delta+\min(\delta,2)}{4}}), \tag{2.12}$$

$$\frac{\mathbb{E}|\mathbf{1}_p^T(X_1+\bar{X}_1-2\mu_1)|^{2+\delta}}{\tau_1^{(2+\delta)/2}} = o(m_1^{\frac{\delta+\min(\delta,2)}{4}}), \tag{2.13}$$

and

$$\frac{\mathbb{E}|\mathbf{1}_p^T(Y_1+\bar{Y}_1-2\mu_2)|^{2+\delta}}{\tau_2^{(2+\delta)/2}} = o(m_2^{\frac{\delta+\min(\delta,2)}{4}}). \tag{2.14}$$

Then, under $H_0 : \mu_1 = \mu_2$, l_m converges in distribution to a chi-square distribution with two degrees of freedom as $\min\{n_1, n_2\} \to \infty$.

Based on the above theorem, one can test $H_0 : \mu_1 = \mu_2$ against $H_a : \mu_1 \ne \mu_2$ by rejecting H_0 when $l_m \ge \chi^2_{2,\gamma}$, where $\chi^2_{2,\gamma}$ denotes the $(1-\gamma)-$quantile of a chi-square distribution with two degrees of freedom and γ is the significant level.

Remark 2.2.1. Conditions (2.11)–(2.14) can be rephrased as $(X_1 - \mu_1)^T(\bar{X}_1 - \mu_1)$, $(Y_1 - \mu_2)^T(\bar{Y}_1 - \mu_2)$ $\mathbf{1}_p^T(X_1 + \bar{X}_1 - 2\mu_1)$ and $\mathbf{1}_p^T(Y_1 + \bar{Y}_1 - 2\mu_2)$ satisfy condition (P).

Remark 2.2.2. In (2.11)–(2.14), the restrictions are put on $\mathbb{E}|W|^{2+\delta}/(\mathbb{E}W^2)^{(2+\delta)/2}$ for some random variables W, which are necessary for the CLT to hold for random arrays. Later we will see those conditions are easily satisfied by imposing some conditions on the higher-order moments or special dependence structure.

Remark 2.2.3. The proposed test has the following merits:

1. The limiting distribution is always chi-square without estimating the asymptotic covariance.

2. It does not require any specific structure such as the one used in Bai and Saranadasa [5] and Chen and Qin [16], which will be discussed later.

3. With higher-order moment condition or special dependence structure of $\{X_i\}$ and $\{Y_i\}$, p can be very large.

4. There is no restriction imposed on the relation between n_1 and n_2 except that $\min\{n_1, n_2\} \to \infty$. That is, no need to assume a limit or bound on the ratio n_1/n_2. Moreover, no assumptions are needed on ρ_1/ρ_2 or τ_1/τ_2. Hence the covariance matrices Σ_1 and Σ_2 can be arbitrary as long as τ_1,$\tau_2 > 0$, which are simply equivalent to $\sum_{i=1}^p X_{1,i}$ and $\sum_{i=1}^p Y_{1,i}$ are non-degenerate random variables.

Next we verify Theorem 2.2.1 with model (A) and (B).

Corollary 2.2.2. *Assume* $\min\{n_1, n_2\} \to \infty$*, X and Y satisfy (A). Then, under $H_0 : \mu_1 = \mu_2$, conditions* (2.11) – (2.14) *are satisfied, i.e., Theorem 2.2.1 holds.*

Theorem 2.2.3. *Assume τ_1 and τ_2 in (2.10) are positive and X and Y satisfy* (B)*. Then under $H_0 : \mu_1 = \mu_2$, l_m converges in distribution to a chi-square distribution with two degrees of freedom as* $\min\{n_1, n_2\} \to \infty$.

Remark 2.2.4. It can be seen from the proof of Theorem 2.2.3 that assumptions $\mathbb{E}B_{i,j}^4 = 3 + \xi_1 < \infty$ in model (B) can be replaced by the much weaker conditions $\max_{1\le j\le k} \mathbb{E}B_{1,j}^4 = o(m)$. Unlike Bai and Saranadasa [5] and Chen and Qin [16], there is no restriction on p and k for our proposed method. The only constraint imposed on matrices Γ_1 and Γ_2 is that both $\sum_{i=1}^p X_{1,i}$ and $\sum_{i=1}^p Y_{1,i}$ are non-degenerate, which is very weak.

2.2.3 Simulation study

We investigate the finite sample behavior of the proposed jackknife empirical likelihood test (JEL) and compare it with the test statistic in (2.6) proposed by Chen and Qin [16] in terms of both size and power.

Let $W_1, \cdots, W_p$ be iid random variables with distribution function $N(0,1)$, and let $\bar{W}_1, \cdots, \bar{W}_p$, independent of $W_i's$ be iid random variables with distribution function $t(8)$. Put $X_{1,1} = W_1, X_{1,2} = W_1 + W_2, \cdots, X_{1,p} = W_{d-1} + W_p$, $Y_{1,1} = \bar{W}_1 + \mu_{2,1}, Y_{1,2} = \bar{W}_1 + \bar{W}_2 + \mu_{2,2}, \cdots, Y_{1,p} = \bar{W}_{d-1} + \bar{W}_p + \mu_{2,p}$, where $\mu_{2,i} = c_1$ if $i \le [c_2 p]$, and $\mu_{2,i} = 0$ if $i > [c_2 p]$. That is, $100c_2\%$ of the components of Y_1 have a shifted mean compared to that of X_1.

Since we test $H_0 : \mathbb{E}X_1 = \mathbb{E}Y_1$ against $H_a : \mathbb{E}X_1 \ne \mathbb{E}Y_1$, the case of $c_1 = 0$ denotes the size of tests. By drawing $1,000$ random samples of sizes $n_1 = 30, 100, 150$ from $X = (X_{1,1}, \cdots, X_{1,p})^T$ and independently drawing $1,000$ random samples of sizes $n_2 = 30, 100, 200$ from $Y = (Y_{1,1}, \cdots, Y_{1,p})^T$ with $d = 10, 20, \cdots, 100, 300, 500$, $c_1 = 0, 0.1$ and $c_2 = 0.25, 0.75$, we calculate the powers of the two tests mentioned above.

In Tables 2.1–2.3, we report the empirical sizes and powers for the proposed jackknife empirical likelihood test and the test in Chen and Qin [16] at level 5%. Results for level 10% are similar. From these three tables, we observe that (i) the size of both tests, i.e., results for $c_1 = 0$ is quite stable with respect to the dimension p; (ii) the

Table 2.1: Sizes and powers of the proposed jackknife empirical likelihood test (JEL) and the test in Chen and Qin [16] (CQ) are reported for the case of $(n_1, n_2) = (30, 30)$ at level 5%.

p	JEL $c_1 = 0$ $c_2 = 0.25$	CQ $c_1 = 0$ $c_2 = 0.25$	JEL $c_1 = 0.1$ $c_2 = 0.25$	CQ $c_1 = 0.1$ $c_2 = 0.25$	JEL $c_1 = 0.1$ $c_2 = 0.75$	CQ $c_1 = 0.1$ $c_2 = 0.75$
10	0.070	0.049	0.071	0.049	0.072	0.062
20	0.056	0.037	0.057	0.049	0.096	0.060
30	0.064	0.047	0.066	0.049	0.113	0.066
40	0.070	0.052	0.069	0.058	0.116	0.072
50	0.067	0.049	0.083	0.054	0.138	0.067
60	0.063	0.039	0.069	0.043	0.174	0.055
70	0.053	0.053	0.076	0.065	0.190	0.081
80	0.056	0.059	0.063	0.067	0.191	0.082
90	0.056	0.044	0.080	0.054	0.204	0.071
100	0.066	0.060	0.082	0.064	0.229	0.091
300	0.056	0.045	0.114	0.054	0.537	0.092
500	0.049	0.051	0.160	0.063	0.731	0.110

proposed jackknife empirical likelihood test is more powerful than the test in Chen and Qin [16] for the case of $c_2 = 0.75$ and the case when the data is sparse, but p is large (i.e., the case of $c_1 = 0.1, c_2 = 0.25$). Since equation (2.8) has nothing to do with sparsity, it is expected that the proposed jackknife empirical likelihood method is not powerful when the data is sparse. Hence, it would be of interest to connect sparsity with some estimating equations so as to improve the power of the proposed jackknife empirical likelihood test.

In conclusion, the proposed jackknife empirical likelihood test has a very stable size with respect to the dimension and is powerful under the dense model. Moreover, the new test is easy to compute, flexible to take other information into account, and works for both fixed dimension and divergent dimension.

Table 2.2: Sizes and powers of the proposed jackknife empirical likelihood test (JEL) and the test in Chen and Qin [16] (CQ) are reported for the case of $(n_1, n_2) = (100, 100)$ at level 5%.

p	JEL $c_1 = 0$ $c_2 = 0.25$	CQ $c_1 = 0$ $c_2 = 0.25$	JEL $c_1 = 0.1$ $c_2 = 0.25$	CQ $c_1 = 0.1$ $c_2 = 0.25$	JEL $c_1 = 0.1$ $c_2 = 0.75$	CQ $c_1 = 0.1$ $c_2 = 0.75$
10	0.074	0.054	0.072	0.063	0.099	0.090
20	0.043	0.047	0.053	0.055	0.145	0.098
30	0.047	0.047	0.056	0.063	0.191	0.115
40	0.051	0.050	0.063	0.062	0.264	0.125
50	0.055	0.040	0.077	0.061	0.326	0.131
60	0.055	0.044	0.077	0.067	0.374	0.151
70	0.043	0.051	0.063	0.086	0.395	0.150
80	0.042	0.059	0.082	0.079	0.474	0.171
90	0.043	0.040	0.098	0.065	0.527	0.163
100	0.049	0.054	0.091	0.088	0.575	0.194
300	0.048	0.054	0.217	0.102	0.974	0.389
500	0.049	0.041	0.353	0.115	0.999	0.544

Table 2.3: Sizes and powers of the proposed jackknife empirical likelihood test (JEL) and the test in Chen and Qin [16] (CQ) are reported for the case of $(n_1, n_2) = (150, 200)$ at level 5%.

p	JEL $c_1 = 0$ $c_2 = 0.25$	CQ $c_1 = 0$ $c_2 = 0.25$	JEL $c_1 = 0.1$ $c_2 = 0.25$	CQ $c_1 = 0.1$ $c_2 = 0.25$	JEL $c_1 = 0.1$ $c_2 = 0.75$	CQ $c_1 = 0.1$ $c_2 = 0.75$
10	0.048	0.054	0.054	0.062	0.129	0.116
20	0.055	0.042	0.078	0.075	0.237	0.166
30	0.052	0.054	0.079	0.081	0.330	0.207
40	0.039	0.035	0.070	0.068	0.430	0.212
50	0.039	0.048	0.071	0.094	0.480	0.231
60	0.047	0.051	0.092	0.095	0.598	0.273
70	0.046	0.051	0.086	0.107	0.658	0.309
80	0.042	0.047	0.113	0.109	0.753	0.327
90	0.046	0.043	0.148	0.098	0.781	0.346
100	0.048	0.059	0.141	0.117	0.821	0.365
300	0.044	0.040	0.370	0.163	1	0.703
500	0.047	0.045	0.555	0.235	1	0.899

2.2.4 Proofs

In the proofs we use $||\cdot||$ to denote the L_2 norm of a vector or matrix. Since $\mu_1-\mu_2$ is our target and under null hypothesis $\mu_1-\mu_2=0$, without loss of generality we assume $\mu_1=\mu_2=0$. Write $u_{ij}=(X_i-Y_j)^T(\bar{X}_i-\bar{Y}_j)$ and $v_{ij}=(X_i-Y_j)^T\mathbf{1}_p+(\bar{X}_i-\bar{Y}_j)^T\mathbf{1}_p$ for $1\le i\le m_1, 1\le j\le m_2$. Then it is easily verified that for $1\le i,k\le m_1, 1\le j,l\le m_2$,

$$\mathbb{E}(u_{ij})=\mathbb{E}(v_{kl})=\mathbb{E}(u_{ij}v_{kl})=0,$$

$$\mathrm{Var}(u_{kl})=\sum_{i,j=1}^{p}(\sigma_{i,j}^2+\bar{\sigma}_{i,j}^2)=\rho_1+\rho_2,$$

and

$$\mathrm{Var}(v_{kl})=2\sum_{i,j=1}^{p}(\sigma_{i,j}+\bar{\sigma}_{i,j})=\tau_1+\tau_2.$$

Lemma 2.2.4. *Under conditions of Theorem 1, we have as* $\min\{n_1,n_2\}\to\infty$

$$\frac{1}{\sqrt{m_1}}\sum_{i=1}^{m_1}\frac{X_i^T\bar{X}_i}{\sqrt{\rho_1}}\xrightarrow{d}N(0,1), \tag{2.15}$$

$$\frac{1}{\sqrt{m_2}}\sum_{j=1}^{m_2}\frac{Y_j^T\bar{Y}_j}{\sqrt{\rho_2}}\xrightarrow{d}N(0,1), \tag{2.16}$$

$$\frac{1}{\sqrt{m_1}}\sum_{i=1}^{m_1}\frac{\mathbf{1}_p^T(X_i+\bar{X}_i)}{\sqrt{\tau_1}}\xrightarrow{d}N(0,1), \tag{2.17}$$

and

$$\frac{1}{\sqrt{m_2}}\sum_{j=1}^{m_2}\frac{\mathbf{1}_p^T(Y_j+\bar{Y}_j)}{\sqrt{\tau_2}}\xrightarrow{d}N(0,1). \tag{2.18}$$

Proof. Since $\mathrm{Var}(X_i^T\bar{X}_i)=\rho_1$ and $X_1^T\bar{X}_1,\cdots,X_{m_1}^T\bar{X}_{m_1}$ are i.i.d. for fixed m_1, equation (2.15) follows from (2.11) and the Lyapunov central limit theorem. The rest can be shown in the same way. □

From now on we denote

$$\rho=\frac{m}{m_1}\rho_1+\frac{m}{m_2}\rho_2 \quad \text{anp} \quad \tau=\frac{m}{m_1}\tau_1+\frac{m}{m_2}\tau_2.$$

Lemma 2.2.5. *Under conditions of Theorem 1, we have as* $\min\{n_1, n_2\} \to \infty$

$$\frac{\sqrt{m}}{m_1 m_2 \sqrt{\rho}} \sum_{i=1}^{m_1} X_i^T \sum_{j=1}^{m_2} \bar{Y}_j \xrightarrow{p} 0, \tag{2.19}$$

$$\frac{1}{m_1 \sqrt{\tau}} \sum_{i=1}^{m_1} \mathbf{1}_p^T X_i \xrightarrow{p} 0, \tag{2.20}$$

$$\frac{1}{m_2 \sqrt{\tau}} \sum_{j=1}^{m_2} \mathbf{1}_p^T Y_j \xrightarrow{p} 0, \tag{2.21}$$

$$\frac{1}{m_1} \sum_{i=1}^{m_1} \frac{(X_i^T \bar{X}_i)^2}{\rho_1} \xrightarrow{p} 1, \tag{2.22}$$

$$\frac{1}{m_2} \sum_{j=1}^{m_2} \frac{(Y_j^T \bar{Y}_j)^2}{\rho_2} \xrightarrow{p} 1, \tag{2.23}$$

$$\frac{1}{m_1} \sum_{i=1}^{m_1} \frac{[\mathbf{1}_p^T (X_i + \bar{X}_i)]^2}{\tau_1} \xrightarrow{p} 1, \tag{2.24}$$

$$\frac{1}{m_2} \sum_{j=1}^{m_2} \frac{[\mathbf{1}_p^T (Y_j + \bar{Y}_j)]^2}{\tau_2} \xrightarrow{p} 1, \tag{2.25}$$

$$\frac{1}{m_1} \sum_{i=1}^{m_1} \frac{X_i^T \bar{X}_i [\mathbf{1}_p^T (X_i + \bar{X}_i)]}{\sqrt{\rho_1 \tau_1}} \xrightarrow{p} 0, \tag{2.26}$$

and

$$\frac{1}{m_2} \sum_{i=1}^{m_2} \frac{Y_j^T \bar{Y}_j [\mathbf{1}_p^T (Y_j + \bar{Y}_j)]}{\sqrt{\rho_2 \tau_2}} \xrightarrow{p} 0. \tag{2.27}$$

Proof. Note that $\mu_1 = \mu_2 = 0$ are assumed in Section 4. Then (2.19) follows from the fact that

$$
\begin{aligned}
\operatorname{Var}(\frac{\sqrt{m}}{m_1 m_2\sqrt{\rho}}\sum_{i=1}^{m_1} X_i^T \sum_{j=1}^{m_2} Y_j) &= \mathbb{E}\left[\frac{m}{m_1^2 m_2^2 \rho}\left(\sum_{i=1}^{m_1}\sum_{j=1}^{m_2} X_i^T Y_j\right)^2\right] \\
&= \mathbb{E}\left[\frac{m}{m_1^2 m_2^2 \rho}\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}(X_i^T Y_j)^2\right] \\
&= \mathbb{E}\left[\frac{m}{m_1 m_2 \rho}(X_1^T Y_1)^2\right] \\
&= \frac{m}{m_1 m_2 \rho}\sum_{i,j=1}^{p}\sigma_{ij}\bar{\sigma}_{ij} \\
&\leq \frac{m}{m_1+m_2}\frac{\rho_1+\rho_2}{2\rho} \\
&\leq \frac{1}{2}(\frac{1}{m_1}+\frac{1}{m_2}) \\
&= o(1).
\end{aligned}
$$

In the same way, we can show (2.20) and (2.21).

To show (2.22), write $u_i = X_i^T \bar{X}_i$. We need to estimate $\mathbb{E}|\sum_{i=1}^{m_1} u_i^2 - m_1\rho_1|^{(2+\delta)/2}$. Note that $\rho_1 = \mathbb{E}u_1^2$. When $0 < \delta \leq 2$, it follows from von Bahr and Esseen [98] that

$$
\mathbb{E}|\sum_{i=1}^{m_1} u_i^2 - m_1\rho_1|^{(2+\delta)/2} \leq 2m_1\mathbb{E}|u_1^2 - \mathbb{E}(u_1^2)|^{(2+\delta)/2} = O(m_1\mathbb{E}|u_1|^{2+\delta}). \tag{2.28}
$$

When $\delta > 2$, it follows from Dharmadhikari and Jogdeo [28] that

$$
\mathbb{E}|\sum_{i=1}^{m_1} u_i^2 - m_1\rho_1|^{(2+\delta)/2} = O(m_1^{(2+\delta)/4}\mathbb{E}|u_1^2 - \mathbb{E}(u_1^2)|^{(2+\delta)/2}) = O(m_1^{(2+\delta)/4}\mathbb{E}|u_1|^{2+\delta}). \tag{2.29}
$$

Therefore, by (2.28), (2.29) and (2.11) we have for any $\varepsilon > 0$

$$
\begin{aligned}
&\mathbb{P}(|\frac{\sum_{i=1}^{m_1} u_i^2}{m_1\rho_1} - 1| > \varepsilon) \\
\leq\ & \varepsilon^{-(2+\delta)/2}\frac{\mathbb{E}|\sum_{i=1}^{m_1} u_i^2 - m\rho_1|^{(2+\delta)/2}}{(m_1\rho_1)^{(2+\delta)/2}} \\
=\ & O(m_1^{-(\delta+\min(\delta,2))/4}\mathbb{E}|\frac{u_1}{\sqrt{\rho_1}}|^{2+\delta}) \\
=\ & o(1),
\end{aligned}
$$

which implies (2.22). The rest can be shown in the same way. □

Lemma 2.2.6. *Under conditions of Theorem 1, we have as* $\min\{n_1, n_2\} \to \infty$

$$\frac{\sqrt{m}}{m_1 m_2} \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} \begin{pmatrix} \frac{u_{ij}}{\sqrt{\rho}} \\ \frac{v_{ij}}{\sqrt{\tau}} \end{pmatrix} \xrightarrow{d} N(0, I_2), \tag{2.30}$$

$$\frac{m}{m_1^2 m_2^2 \rho} \sum_{k=1}^{m_1} \left(\sum_{j=1}^{m_2} u_{kj} \right)^2 - \frac{m\rho_1}{m_1 \rho} \xrightarrow{p} 0, \tag{2.31}$$

$$\frac{m}{m_1^2 m_2^2 \rho} \sum_{k=1}^{m_2} \left(\sum_{i=1}^{m_1} u_{ik} \right)^2 - \frac{m\rho_2}{m_2 \rho} \xrightarrow{p} 0, \tag{2.32}$$

$$\frac{m}{m_1^2 m_2^2 \tau} \sum_{k=1}^{m_1} \left(\sum_{j=1}^{m_2} v_{kj} \right)^2 - \frac{m\tau_1}{m_1 \tau} \xrightarrow{p} 0, \tag{2.33}$$

$$\frac{m}{m_1^2 m_2^2 \tau} \sum_{k=1}^{m_2} \left(\sum_{i=1}^{m_1} v_{ik} \right)^2 - \frac{m\tau_2}{m_1 \tau} \xrightarrow{p} 0, \tag{2.34}$$

$$\frac{m}{m_1^2 m_2^2 \sqrt{\rho\tau}} \sum_{k=1}^{m_1} \left(\sum_{i=1}^{m_2} u_{ki} \sum_{j=1}^{m_2} v_{kj} \right) \xrightarrow{p} 0, \tag{2.35}$$

$$\frac{m}{m_1^2 m_2^2 \sqrt{\rho\tau}} \sum_{k=1}^{m_2} \left(\sum_{i=1}^{m_1} u_{ik} \sum_{j=1}^{m_1} v_{jk} \right) \xrightarrow{p} 0, \tag{2.36}$$

where I_2 *is the* 2×2 *identity matrix.*

Proof. It follows from Lemma 2.2.5 that

$$\begin{aligned}
& \frac{\sqrt{m}}{m_1 m_2} \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} \frac{u_{ij}}{\sqrt{\rho}} \\
= & \frac{\sqrt{m}}{m_1 m_2 \sqrt{\rho}} \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} (X_i^T \bar{X}_i + Y_j^T \bar{Y}_j - X_i^T \bar{Y}_j - Y_j^T \bar{X}_i) \\
= & \frac{\sqrt{m}}{m_1 \sqrt{\rho}} \sum_{i=1}^{m_1} X_i^T \bar{X}_i + \frac{\sqrt{m}}{m_2 \sqrt{\rho}} \sum_{j=1}^{m_2} Y_j^T \bar{Y}_j - \frac{\sqrt{m}}{m_1 m_2 \sqrt{\rho}} \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} (X_i^T \bar{Y}_j + Y_j^T \bar{X}_i) \\
= & \frac{\sqrt{m\rho_1}}{\sqrt{m_1 \rho}} \frac{1}{\sqrt{m_1}} \sum_{i=1}^{m_1} \frac{X_i^T \bar{X}_i}{\sqrt{\rho_1}} + \frac{\sqrt{m\rho_2}}{\sqrt{m_2 \rho}} \frac{1}{\sqrt{m_2}} \sum_{j=1}^{m_2} \frac{Y_j^T \bar{Y}_j}{\sqrt{\rho_2}} + o_p(1) \\
= & a_m A_m + b_m B_m + o_p(1),
\end{aligned}$$

where $a_m = \frac{\sqrt{m\rho_1}}{\sqrt{m_1 \rho}}$, $b_m = \frac{\sqrt{m\rho_2}}{\sqrt{m_2 \rho}}$,

$$A_m = \frac{1}{\sqrt{m_1}} \sum_{i=1}^{m_1} \frac{X_i^T \bar{X}_i}{\sqrt{\rho_1}} \xrightarrow{d} N(0, 1)$$

and

$$B_m = \frac{1}{\sqrt{m_2}} \sum_{j=1}^{m_2} \frac{Y_j^T \bar{Y}_j}{\sqrt{\rho_2}} \xrightarrow{d} N(0,1).$$

Obviously $a_m^2 + b_m^2 = 1$ and A_m, B_m are independent. Denote the characteristic functions of A_m and B_m by Φ_m and Ψ_m, respectively. Then,

$$\begin{aligned}
\mathbb{E}\exp(it(a_m A_m + b_m B_m)) &= \mathbb{E}\exp(ita_m A_m)\mathbb{E}\exp(itb_m B_m) \\
&= \Phi_m(ta_m)\Psi_m(tb_m) \\
&= [\exp(-\frac{(ta_m)^2}{2}) + o(1)][\exp(-\frac{(tb_m)^2}{2}) + o(1)] \\
&= \exp(-\frac{t^2}{2}) + o(1),
\end{aligned}$$

i.e.,

$$\frac{\sqrt{m}}{m_1 m_2} \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} \frac{u_{ij}}{\sqrt{\rho}} \xrightarrow{d} N(0,1). \tag{2.37}$$

Similarly, we have

$$\begin{aligned}
\frac{\sqrt{m}}{m_1 m_2} \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} \frac{v_{ij}}{\sqrt{\tau}} &= \frac{\sqrt{m\tau_1}}{\sqrt{m_1\tau}} \frac{1}{\sqrt{m_1}} \sum_{i=1}^{m_1} \frac{(X_i + \bar{X}_i)^T \mathbf{1}_p}{\sqrt{\tau_1}} - \frac{\sqrt{m\tau_2}}{\sqrt{m_2\tau}} \frac{1}{\sqrt{m_2}} \sum_{j=1}^{m_2} \frac{(Y_j + \bar{Y}_j)^T \mathbf{1}_p}{\sqrt{\tau_2}} \\
&\xrightarrow{d} N(0,1).
\end{aligned}$$

Let a and b be two real numbers with $a^2 + b^2 \neq 0$. Note that

$$\begin{aligned}
&\frac{\sqrt{m}}{m_1 m_2} \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} (a\frac{u_{ij}}{\sqrt{\rho}} + b\frac{v_{ij}}{\sqrt{\tau}}) \\
= \; & a\left(\frac{\sqrt{m\rho_1}}{\sqrt{m_1\rho}} \frac{1}{\sqrt{m_1}} \sum_{i=1}^{m_1} \frac{X_i^T \bar{X}_i}{\sqrt{\rho_1}} - \frac{\sqrt{m\rho_2}}{\sqrt{m_2\rho}} \frac{1}{\sqrt{m_2}} \sum_{j=1}^{m_2} \frac{Y_j^T \bar{Y}_j}{\sqrt{\rho_2}} \right) \\
& +b\left(\frac{\sqrt{m\tau_1}}{\sqrt{m_1\tau}} \frac{1}{\sqrt{m_1}} \sum_{i=1}^{m_1} \frac{(X_i + \bar{X}_i)\mathbf{1}_p}{\sqrt{\tau_1}} + \frac{\sqrt{m\tau_2}}{\sqrt{m_2\tau}} \frac{1}{\sqrt{m_2}} \sum_{j=1}^{m_2} \frac{(Y_j + \bar{Y}_j)^T \mathbf{1}_p}{\sqrt{\tau_2}} \right) + o_p(1) \\
= \; & \left(\frac{a\sqrt{m\rho_1}}{\sqrt{m_1\rho}} \frac{1}{\sqrt{m_1}} \sum_{i=1}^{m_1} \frac{X_i^T \bar{X}_i}{\sqrt{\rho_1}} + \frac{b\sqrt{m\tau_1}}{\sqrt{m_1\tau}} \frac{1}{\sqrt{m_1}} \sum_{i=1}^{m_1} \frac{(X_i + \bar{X}_i)\mathbf{1}_p}{\sqrt{\tau_1}} \right) \\
& + \left(\frac{a\sqrt{m\rho_2}}{\sqrt{m_2\rho}} \frac{1}{\sqrt{m_2}} \sum_{j=1}^{m_2} \frac{Y_j^T \bar{Y}_j}{\sqrt{\rho_2}} - \frac{b\sqrt{m\tau_2}}{\sqrt{m_2\tau}} \frac{1}{\sqrt{m_2}} \sum_{j=1}^{m_2} \frac{(Y_j + \bar{Y}_j)^T \mathbf{1}_p}{\sqrt{\tau_2}} \right) + o_p(1) \\
= \; & I_1 + I_2 + o_p(1).
\end{aligned}$$

Since $\frac{\sqrt{m\rho_1}}{\sqrt{m_1\rho}}, \frac{\sqrt{m\rho_2}}{\sqrt{m_2\rho}}, \frac{\sqrt{m\tau_1}}{\sqrt{m_1\tau}}, \frac{\sqrt{m\tau_2}}{\sqrt{m_2\tau}}$ are all bounded by one, it is easy to check that I_1 and I_2 satisfy the Lyapunov condition by (2.11) - (2.14). Therefore

$$\{a^2\frac{m\rho_1}{m_1\rho} + b^2\frac{m\tau_1}{m_1\tau}\}^{-1/2} I_1 \xrightarrow{p} N(0,1)$$

and

$$\{a^2\frac{m\rho_2}{m_2\rho} + b^2\frac{m\tau_2}{m_2\tau}\}^{-1/2} I_2 \xrightarrow{p} N(0,1).$$

Since $X_i's$ are independent of $Y_i's$, it follows from the same arguments in proving (2.37) that

$$I_1 + I_2 \xrightarrow{p} N(0, a^2 + b^2),$$

i.e., (2.30) holds.

To prove (2.31), we write

$$\begin{aligned} & \tfrac{m}{m_1^2 m_2^2 \rho} \textstyle\sum_{k=1}^{m_1} \left(\sum_{j=1}^{m_2} u_{kj}\right)^2 \\ = \ & \tfrac{m}{m_1^2 m_2^2 \rho} \textstyle\sum_{k=1}^{m_1} \left(\sum_{j=1}^{m_2} (X_k^T \bar{X}_k + Y_j^T \bar{Y}_j - Y_j^T \bar{X}_k - X_k^T \bar{Y}_j)\right)^2 \\ = \ & \tfrac{m}{m_1^2 \rho} \textstyle\sum_{k=1}^{m_1} \left(X_k^T \bar{X}_k + \frac{1}{m_2}\sum_{j=1}^{m_2} Y_j^T \bar{Y}_j - \frac{1}{m_2}\sum_{j=1}^{m_2} Y_j^T \bar{X}_k - X_k^T \frac{1}{m_2}\sum_{j=1}^{m_2} \bar{Y}_j\right)^2. \end{aligned} \tag{2.38}$$

Since $m\rho_1/m_1\rho \leq 1$, it follows from Lemma 2.2.5 that

$$\frac{m}{m_1^2\rho}\sum_{k=1}^{m_1}(X_k^T\bar{X}_k)^2 - \frac{m\rho_1}{m_1\rho} \xrightarrow{p} 0. \tag{2.39}$$

By Lemma 2.2.4, we have

$$\frac{m}{m_1^2\rho}\sum_{k=1}^{m_1}\left(\frac{1}{m_2}\sum_{j=1}^{m_2} Y_j^T\bar{Y}_j\right)^2 = O_p(\frac{m\rho_2}{m_1 m_2 \rho}) = o_p(1). \tag{2.40}$$

A direct calculation shows that

$$\begin{aligned}
&\mathbb{E}\{\tfrac{1}{m_2}\textstyle\sum_{j=1}^{m_2} Y_j^T\bar{X}_k\}^2\\
=\ &\mathbb{E}\{(\tfrac{1}{m_2}\textstyle\sum_{j=1}^{m_2} Y_j^T)\bar{X}_k\bar{X}_k^T(\tfrac{1}{m_2}\sum_{j=1}^{m_2} Y_j)\}\\
=\ &\mathbb{E}\mathrm{tr}\{\tfrac{1}{m_2}\textstyle\sum_{j=1}^{m_2} Y_j^T\bar{X}_k\bar{X}_k^T(\tfrac{1}{m_2}\sum_{j=1}^{m_2} Y_j)\}\\
=\ &\mathbb{E}\mathrm{tr}\{\bar{X}_k\bar{X}_k^T(\tfrac{1}{m_2}\textstyle\sum_{j=1}^{m_2} Y_j)(\tfrac{1}{m_2}\sum_{i=1}^{m_2} Y_i^T)\}\\
=\ &\mathrm{tr}\mathbb{E}\{\bar{X}_k\bar{X}_k^T(\tfrac{1}{m_2}\textstyle\sum_{j=1}^{m_2} Y_j)(\tfrac{1}{m_2}\sum_{i=1}^{m_2} Y_i^T)\}\\
=\ &\mathrm{tr}\{\Sigma\tfrac{1}{m_2}\bar{\Sigma}\}\\
=\ &O(\tfrac{\rho_1+\rho_2}{m_2})\\
=\ &O(\tfrac{m_1\rho}{m_2 m})+O(\tfrac{\rho_2}{m_2})\\
=\ &o(\tfrac{m_1\rho}{m}),
\end{aligned}$$

which implies that

$$\frac{m}{m_1^2\rho}\sum_{k=1}^{m_1}\{\frac{1}{m_2}\sum_{j=1}^{m_2} Y_j^T\bar{X}_k\}^2 = o_p(1). \tag{2.41}$$

Here tr means trace for a matrix. Similarly we have

$$\frac{m}{m_1^2\rho}\sum_{k=1}^{m_1}\{X_k^T\frac{1}{m_2}\sum_{j=1}^{m_2} \bar{Y}_j\}^2 = o_p(1). \tag{2.42}$$

It follows from (2.39) and (2.41) that

$$\begin{aligned}
&|\tfrac{m}{m_1^2\rho}\textstyle\sum_{k=1}^{m_1}(X_k^T\bar{X}_k)(\tfrac{1}{m_2}\sum_{j=1}^{m_2} Y_j^T\bar{X}_k)|\\
\le\ &\{\tfrac{m}{m_1^2\rho}\textstyle\sum_{k=1}^{m_1}(X_k^T\bar{X}_k)^2\}^{1/2}\{\tfrac{m}{m_1^2\rho}\sum_{k=1}^{m_1}(\tfrac{1}{m_2}\sum_{j=1}^{m_2} Y_j^T\bar{X}_k)^2\}^{1/2}\\
=\ &O_p(1)o_p(1)=o_p(1).
\end{aligned} \tag{2.43}$$

Similarly we can show that

$$\begin{cases}
\tfrac{m}{m_1^2\rho}\sum_{k=1}^{m_1}(X_k^T\bar{X}_k)(\tfrac{1}{m_2}\sum_{j=1}^{m_2} Y_j^T\bar{Y}_j)=o_p(1)\\
\tfrac{m}{m_1^2\rho}\sum_{k=1}^{m_1}(X_k^T\bar{X}_k)(X_k^T\tfrac{1}{m_2}\sum_{j=1}^{m_2} \bar{Y}_j^T)=o_p(1)\\
\tfrac{m}{m_1^2\rho}\sum_{k=1}^{m_1}(\tfrac{1}{m_2}\sum_{j=1}^{m_2} Y_j^T\bar{Y}_j)(\tfrac{1}{m_2}\sum_{i=1}^{m_2} Y_i^T\bar{X}_k)=o_p(1)\\
\tfrac{m}{m_1^2\rho}\sum_{k=1}^{m_1}(\tfrac{1}{m_2}\sum_{j=1}^{m_2} Y_j^T\bar{Y}_j)(X_k^T\tfrac{1}{m_2}\sum_{i=1}^{m_2} \bar{Y}_i)=o_p(1)\\
\tfrac{m}{m_1^2\rho}\sum_{k=1}^{m_1}(\tfrac{1}{m_2}\sum_{j=1}^{m_2} Y_j^T\bar{X}_k)(X_k^T\tfrac{1}{m_2}\sum_{i=1}^{m_2} Y_i)=o_p(1).
\end{cases} \tag{2.44}$$

Hence (2.31) follows from (2.38)–(2.44). The rest can be shown in the same way as proving (2.31). □

Lemma 2.2.7. *Under conditions of Theorem 1, we have as* $\min\{n_1, n_2\} \to \infty$

$$\frac{1}{\sqrt{m}} \sum_{k=1}^{m} \begin{pmatrix} \frac{Z_{k,1}}{\sqrt{\rho}} \\ \frac{Z_{k,2}}{\sqrt{\tau}} \end{pmatrix} \xrightarrow{d} N(0, I_2), \tag{2.45}$$

$$\frac{1}{m\rho} \sum_{k=1}^{m} Z_{k,1}^2 - 1 \xrightarrow{p} 0, \tag{2.46}$$

$$\frac{1}{m\tau} \sum_{k=1}^{m} Z_{k,2}^2 - 1 \xrightarrow{p} 0, \tag{2.47}$$

$$\frac{1}{m\sqrt{\rho\tau}} \sum_{k=1}^{m} Z_{k,1} Z_{k,2} \xrightarrow{p} 0. \tag{2.48}$$

Moreover, we have

$$\max_{1 \le k \le m} \left|\frac{Z_{k,1}}{\sqrt{\rho}}\right| = o_p(m^{1/2}) \quad \textit{and} \quad \max_{1 \le k \le m} \left|\frac{Z_{k,2}}{\sqrt{\tau}}\right| = o_p(m^{1/2}). \tag{2.49}$$

Proof. Note that for $1 \le k \le m_1$,

$$Z_{k,1} = \frac{-1}{(m_1 - 1)m_1} \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} u_{ij} + \frac{m_1 + m_2 - 1}{(m_1 - 1)m_2} \sum_{j=1}^{m_2} u_{kj},$$

$$Z_{k,2} = \frac{-1}{(m_1 - 1)m_1} \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} v_{ij} + \frac{m_1 + m_2 - 1}{(m_1 - 1)m_2} \sum_{j=1}^{m_2} v_{kj},$$

and for $m_1 + 1 \le k \le m$,

$$Z_{k,1} = \frac{-1}{(m_2 - 1)m_2} \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} u_{ij} + \frac{m_1 + m_2 - 1}{(m_2 - 1)m_1} \sum_{i=1}^{m_1} u_{i,k-m_1},$$

$$Z_{k,2} = \frac{-1}{(m_2 - 1)m_2} \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} u_{ij} + \frac{m_1 + m_2 - 1}{(m_2 - 1)m_1} \sum_{i=1}^{m_1} v_{i,k-m_1}.$$

Thus

$$\begin{aligned} \frac{1}{\sqrt{m}} \sum_{k=1}^{m} \frac{Z_{k,1}}{\sqrt{\rho}} &= \frac{1}{\sqrt{m}} \left(\frac{-1}{m_2 - 1} + \frac{-1}{m_1 - 1} + \frac{m_1 + m_2 - 1}{(m_1 - 1)m_2} + \frac{m_1 + m_2 - 1}{(m_2 - 1)m_1} \right) \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} \frac{u_{ij}}{\sqrt{\rho}} \\ &= \frac{\sqrt{m}}{m_1 m_2} \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} \frac{u_{ij}}{\sqrt{\rho}} \end{aligned}$$

and

$$\frac{1}{\sqrt{m}} \sum_{k=1}^{m} \frac{Z_{k,2}}{\sqrt{\tau}} = \frac{\sqrt{m}}{m_1 m_2} \sum_{j=1}^{m_2} \sum_{i=1}^{m_1} \frac{v_{ij}}{\sqrt{\tau}},$$

which imply (2.45) by using Lemma 2.2.6.

It follows from Lemma 2.2.6 that

$$
\begin{aligned}
& \frac{1}{m\rho}\sum_{k=1}^{m} Z_{k,1}^2 \\
= & \frac{1}{m\rho}\sum_{k=1}^{m_1}\left(\frac{-1}{(m_1-1)m_1}\sum_{j=1}^{m_2}\sum_{i=1}^{m_1} u_{ij} + \frac{m_1+m_2-1}{(m_1-1)m_2}\sum_{j=1}^{m_2} u_{kj}\right)^2 \\
& +\frac{1}{m\rho}\sum_{k=1}^{m_2}\left(\frac{-1}{(m_2-1)m_2}\sum_{j=1}^{m_2}\sum_{i=1}^{m_1} u_{ij} + \frac{m_1+m_2-1}{(m_2-1)m_1}\sum_{i=1}^{m_1} u_{ik}\right)^2 \\
= & \{\frac{1}{(m_1-1)\sqrt{m_1 m\rho}}\sum_{j=1}^{m_2}\sum_{i=1}^{m_1} u_{ij}\}^2 + \frac{(m-1)^2}{(m_1-1)^2 m_2^2 m\rho}\sum_{k=1}^{m_1}(\sum_{j=1}^{m_2} u_{kj})^2 \\
& -2\{(\frac{m-1}{m\rho(m_1-1)^2 m_1 m_2})^{1/2}\sum_{j=1}^{m_2}\sum_{i=1}^{m_1} u_{ij}\}^2 + \{\frac{1}{(m_2-1)\sqrt{m_2 m\rho}}\sum_{j=1}^{m_2}\sum_{i=1}^{m_1} u_{ij}\}^2 \\
& +\frac{(m-1)^2}{m\rho(m_2-1)^2 m_1^2}\sum_{k=1}^{m_2}(\sum_{i=1}^{m_1} u_{ik})^2 - 2\{(\frac{m-1}{m\rho(m_2-1)^2 m_2 m_1})^{1/2}\sum_{j=1}^{m_2}\sum_{i=1}^{m_1} u_{ij}\}^2 \\
= & \{O_p(\frac{1}{m_1\sqrt{m_1 m}}\frac{m_1 m_2}{\sqrt{m}})\}^2 + \frac{(m-1)^2 m_1^2}{(m_1-1)^2 m^2}\{\frac{m\rho_1}{m_1\rho} + o_p(1)\} + \{O_p(\frac{1}{m_1\sqrt{m_1 m_2}}\frac{m_1 m_2}{\sqrt{m}})\}^2 \\
& +\{O_p(\frac{1}{m_2\sqrt{m_2 m}}\frac{m_1 m_2}{\sqrt{m}})\}^2 + \frac{(m-1)^2 m_2^2}{m^2(m_2-1)^2}\{\frac{m\rho_2}{m_2\rho} + o_p(1)\} + \{O_p(\frac{1}{m_2\sqrt{m_2 m_1}}\frac{m_1 m_2}{\sqrt{m}})\}^2 \\
= & \frac{m\rho_1}{m_1\rho} + \frac{m\rho_2}{m_2\rho} + o_p(1) \\
= & 1 + o_p(1),
\end{aligned}
$$

i.e., (2.46) holds. Similarly we can show (2.47) and (2.48).

Since $\mathrm{Var}(\sum_{i=1}^{m_1} u_{ij}) = m_1(\rho_1+\rho_2)$, we have

$$\lim_{x\to\infty} xP(|\sum_{i=1}^{m_1} u_{ij}| > \sqrt{x m_1(\rho_1+\rho_2)}) = 0,$$

which implies that

$$\max_{1\le j\le m_2} |\sum_{i=1}^{m_1} u_{ij}| = o_p(\sqrt{m_2 m_1(\rho_1+\rho_2)}).$$

Similarly we have

$$\max_{1\le 1\le m_1} |\sum_{j=1}^{m_2} u_{ij}| = o_p(\sqrt{m_2 m_1(\rho_1+\rho_2)}).$$

Hence by Lemma 2.2.6 and the expression for $Z_{k,1}$, we have

$$\begin{aligned}
\max_{1\le k\le m}|\tfrac{Z_{k,1}}{\sqrt{\rho}}| &\le \tfrac{1}{(m_1-1)m_1}|\textstyle\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\tfrac{u_{ij}}{\sqrt{\rho}}| + \max_{1\le k\le m_1}|\tfrac{m-1}{(m_1-1)m_2}\textstyle\sum_{j=1}^{m_2}\tfrac{u_{kj}}{\sqrt{\rho}}| \\
&\quad + \tfrac{1}{(m_2-1)m_2}|\textstyle\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}\tfrac{u_{ij}}{\sqrt{\rho}}| + \max_{1\le k\le m_2}|\tfrac{m-1}{(m_2-1)m_1}\textstyle\sum_{j=1}^{m_1}\tfrac{u_{jk}}{\sqrt{\rho}}| \\
&= o_p(1) + o_p(\tfrac{m-1}{(m_1-1)m_2\sqrt{\rho}}\{m_1m_2(\rho_1+\rho_2)\}^{1/2}) \\
&\quad + o_p(1) + o_p(\tfrac{m-1}{(m_2-1)m_1\sqrt{\rho}}\{m_1m_2(\rho_1+\rho_2)\}^{1/2}) \\
&= o_p(m^{1/2}).
\end{aligned}$$

Similarly we can show that

$$\max_{1\le k\le m}|\frac{Z_{k,2}}{\sqrt{\tau}}| = o_p(m^{1/2}).$$

□

Proof of Theorem 2.2.1. It follows from Lemma 2.2.7 and the standard arguments in empirical likelihood method (see Owen [72]). □

To show Corollary 2.2.2 and Theorem 2.2.3, we first prove the following lemmas.

Lemma 2.2.8. $\mathrm{tr}(\Sigma^4) = O\big((\mathrm{tr}(\Sigma^2))^2\big)$, $\rho_1 = \sum_{j=1}^{p}\lambda_j^2$, *and* $2p\lambda_1 \le \tau_1 \le 2p\lambda_p$.

Proof. Since $\mathrm{tr}(\Sigma^j) = \sum_{i=1}^{p}\lambda_i^j$ for any positive integer j, the first equality follows immediately. The second equality follows since $\rho_1 = \mathrm{tr}(\Sigma^2)$. The third inequalities on τ_1 are obvious. □

Lemma 2.2.9. *For any* $\delta > 0$

$$\mathbb{E}|X_1^T\bar{X}_1|^{2+\delta} \le p^\delta\left(\sum_{i=1}^{p}\mathbb{E}|X_{1,i}|^{2+\delta}\right)^2$$

and

$$\mathbb{E}|\mathbf{1}_p^T(X_1+\bar{X}_1)|^{2+\delta} \le 2^{4+\delta}p^{1+\delta}\sum_{i=1}^{p}\mathbb{E}|X_{1,i}|^{2+\delta}.$$

Proof. It follows from the Cauchy-Schwarz inequality that

$$|X_1^T\bar{X}_1|^2 \le ||X_1||^2||\bar{X}_1||^2.$$

Then by using the C_r inequality we conclude that

$$\begin{aligned}\mathbb{E}|X_1^T\bar{X}_1|^{2+\delta} &\leq \mathbb{E}\left(\sum_{i=1}^p X_{1,i}^2\right)^{(2+\delta)/2} \mathbb{E}\left(\sum_{i=1}^p \bar{X}_{1,i}^2\right)^{(2+\delta)/2} \\ &= \left(\mathbb{E}\left(\sum_{i=1}^p X_{1,i}^2\right)^{(2+\delta)/2}\right)^2 \\ &\leq \left(p^{\delta/2}\sum_{i=1}^p \mathbb{E}|X_{1,i}|^{2+\delta}\right)^2 \\ &= p^{\delta}\left(\sum_{i=1}^p \mathbb{E}|X_{1,i}|^{2+\delta}\right)^2.\end{aligned}$$

Similarly, from the C_r inequality we have

$$\mathbb{E}|\mathbf{1}_p^T(X_1+\bar{X}_1)|^{2+\delta} \leq 2^{4+\delta}\mathbb{E}\left(\sum_{i=1}^p |X_{1,i}|\right)^{2+\delta} \leq 2^{4+\delta}p^{1+\delta}\sum_{i=1}^p \mathbb{E}|X_{1,i}|^{2+\delta}.$$

This completes the proof. □

Proof of Corollary 2.2.2. Equations (2.11) and (2.13) follow from conditions (A1)–(A3) by using Lemmas 2.2.8 and 2.2.9. So do equations (2.12) and (2.14), since we have the same assumptions on $\{X_i\}$ and $\{Y_j\}$. □

Proof of Theorem 2.2.3. If suffices to verify conditions (2.11) and (2.13) with $\delta = 2$ in Theorem 2.2.1. Recall we assume that $\mu_1 = \mu_2 = 0$. Note that $\mathrm{Var}(X_1) = \Sigma = \Gamma_1\Gamma_1^T$. Denote $\mathbf{1}_p^T\Gamma_1 = (a_1, \cdots, a_k)$ and $\Sigma' = \Gamma_1^T\Gamma_1 = (\sigma'_{j,l})_{1\leq j,l\leq k}$. Then

$$X_1^T\bar{X}_1 = \sum_{j=1}^k\sum_{l=1}^k \sigma'_{j,l}B_{1,j}B_{1+m_1,l},$$

and

$$\mathbf{1}_p^T(X_1+\bar{X}_1) = \sum_{j=1}^k a_j(B_{1,j}+B_{1+m_1,j}).$$

Set $\delta_{j_1,j_2,j_3,j_4} = \mathbb{E}(B_{1,j_1}B_{1,j_2}B_{1,j_3}B_{1,j_4})$. Then δ_{j_1,j_2,j_3,j_4} equals $3+\xi_1$ if $j_1 = j_2 = j_3 = j_4$, equals 1 if j_1, j_2, j_3 and j_4 form two different pairs of integers, and is zero otherwise. By Lemma 2.2.8, we have

$$
\begin{aligned}
\mathbb{E}(X_1^T\bar{X}_1)^4 &= \sum_{j_1,j_2,j_3,j_4=1}^{k}\sum_{l_1,l_2,l_3,l_4=1}^{k}\sigma'_{j_1,l_1}\sigma'_{j_2,l_2}\sigma'_{j_3,l_3}\sigma'_{j_4,l_4}\delta_{j_1,j_2,j_3,j_4}\delta_{l_1,l_2,l_3,l_4}\\
&= O\left(|\textstyle\sum_{j_1\neq j_2}\sum_{l_1\neq l_2}\sigma'_{j_1,l_1}\sigma'_{j_1,l_2}\sigma'_{j_2,l_1}\sigma'_{j_2,l_2}|\right)+O\left(\textstyle\sum_{j_1\neq j_2}\sum_{l=1}^{k}\sigma'^2_{j_1,l}\sigma'^2_{j_2,l}\right)\\
&\quad +O\left(\textstyle\sum_{j=1}^{k}\sum_{l_1\neq l_2}\sigma'^2_{j,l_1}\sigma'^2_{j,l_2}\right)+O\left(\textstyle\sum_{j=1}^{k}\sum_{l=1}^{k}\sigma'^4_{j,l}\right)\\
&= O\left(|\textstyle\sum_{j_1=1}^{k}\sum_{j_2=1}^{k}\sum_{l_1=1}^{k}\sum_{l_2=1}^{k}\sigma'_{j_1,l_1}\sigma'_{j_1,l_2}\sigma'_{j_2,l_1}\sigma'_{j_2,l_2}|\right)\\
&\quad +O\left(\textstyle\sum_{j_1=1}^{k}\sum_{j_2=1}^{k}\sum_{l=1}^{k}\sigma'^2_{j_1,l}\sigma'^2_{j_2,l}\right)+O\left(\textstyle\sum_{j=1}^{k}\sum_{l=1}^{k}\sigma'^4_{j,l}\right)\\
&= O\left(\mathbf{tr}(\Sigma'^4)\right)+O\left(\left(\textstyle\sum_{j=1}^{k}\sum_{l=1}^{k}\sigma'^2_{j,l}\right)^2\right)\\
&= O\left(\mathbf{tr}(\Sigma'^4)\right)+O\left((\mathbf{tr}(\Sigma'^2))^2\right)\\
&= O\left(\mathbf{tr}(\Sigma^4)\right)+O\left((\mathbf{tr}(\Sigma^2))^2\right)\\
&= o\big(m(\mathbf{tr}(\Sigma^2))^2\big),
\end{aligned}
$$

i.e., (2.11) holds with $\delta = 2$.

Similarly we have

$$
\begin{aligned}
\mathbb{E}(\mathbf{1}_p^T(X_1+\bar{X}_1))^4 &\leq 2^4\mathbb{E}\Big(\sum_{j=1}^{k}a_jB_{1,j}\Big)^4\\
&= O\left(\sum_{j_1,j_2=1}^{k}a_{j_1}^2a_{j_2}^2\right)+O\left(\sum_{j=1}^{k}a_j^4\right)\\
&= O\left(\Big(\sum_{j=1}^{k}a_j^2\Big)^2\right)\\
&= O\left(\Big(\mathbf{1}_p^T\Gamma_1\Gamma_1^T\mathbf{1}_p\Big)^2\right)\\
&= O\left(\Big(\sum_{i=1}^{p}\sum_{j=1}^{p}\sigma_{i,j}\Big)^2\right),
\end{aligned}
$$

which yields (2.13) with $\delta = 2$. Equations (2.12) and (2.14) can be shown in the same way. Hence Theorem 2.2.3 follows from Theorem 2.2.1. □

2.3 *Test for High-dimensional Linear Models*

Linear model is a common technique to fit the relationship between responses and covariates. Statistical inference can be based on either the least squares estimator or

M-estimator for the coefficients. However, the asymptotic behavior generally depends on whether the number of covariates is fixed or goes to infinity as the sample size tends to infinity. In this section, we propose an empirical likelihood method for testing whether the coefficients are equal to the given values. The asymptotic distribution of the proposed test is independent of the number of covariates in the linear model. A simulation study shows that the proposed test performs well in terms of both size and power.

2.3.1 Introduction

In order to model the relationship between responses and covariates, regression model is a commonly employed technique. Consider the following classic linear regression model

$$Y_i = \beta^T X_i + \epsilon_i, \quad i = 1, \cdots, n, \tag{2.50}$$

where $\beta = (\beta_1, \cdots, \beta_p)^T$ is the vector of unknown parameters, $X_1 = (X_{1,1}, \cdots, X_{1,p})^T$, $\cdots$, $X_n = (X_{n,1}, \cdots, X_{n,p})^T$ are i.i.d random vectors, $\epsilon_1, \cdots, \epsilon_n$ are independent and identically distributed random variables with zero mean and variance σ^2, and $X_i's$ and $\epsilon_i's$ are independent. Statistical inference for β can be based on either least squares estimator or M-estimator when p is fixed. When p depends on the sample size n and goes to infinity as $n \to \infty$, Portnoy [79, 80] studied the consistency and asymptotic normality of M-estimators for β, which requires that p can not be too large.

Motivated by the studies in bioinformatics and other fields, statistical inference for the linear model (2.50) is needed for the case when p is of an exponential order of n, but many of $\beta_i's$ are zero. To deal with this case, one first selects variables with nonzero $\beta_i's$ and then makes statistical inference for the selected nonzero $\beta_i's$. It is not surprising that the order of the number of nonzero $\beta_i's$ can not be larger than the optimal one in Portnoy [80]. We refer to Bradic, Fan and Wang [10] for more details and references on the ultrahigh dimensional situation. In this section we are interested in testing $H_0 : \beta = \beta_0$ against $\beta \neq \beta_0$ for a given value $\beta_0 \in \mathbb{R}^p$ when p is either fixed or goes to infinity as $n \to \infty$.

When p is fixed, a traditional test is the Hotelling's T^2 test defined as

$$HT = \frac{1}{\hat{\sigma}^2}(\hat{\beta} - \beta_0)^T \left(\frac{1}{n} \sum_{i=1}^{n} X_i X_i^T \right)^{-1} (\hat{\beta} - \beta_0), \tag{2.51}$$

where $\hat{\beta} = (\frac{1}{n} \sum_{i=1}^{n} X_i X_i^T)^{-1} \frac{1}{n} \sum_{i=1}^{n} Y_i X_i$ and $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\beta}^T X_i)^2$. It is known that $HT \xrightarrow{d} \chi_p^2$ as $n \to \infty$. However, when p is large, finding the inverse matrix in (2.51) becomes problematic.

As a powerful nonparametric likelihood approach, empirical likelihood test is another useful method. More specifically, write $z_i = X_i(Y_i - X_i^T \beta)$ for $i = 1, \cdots, n$ and define the empirical likelihood function for β as

$$L_{n1}(\beta) = \sup\{\prod_{i=1}^{n} (nq_i) : q_1 \geq 0, \cdots, q_n \geq 0, \sum_{i=1}^{n} q_i = 1, \sum_{i=1}^{n} q_i z_i = 0\}.$$

Under some regularity conditions, one can show that the Wilks' Theorem holds, i.e., $-2 \log L_{n1}(\beta_0)$ converges in distribution to a chi-square limit with p degrees of freedom. Therefore, the empirical likelihood test can be constructed by using the test statistic $-2 \log L_{n1}(\beta)$. See Owen [73] for more details on empirical likelihood methods. However, the maximization in computing $L_{n1}(\beta)$ becomes nontrivial and even unavailable when p is large; see Chen, Variyath, and Abraham [20] for discussions on this phenomena. Empirical likelihood method for high dimensional data can be found in Chen, Peng and Qin [15] and Hjort, McKeague and Van Keilegom [45].

Considering the difficulties in the above methods, in this section we propose a new empirical likelihood test for testing $H_0 : \beta = \beta_0$ by splitting the data into two parts. It turns out the new method works for both fixed and divergent p.

We organize the whole section as follows. Section 2.3.2 presents the new methodology and main results. A simulation study is given in Section 2.3.3. All proofs are put in Section 2.3.4.

2.3.2 Methodology

Put $m = [n/2]$, the integer part of $n/2$, and define $\tilde{X}_i = X_{m+i}$, $\tilde{Y}_i = Y_{i+m}$, $\tilde{\epsilon}_i = \epsilon_{i+m}$,

$$W_i(\beta) = (Y_i X_i - X_i X_i^T \beta)^T (\tilde{Y}_i \tilde{X}_i - \tilde{X}_i \tilde{X}_i^T \beta)$$

for $i = 1, \cdots, m$. Then

$$\mathbb{E}W_i(\beta) = \mathbb{E}\{(X_i X_i^T(\beta_0 - \beta) + X_i\epsilon_i)^T(\tilde{X}_i\tilde{X}_i^T(\beta_0 - \beta) + \tilde{X}_i\tilde{\epsilon}_i)\} = (\beta_0 - \beta)^T \Sigma^2 (\beta_0 - \beta),$$

where $\Sigma = \mathbb{E}(X_1 X_1^T)$. When Σ is positive definite, testing $H_0 : \beta = \beta_0$ against $H_a : \beta \neq \beta_0$ is equivalent to testing $H_0 : \mathbb{E}W_1(\beta) = 0$ against $H_a : \mathbb{E}W_1(\beta) \neq 0$. This motivates us to apply the empirical likelihood method in Qin and Lawless [82] to the estimating equation $\mathbb{E}W_1(\beta_0) = 0$. However this direct application results in a poor power in general by noting that $\mathbb{E}W_1(\beta) = O(||\beta - \beta_0||^2)$ instead of $O(||\beta - \beta_0||)$ when $||\beta - \beta_0||$ is small, where $||\cdot||$ denotes the L_2 norm of a vector.

To improve the power, we propose to add one more linear equation $\mathbb{E}W_1^*(\beta_0) = 0$ where $\mathbb{E}W_1^*(\beta) = O(||\beta - \beta_0||)$ and thus it catches the small change of $\beta - \beta_0$. More specifically, define

$$W_i^*(\beta) = (Y_i X_i - X_i X_i^T \beta)^T \mathbf{1}_p + (\tilde{Y}_i \tilde{X}_i - \tilde{X}_i \tilde{X}_i^T \beta)^T \mathbf{1}_p$$

for $i = 1, \cdots, m$, where $\mathbf{1}_p = (1, 1, \cdots, 1)^T \in \mathbb{R}^p$, and then define the empirical likelihood function for β as

$$L_{n2}(\beta) = \sup\{\prod_{i=1}^m (mq_i) : q_1 \geq 0, \cdots, q_m \geq 0, \sum_{i=1}^m q_i = 1, \sum_{i=1}^m q_i W_i(\beta) = 0, \sum_{i=1}^m q_i W_i^*(\beta) = 0\}.$$

By the Lagrange multiplier technique, we have

$$-2 \log L_{n2}(\beta) = 2 \sum_{i=1}^m \log\{1 + b_1 W_i(\beta) + b_2 W_i^*(\beta)\}, \tag{2.52}$$

where $b_1 = b_1(\beta)$ and $b_2 = b_2(\beta)$ satisfy that

$$\begin{cases} \sum_{i=1}^m \frac{W_i(\beta)}{1+b_1 W_i(\beta)+b_2 W_i^*(\beta)} = 0, \\ \sum_{i=1}^m \frac{W_i^*(\beta)}{1+b_1 W_i(\beta)+b_2 W_i^*(\beta)} = 0. \end{cases} \tag{2.53}$$

The following theorem shows that the Wilks' Theorem holds for the above empirical likelihood method. As in Section 2.2, we use $\mathrm{tr}(A)$ to denote the trace of a matrix A.

Theorem 2.3.1. *Let β_0 be the true value of the parameter β. Assume Σ is positively definite and and there exists some $\delta > 0$ such that*

$$\frac{\mathbb{E}|X_1^T \tilde{X}_1|^{2+\delta}}{\{\mathrm{tr}(\Sigma^2)\}^{(2+\delta)/2}} \left(\frac{\mathbb{E}|\epsilon_1|^{2+\delta}}{\sigma^{2+\delta}} \right)^2 = o(m^{\frac{\delta+\min(\delta,2)}{4}}), \tag{2.54}$$

and

$$\frac{\mathbb{E}|X_1^T \mathbf{1}_p|^{2+\delta}}{\{\mathbb{E}(X_1^T \mathbf{1}_p)^2\}^{(2+\delta)/2}} \left(\frac{\mathbb{E}|\epsilon_1|^{2+\delta}}{\sigma^{2+\delta}} \right) = o(m^{\frac{\delta+\min(\delta,2)}{4}}), \tag{2.55}$$

where $\sigma^2 = \mathrm{Var}(\epsilon_1)$. Then $-2 \log L_{n2}(\beta_0)$ converges in distribution to a chi-square limit with 2 degrees of freedom.

Remark 2.3.1. The conditions (2.54) and (2.55) can be rephrased as $X_1^T \tilde{X}_1 \epsilon \tilde{\epsilon}$ and $X_1^T \mathbf{1}_p \epsilon$ satisfy condition (P).

Remark 2.3.2. The distributions of X_1 varies with n as the dimension of X_1 changes with n. In general, the distribution of the error term ϵ_1 may also change with n and thus the moments of ϵ_1 may not be constants.

Remark 2.3.3. Theorem 2.3.1 deals with large p since the high-dimensional model is of our interest. When p is small and fixed, the traditional empirical likelihood test L_{n1} defined in the introduction may perform better since the sample size in our test is $n/2$ instead of n.

Remark 2.3.4. In Theorem 2.3.1, the condition that Σ is positively definite simply requires the random variable X_1 not to be degenerate. Conditions (2.54) and (2.55) may impose some restriction on p implicitly.

In the following we will give two examples where little restriction on p is required.

Example 2.3.1. Let X_1 be a Gaussian random vector with mean 0 and covariance matrix $\Sigma = (\sigma_{i,j})_{1 \le i,j \le p}$, where Σ is an arbitrary p by p positively definite matrix. Assume $\mathbb{E}(\epsilon_1^4)/\sigma^4 = o(m^{1/2})$, then conditions (2.54) and (2.55) hold.

Example 2.3.2. Assume $(X_1, \cdots, X_n)$ and $(\epsilon_1, \cdots, \epsilon_n)$ satisfy (A), then conditions (2.54) and (2.55) hold.

The proof of the examples will be put in Section 2.3.4.

Remark 2.3.5. Example 2.3.1 assumes a special dependence structure and it is a special case of (B). Similar to the test in Chapter 2.2, condition (B) is sufficient for Theorem 2.3.1.

Remark 2.3.6. One advantage of the proposed empirical likelihood method is that one can easily add more equations if one has more information on the alternative hypothesis, or replace $W_1^*(\beta)$ by another statistic $\bar{W}_1(\beta)$ satisfying $\mathbb{E}\bar{W}_1(\beta) = O(||\beta - \beta_0||)$. Although adding more relevant equations may improve the test power, computing the empirical likelihood function becomes more complicated. The simulation study in the next section shows that the test using $\mathbb{E}W_i(\beta) = 0$ and $\mathbb{E}W_i^*(\beta) = 0$ in Theorem 2.3.1 performs well in terms of both size and power in the dense model.

2.3.3 Simulation study

In this section, we examine the finite sample behavior of the proposed empirical likelihood test and compare it with the Hotelling's T^2 test and the standard empirical likelihood method in terms of both size and power.

Draw 10,000 random samples with size $n = 200, 500$ from the linear model (2.50) with $X_i = (X_{i1}, \cdots, X_{ip})^T \sim N(0, \Sigma_0)$, $\Sigma_0 = (0.5^{(|i-j|)})_{1\le i,j\le p}$, $\epsilon_i \sim t_8$ and $\beta = \beta_0 + \delta/\sqrt{n}$, $\beta_0 = \mathbf{1}_p$. Consider testing $H_0 : \beta = \beta_0$ against $H_a : \beta \neq \beta_0$. We use EL1, EL2 and HT to denote the empirical likelihood tests $-2\log L_{n1}(\beta)$, $-2\log L_{n2}(\beta)$ and the Hotelling's T^2 test in (2.51), respectively. We compute the powers of these three tests and plot them against different p at levels 0.1 and 0.05 in Figures 2.1–2.4. Note that $\delta = 0$ corresponds to the size of the tests.

From the first plot of each figure we find that the traditional empirical likelihood method and the Hotelling's T^2 test do not have a consistent size when p is slightly large, while the proposed empirical likelihood test has a very stable size with respect to p. The other three plots in each figure show that the proposed empirical likelihood method is powerful too. Note that the power for the traditional empirical likelihood method and the Hotelling's T^2 test do not make much sense for a slightly large p since their sizes do not converge to the nominal levels. When n becomes large, the proposed empirical likelihood tests have more accurate size.

In summary, the proposed empirical likelihood test has a very stable size with respect to the number of covariates and are powerful too. The proposed new tests are easy to implement by using the R package `emplik`, which does not need to compute the inverse of a high dimensional covariance matrix.

2.3.4 Proofs

Throughout we denote

$$u_i := W_i(\beta_0) = (X_i^T \tilde{X}_i)\epsilon_i\tilde{\epsilon}_i, \qquad v_i := W_i^*(\beta_0) = (X_i^T \mathbf{1}_p)\epsilon_i + (\tilde{X}_i^T \mathbf{1}_p)\tilde{\epsilon}_i,$$

$$\sigma_1 = \sqrt{\mathrm{Var}(u_1)} \quad \text{and} \quad \sigma_2 = \sqrt{\mathrm{Var}(v_1)}.$$

Then it is easy to verify that $\mathbb{E}(u_1) = \mathbb{E}(v_1) = \mathbb{E}(u_1 v_1) = 0$. One can also easily show that conditions (2.54) and (2.55) are respectively equivalent to

$$\frac{\mathbb{E}|u_1|^{2+\delta}}{\sigma_1^{2+\delta}} = o(m^{\frac{\delta+\min(\delta,2)}{4}}), \tag{2.56}$$

and

$$\frac{\mathbb{E}|v_1|^{2+\delta}}{\sigma_2^{2+\delta}} = o(m^{\frac{\delta+\min(\delta,2)}{4}}). \tag{2.57}$$

Lemma 2.3.2. *Under conditions of Theorem 2.3.1, we have*

$$\frac{1}{\sqrt{m}} \sum_{i=1}^{m} \begin{pmatrix} \frac{u_i}{\sigma_1} \\ \frac{v_i}{\sigma_2} \end{pmatrix} \xrightarrow{d} N(0, I_2), \tag{2.58}$$

$$\frac{\sum_{i=1}^m u_i^2}{m\sigma_1^2} - 1 \xrightarrow{p} 0, \tag{2.59}$$

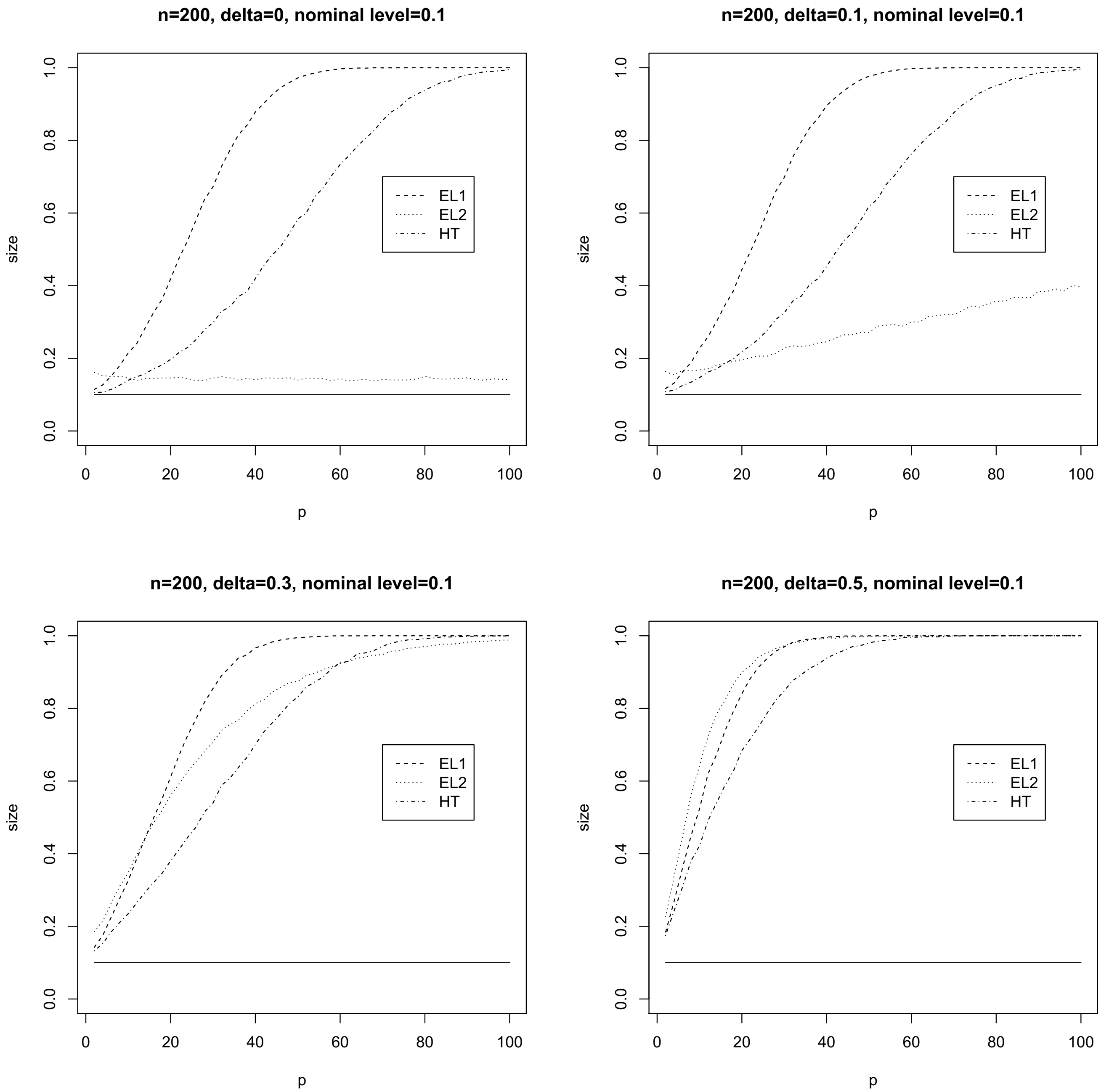

Figure 2.1: Powers of tests are plotted against $p = 2, 4, \cdots, 100$ with level 0.01 and $n = 200$.

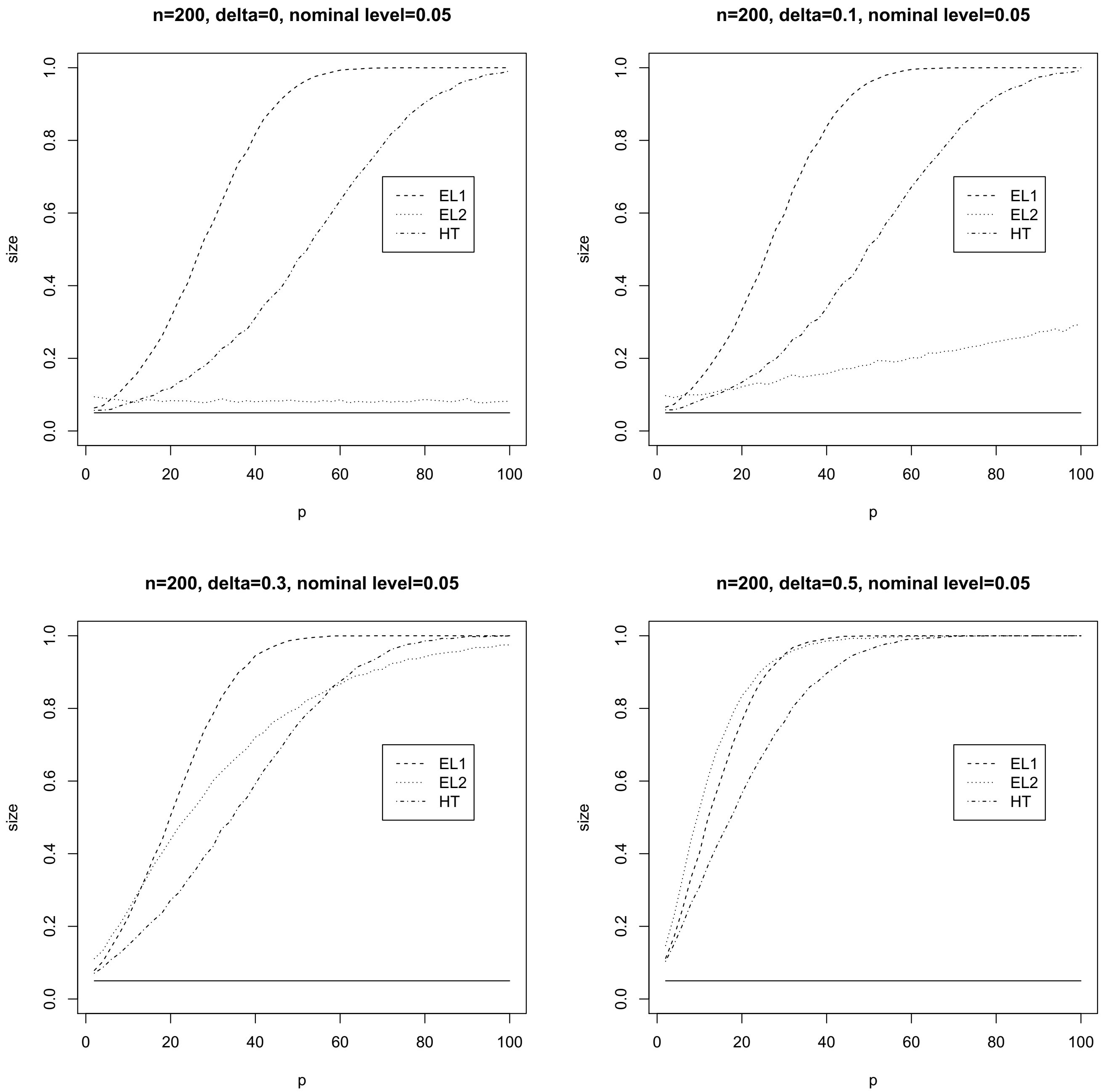

Figure 2.2: Powers of tests are plotted against $p = 2, 4, \cdots, 100$ with level 0.05 and $n = 200$.

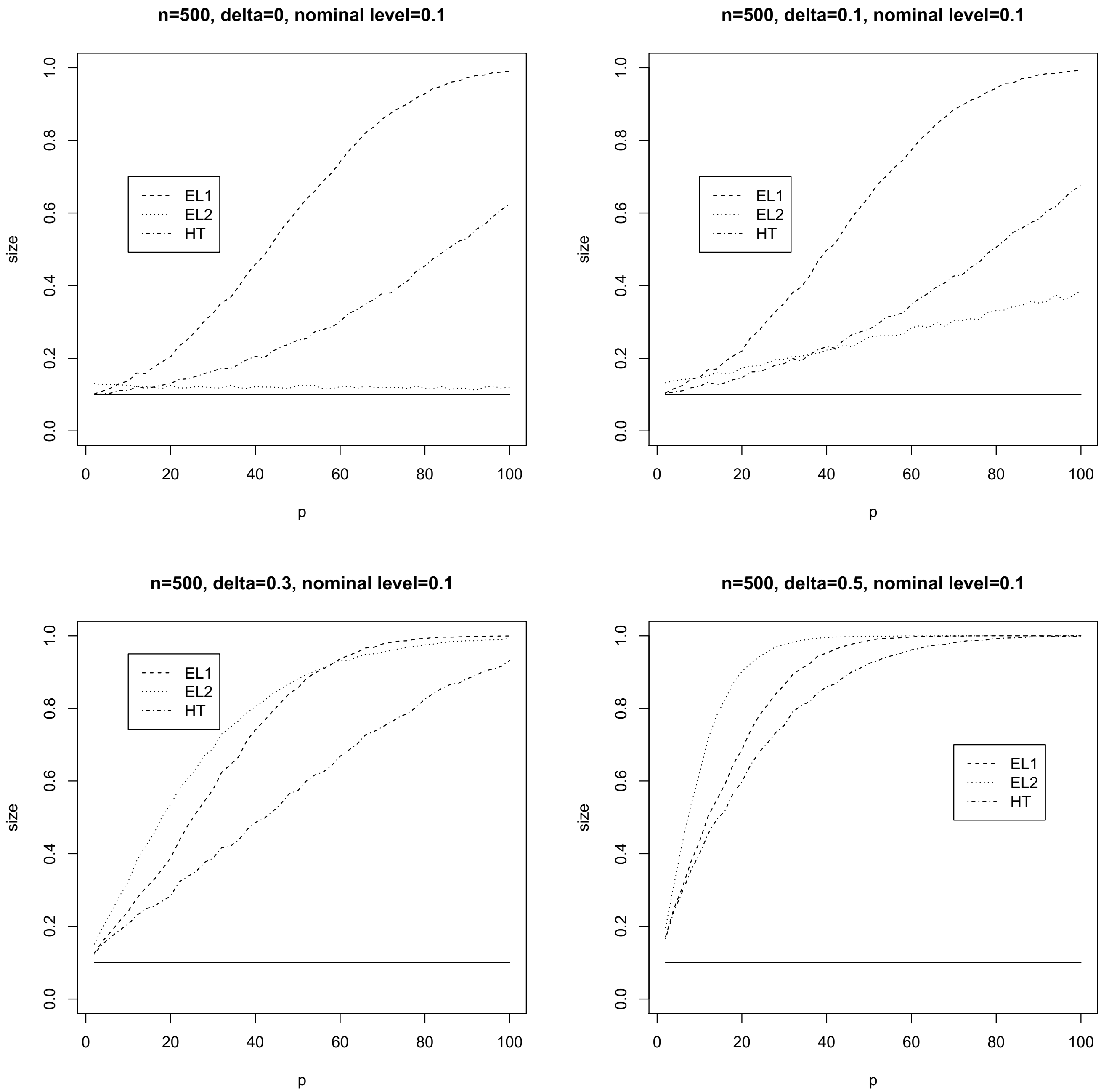

Figure 2.3: Powers of tests are plotted against $p = 2, 4, \cdots, 100$ with level 0.01 and $n = 500$.

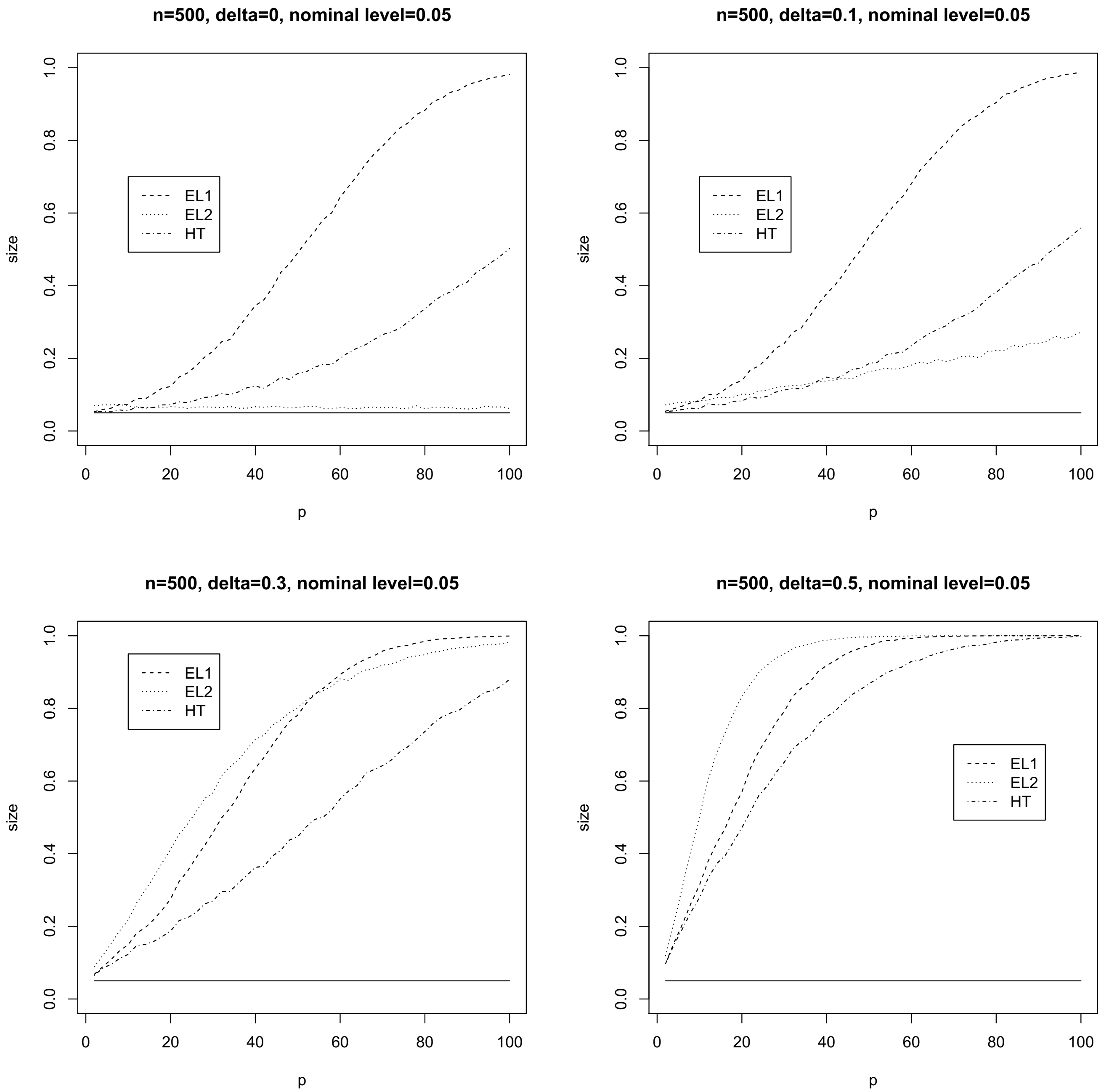

Figure 2.4: Powers of tests are plotted against $p = 2, 4, \cdots, 100$ with level 0.05 and $n = 500$.

$$\frac{\sum_{i=1}^m v_i^2}{m\sigma_2^2} - 1 \xrightarrow{p} 0, \tag{2.60}$$

$$\frac{\sum_{i=1}^m u_i v_i}{m\sigma_1\sigma_2} \xrightarrow{p} 0, \tag{2.61}$$

$$\max_{1\le i\le m} |\frac{u_i}{\sigma_1}| = o_p(m^{1/2}) \quad \textit{and} \quad \max_{1\le i\le m} |\frac{v_i}{\sigma_2}| = o_p(m^{1/2}), \tag{2.62}$$

where I_2 is a 2×2 identity matrix.

Proof. Note that u_1 and v_1 are uncorrelated. To show (2.58) we need to prove that for any constants a and b with $a^2 + b^2 \neq 0$,

$$\frac{1}{\sqrt{m}} \sum_{i=1}^m (a\frac{u_i}{\sigma_1} + b\frac{v_i}{\sigma_2}) \xrightarrow{d} N(0, a^2 + b^2).$$

Apparently $\{a\frac{u_i}{\sigma_1} + b\frac{v_i}{\sigma_2}, 1 \le i \le m\}$ are independent and identically distributed random variables with variance $a^2 + b^2$. Therefore we shall verify the Lindeberg condition for the triangular array $\{a\frac{u_i}{\sigma_1} + b\frac{v_i}{\sigma_2}, 1 \le i \le m\}$. It suffices to show the Lyapunov condition

$$\frac{1}{m^{(2+\delta)/2}} \sum_{i=1}^m \mathbb{E}|a\frac{u_i}{\sigma_1} + b\frac{v_i}{\sigma_2}|^{2+\delta} \to 0 \tag{2.63}$$

holds. This follows from the fact that the left-hand side of (2.63) is dominated by

$$\begin{aligned} \frac{m\mathbb{E}|a\frac{u_i}{\sigma_1} + b\frac{v_i}{\sigma_2}|^{2+\delta}}{m^{(2+\delta)/2}} &\le (|a| + |b|)^{2+\delta} \left(\mathbb{E}|\frac{u_1}{\sigma_1}|^{2+\delta} + \mathbb{E}|\frac{v_1}{\sigma_2}|^{2+\delta} \right) \frac{1}{m^{\delta/2}} \\ &= o(m^{\frac{\delta+\min(\delta,2)}{4} - \frac{\delta}{2}}) \\ &= o(1). \end{aligned}$$

To show (2.59), we need to estimate $\mathbb{E}|\sum_{i=1}^m u_i^2 - m\sigma_1^2|^{(2+\delta)/2}$. We have from von Bahr and Esseen [98] that

$$\mathbb{E}|\sum_{i=1}^m u_i^2 - m\sigma_1^2|^{(2+\delta)/2} \le 2m\mathbb{E}|u_1^2 - \mathbb{E}(u_1^2)|^{(2+\delta/2)} = O(m\mathbb{E}|u_1|^{2+\delta}) \tag{2.64}$$

if $0 < \delta \le 2$, and from Dharmadhikari and Jogdeo [28] that

$$\mathbb{E}|\sum_{i=1}^m u_i^2 - m\sigma_1^2|^{(2+\delta)/2} \le Cm^{(2+\delta)/4}\mathbb{E}|u_1^2 - \mathbb{E}(u_1^2)|^{(2+\delta/2)} = O(m^{(2+\delta)/4}\mathbb{E}|u_1|^{2+\delta}) \tag{2.65}$$

if $\delta > 2$. Therefore, by (2.64), (2.65) and (2.56), we have for any $\varepsilon > 0$

$$\begin{aligned} & \mathbb{P}(|\frac{\sum_{i=1}^m u_i^2}{m\sigma_1^2} - 1| > \varepsilon) \\ \leq\ & \varepsilon^{-(2+\delta)/2} \frac{\mathbb{E}|\sum_{i=1}^m u_i^2 - m\sigma_1^2|^{(2+\delta)/2}}{(m\sigma_1^2)^{(2+\delta)/2}} \\ =\ & O(m^{-(\delta+\min(\delta,2))/4}\mathbb{E}|\frac{u_1}{\sigma_1}|^{2+\delta}) \\ =\ & o(1), \end{aligned}$$

which implies (2.59). Similarly we can show (2.60) and (2.61). Equation (2.62) follows from the Lyapunov condition (2.63) by letting $a = 1$ and $b = 0$ or $a = 0$ and $b = 1$. This completes the proof of the lemma. □

Proof of Theorem 2.3.1. Set $Z_i = (u_i/\sigma_1, v_i/\sigma_2)^T$ for $i = 1\ldots, m$. It follows from Lemma 2.3.2 that

$$\frac{1}{\sqrt{m}} \sum_{i=1}^m Z_i \xrightarrow{d} N(0, I_2), \tag{2.66}$$

$$||\frac{1}{m} \sum_{i=1}^m Z_i(Z_i)^T - I_2|| \xrightarrow{p} 0, \tag{2.67}$$

$$\max_{1\leq i\leq m} ||Z_i|| = o_p(m^{1/2}). \tag{2.68}$$

Put $\rho = (\rho_1, \rho_2)^T = (b_1\sigma_1, b_2\sigma_2)^T$ with b_1 and b_2 being given in (2.52) and (3.34). Then we have

$$-2\log L_{n2}(\beta_0) = 2\sum_{i=1}^m \log(1 + \rho^T Z_i),$$

where ρ solves

$$\sum_{i=1}^m \frac{Z_i}{1 + \rho^T Z_i} = 0. \tag{2.69}$$

Similar to the proof of (2.14) in Owen [72] we can show $\sqrt{\rho_1^2 + \rho_2^2} = O_p(m^{-1/2})$. Then it follows from (2.68) that

$$\max_{1\leq i\leq m} ||\frac{\rho^T Z_i}{1 + \rho^T Z_i}|| = o_p(1).$$

By (2.69), we have

$$\begin{aligned}
0 &= \frac{1}{m}\sum_{i=1}^{m}\frac{\rho^T Z_i}{1+\rho^T Z_i} \\
&= \frac{1}{m}\sum_{i=1}^{m}\rho^T Z_i(1-\rho Z_i+\frac{(\rho^T Z_i)^2}{1+\rho^T Z_i}) \\
&= \frac{1}{m}\sum_{i=1}^{m}\rho^T Z_i-\frac{1}{m}\sum_{i=1}^{m}(\rho^T Z_i)^2+\frac{1}{m}\sum_{i=1}^{m}\frac{(\rho^T Z_i)^3}{1+\rho^T Z_i}) \\
&= \frac{1}{m}\sum_{i=1}^{m}\rho^T Z_i-\frac{(1+o_p(1))}{m}\sum_{i=1}^{m}(\rho^T Z_i)^2,
\end{aligned}$$

which implies

$$\frac{1}{m}\sum_{i=1}^{m}\rho^T Z_i=\frac{(1+o_p(1))}{m}\sum_{i=1}^{m}(\rho^T Z_i)^2. \tag{2.70}$$

Using (2.69) and (2.67) we obtain

$$\begin{aligned}
0 &= \frac{1}{m}\sum_{i=1}^{m}\frac{Z_i}{1+\rho^T Z_i} \\
&= \frac{1}{m}\sum_{i=1}^{m}Z_i(1-\rho Z_i+\frac{(\rho^T Z_i)^2}{1+\rho^T Z_i}) \\
&= \frac{1}{m}\sum_{i=1}^{m}Z_i-\frac{1}{m}\sum_{i=1}^{m}Z_i(Z_i)^T\rho+\frac{1}{m}\sum_{i=1}^{n}\frac{Z_i(\rho^T Z_i)^2}{1+\rho^T Z_i}) \\
&= \frac{1}{m}\sum_{i=1}^{m}Z_i-\frac{1}{m}\sum_{i=1}^{m}Z_i(Z_i)^T\rho+O_p\left(\max_{1\le i\le m}||\frac{Z_i}{1+\rho^T Z_i}||\frac{1}{m}\sum_{i=1}^{n}(\rho^T Z_i)^2\right) \\
&= \frac{1}{m}\sum_{i=1}^{m}Z_i-\frac{1}{m}\sum_{i=1}^{m}Z_i(Z_i)^T\rho+o_p\left(m^{1/2}\rho^T(\frac{1}{m}\sum_{i=1}^{n}Z_i(Z_i)^T)\rho\right) \\
&= \frac{1}{m}\sum_{i=1}^{m}Z_i-\frac{1}{m}\sum_{i=1}^{m}Z_i(Z_i)^T\rho+o_p(m^{-1/2})
\end{aligned}$$

which implies

$$\rho=\left(\frac{1}{m}\sum_{i=1}^{m}Z_i(Z_i)^T\right)^{-1}\frac{1}{m}\sum_{i=1}^{m}Z_i+o_p(m^{-1/2}). \tag{2.71}$$

Finally by using Taylor's expansion, (2.70), (2.71), (2.66) and (2.67) we obtain

$$
\begin{aligned}
-2\log L_{n2}(\beta_0) &= 2\sum_{i=1}^{m}\rho^T Z_i - (1+o_p(1))\sum_{i=1}^{m}(\rho^T Z_i)^2) \\
&= (1+o_p(1))\rho^T\left(\sum_{i=1}^{m} Z_i(Z_i)^T\right)\rho \\
&= (1+o_p(1))(\frac{1}{\sqrt{m}}\sum_{i=1}^{m} Z_i)^T\left(\frac{1}{m}\sum_{i=1}^{m} Z_i(Z_i)^T\right)^{-1}\frac{1}{\sqrt{m}}\sum_{i=1}^{m} Z_i + o_p(1) \\
&\xrightarrow{d} \chi_2^2.
\end{aligned}
$$

This completes the proof of Theorem 2.3.1. □

Proof of Example 1. Set

$$(x_1,\cdots,x_p)^T = \Sigma^{-1/2}X_1 \quad \text{and} \quad (y_1,\cdots,y_p)^T = \Sigma^{-1/2}\tilde{X}_1.$$

Then $x_1,\cdots,x_p$, $y_1,\cdots,y_p$ are independent standard normal random variables. Therefore we have $X_1 = \Sigma^{1/2}(x_1,\cdots,x_p)^T$ and $\tilde{X}_1 = \Sigma^{1/2}(y_1,\cdots,y_p)^T$, and

$$X_1^T\tilde{X}_1 = (x_1,\cdots,x_p)\Sigma(y_1,\cdots,y_p)^T = \sum_{1\le i,j\le p}\sigma_{i,j}x_iy_j.$$

In order to estimate $\mathbb{E}(X_1^T\tilde{X}_1)^4$, we set $\delta_{j_1,j_2,j_3,j_4} = \mathbb{E}(x_{j_1}x_{j_2}x_{j_3}x_{j_4}) = \mathbb{E}(y_{j_1}y_{j_2}y_{j_3}y_{j_4})$. Then δ_{j_1,j_2,j_3,j_4} is equal to 3 if $j_1=j_2=j_3=j_4$, 1 if j_1,j_2,j_3 and j_4 are two different pairs of integers, and 0 otherwise.

Then we have

$$
\begin{aligned}
\mathbb{E}(X_1^T\tilde{X}_1)^4 &= \sum_{1\le j_1,j_2,j_3,j_4\le p}\sum_{1\le l_1,l_2,l_3,l_4\le p}\sigma_{j_1,l_1}\sigma_{j_2,l_2}\sigma_{j_3,l_3}\sigma_{j_4,l_4}\delta_{j_1,j_2,j_3,j_4}\delta_{l_1,l_2,l_3,l_4}\\
&= O\left(|\sum_{j_1\neq j_2}\sum_{l_1\neq l_2}\sigma_{j_1,l_1}\sigma_{j_1,l_2}\sigma_{j_2,l_1}\sigma_{j_2,l_2}|\right)+O\left(\sum_{j_1\neq j_2}\sum_{l=1}^{p}\sigma_{j_1,l}^2\sigma_{j_2,l}^2\right)\\
&\quad +O\left(\sum_{j=1}^{p}\sum_{l_1\neq l_2}\sigma_{j,l_1}^2\sigma_{j,l_2}^2\right)+O\left(\sum_{j=1}^{p}\sum_{l=1}^{p}\sigma_{j,l}^4\right)\\
&= O\left(|\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}\sum_{l_1=1}^{p}\sum_{l_2=1}^{p}\sigma_{j_1,l_1}\sigma_{j_1,l_2}\sigma_{j_2,l_1}\sigma_{j_2,l_2}|\right)\\
&\quad +O\left(\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}\sum_{l=1}^{p}\sigma_{j_1,l}^2\sigma_{j_2,l}^2\right)+O\left(\sum_{j=1}^{p}\sum_{l=1}^{p}\sigma_{j,l}^4\right)\\
&= O\left(\mathrm{tr}(\Sigma^4)\right)+O\left((\sum_{j=1}^{p}\sum_{l=1}^{p}\sigma_{j,l}^2)^2\right)\\
&= O\left(\mathrm{tr}(\Sigma^4)\right)+O\left((\mathrm{tr}(\Sigma^2))^2\right)\\
&= O\left((\mathrm{tr}(\Sigma^2))^2\right).
\end{aligned}
$$

We have used the inequality $\mathrm{tr}(\Sigma^4)\le(\mathrm{tr}(\Sigma^2))^2$, which follows from the identity $\mathrm{tr}(\Sigma^i)=\sum_{j=1}^{p}\lambda_j^i$ for any positive integer i, where $\lambda_1,\cdots,\lambda_p$ are eigenvalues of Σ. Thus we have that $\dfrac{\mathbb{E}(X_1^T\tilde{X}_1)^4}{(\mathrm{tr}(\Sigma^2))^2}=O(1)$ is bounded uniformly for p.

Similarly, we can show that the first term on the left-hand side of (2.55) is also bounded uniformly for p. Therefore, conditions (2.54) and (2.55) will be fulfilled with $\delta=2$ for any p if $\mathbb{E}(\epsilon_1^4)/\sigma^4=o(m^{1/2})$. □

Proof of Example 2. It follows from the same argument in the proof of Corollary 2.2.2. □

2.4 Tests for High-dimensional Covariance Matrices

Testing covariance structure is of importance in many areas of statistical analysis, such as microarray analysis and signal processing. Conventional tests for finite-dimensional covariance can not be applied to high-dimensional data in general, and tests for high-dimensional covariance in the literature usually depend on some special structure of the matrix. In this section, we propose an empirical likelihood method to test the covariance matrix by simply splitting the data into two groups. The asymptotic distribution of the new test is independent of the dimension. A simulation study shows that the new test has a very stable size with respect to the dimension and it is also more powerful than the test proposed by Cai and Jiang [12] for testing the bandedness of a covariance matrix in the dense model.

2.4.1 Introduction

Let $X_i = (X_{i1}, \ldots, X_{ip}),\ i = 1, 2, \ldots, n$ be independent and identically distributed (i.i.d.) random vectors with mean $\mu = (\mu_1, \ldots, \mu_p)$ and covariance $\Sigma = (\sigma_{ij})_{1\le i,j\le p}$. Testing covariance matrix

$$H_0 : \Sigma = \Sigma_0 \text{ against } H_1 : \Sigma \neq \Sigma_0 \tag{2.72}$$

is an important problem in statistical inference and applications. There has been a long history for the study of this problem. Traditional methods for testing (2.72) with finite p include the likelihood ratio test (see Anderson [1]) and the scaled distance measure defined as

$$V = \frac{1}{p}\mathrm{tr}\Big(S_n - I_p\Big)^2, \tag{2.73}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix and S_n is the sample covariance matrix of $\Sigma_0^{-1/2} X_i$ (see John ([49, 50]) and Nagao [67]). When dealing with high-dimensional data, the sample covariance in the likelihood ratio test is no longer invertible with probability one and the tests based on a scaled distance may also fail as demonstrated in Ledoit and Wolf [62].

Since the above conventional tests can not be employed for testing high-dimensional covariance matrix, new methods are needed. When the high-dimensional covariance matrix has a modest dimension p compared to the sample size n, i.e. $p/n \to c$ for some $c \in (0, \infty)$, Ledoit and Wolf [62] proposed a test by modifying the scaled distance measure V defined in (2.73) under the assumption that X_1 is a normal random vector. When the dimension p is much larger than the sample size n, some special structure has to be imposed. Chen, Zhang and Zhong [18] proposed a test which generalizes the result of Ledoit and Wolf [62] to the case of non-normal distribution and large dimension by assuming that

$$X_i = \Gamma Z_i + \mu$$

for some i.i.d. m-dimensional random vectors $\{Z_i\}$ with $\mathrm{E}Z_1 = 0$, $\mathrm{var}(Z_1) = I_m$, and Γ is a $p \times m$ constant matrix with $\Gamma\Gamma^T = \Sigma$.

Another commonly employed special structure is sparsity. High-dimensional sparse data setting, where dimension p is larger than the sample size n, is frequently encountered in signal processing and gene expression experiments, see for example Sebastini, Gussoni, Kohane and Ramoni [89]. Estimating covariance matrix with sparsity has been actively studied in the recent years. Some recent references are Bickel and Levina [9], Cai, Zhang and Zhou [14], and Cai and Liu [11]. When the sparsity assumes that the covariance matrix has a desired banded structure, it becomes important to test whether the covariance matrix possesses such a desired structure, i.e.

$$H_0 : \sigma_{ij} = 0 \text{ for all } |i - j| \geq \tau, \tag{2.74}$$

where $\tau < p$ is given and may depend on n. Recently, Cai and Jiang [12] proposed to use the maximum of the absolute values of sample covariances to test (2.74) when X_1 has a multivariate Gaussian distribution. However, it is known that the convergence rate of the normalized maximum to a Gumbel limit is very slow, which means such a test is not powerful in general.

To get rid of the sparse structure and normality condition in the testing problems (2.72) and (2.74), we propose to construct tests based on the following equivalent testing problem. Write

$$\boldsymbol{a} = (\sigma_{11}, \ldots, \sigma_{1p}, \sigma_{21}, \ldots, \sigma_{2p}, \ldots, \sigma_{p1}, \ldots, \sigma_{pp})^T.$$

Then testing $H_0 : \Sigma = \Sigma_0 =: (\sigma_{ij}^0)$ is equivalent to testing

$$\boldsymbol{a} = \boldsymbol{a}_0 := (\sigma_{11}^0, \ldots, \sigma_{1p}^0, \sigma_{21}^0, \ldots, \sigma_{2p}^0, \ldots, \sigma_{p1}^0, \ldots, \sigma_{pp}^0)^T. \tag{2.75}$$

Put $Y_i = ((X_{i1} - \mu_1)^2, \ldots, (X_{i1} - \mu_1)(X_{ip} - \mu_p), (X_{i2} - \mu_2)(X_{i1} - \mu_1), \ldots, (X_{ip} - \mu_p)^2)^T$, $i = 1, \ldots, n$. Based on the fact that $\mathbb{E}Y_i = \boldsymbol{a}$, one can employ the well-known Hotelling T^2 statistic for finite p or its modified versions for divergent p under some specific models to test (2.75); see for example Bai and Saranadasa [5] and Chen and Qin [16].

Another popular test for a mean vector is the empirical likelihood method proposed by Owen [71, 72]. Unfortunately, the asymptotic distribution of the empirical likelihood ratio test depends on whether the dimension is fixed or diverges; see Hjort, Mckeague and Van Keilegom [45].

Motivated by the empirical likelihood method in Peng, Qi and Wang [78] for testing a high dimensional mean vector, we propose to apply the empirical likelihood method to the following two equations

$$\mathrm{E}[(Y_1 - \boldsymbol{a}_0)^T(Y_2 - \boldsymbol{a}_0)] = 0 \text{ and } \mathrm{E}[\mathbf{1}_{p^2}^T(Y_1 + Y_2 - 2\boldsymbol{a}_0)] = 0, \tag{2.76}$$

where $\mathbf{1}_{p^2} = (1, \ldots, 1)^T \in \mathbb{R}^{p^2}$. The first equation in (2.76) ensures the consistence of the proposed test and the second equation in (2.76) is used to improve the test power, since using only the first equation will lead to a poor power by noting that $\mathrm{E}[(Y_1 - \boldsymbol{a}_0)^T(Y_2 - \boldsymbol{a}_0)] = O(\delta^2)$ rather than $O(\delta)$ if $||\mathrm{E}(Y_1 - \boldsymbol{a}_0)|| = O(\delta)$, where $||\cdot||$ is the Euclidean norm for a vector. It turns out that the proposed empirical likelihood test puts no restriction on the sparse structure of the matrix and normality of X_1. When testing (2.74), a similar procedure can be employed; see Section 2 for more details.

The rest of this section is organized as follows. In Section 2.4.2, we introduce the new methodology and present the main results. A simulation study is given in Section 2.4.3. Section 2.4.4 contains the proofs of the main results.

2.4.2 Methodology

Testing covariance matrix. Let $X_i = (X_{i1}, \ldots, X_{ip})$, $i = 1, \ldots, n$ be independent and identically distributed observations with mean $\mu = (\mu_1, \ldots, \mu_p)$ and covariance $\Sigma = (\sigma_{ij})$. Instead of testing the covariance matrix hypothesis (2.72) directly, we consider testing a p^2-dimensional vector $\boldsymbol{a}$, i.e., testing

$$\begin{aligned} H_0 : \boldsymbol{a} &= (\sigma_{11}, \ldots, \sigma_{1p}, \sigma_{21}, \ldots, \sigma_{2p}, \ldots, \sigma_{p1}, \ldots, \sigma_{pp})^T \\ &= (\sigma_{11}^0, \ldots, \sigma_{1p}^0, \sigma_{21}^0, \ldots, \sigma_{2p}^0, \ldots, \sigma_{p1}^0, \ldots, \sigma_{pp}^0)^T =: \boldsymbol{a}_0. \end{aligned}$$

When μ is known, for $i = 1, \ldots, n$, we define

$$Y_i = ((X_{i1} - \mu_1)^2, \ldots, (X_{i1} - \mu_1)(X_{ip} - \mu_p), (X_{i2} - \mu_2)(X_{i1} - \mu_1), \ldots, (X_{ip} - \mu_p)^2)^T.$$

Then $\mathrm{E}[(Y_1 - \boldsymbol{a}_0)^T (Y_2 - \boldsymbol{a}_0)] = 0$ is equivalent to $H_0 : (\boldsymbol{a} - \boldsymbol{a}_0)^T (\boldsymbol{a} - \boldsymbol{a}_0) = 0$, which is equivalent to $H_0 : \boldsymbol{a} = \boldsymbol{a}_0$. A direct application of the empirical likelihood method to the above estimating equation results in a poor power as explained in the introduction. A brief simulation study confirms this fact. In order to improve the test power, we propose to add one more linear equation. Note that with prior information on the model or more specific alternative hypothesis, a more proper linear equation may be obtained. With no additional information, any linear equation that detects the change of order $||\boldsymbol{a} - \boldsymbol{a}_0||$ is a possible choice theoretically. Here we simply choose the following functional $\mathbf{1}_{p^2}^T (Y_1 + Y_2 - 2\boldsymbol{a}_0)$. More specifically, we propose to apply the empirical likelihood method to the following two equations

$$\mathrm{E}\{(Y_1 - \boldsymbol{a}_0)^T (Y_2 - \boldsymbol{a}_0)\} = 0 \quad \text{and} \quad \mathrm{E}\{\mathbf{1}_{p^2}^T (Y_1 + Y_2 - 2\boldsymbol{a}_0)\} = 0.$$

Of course one can try other linear equations or add more equations to further improve the power. Simulation study in Section 3 shows that with the above two estimating equations, the proposed test performs well in terms of both size and power.

In order to obtain an independent paired data (Y_1, Y_2), we split the sample into two subsamples with size $N = [n/2]$. That is, for $i = 1, 2, \ldots, N$, we define $\boldsymbol{R}_i(\boldsymbol{a}) =$

$(e_i(\boldsymbol{a}), v_i(\boldsymbol{a}))^T$, where

$$e_i(\boldsymbol{a}) = (Y_i - \boldsymbol{a})^T(Y_{i+N} - \boldsymbol{a}) \text{ and } v_i(\boldsymbol{a}) = \mathbf{1}_{p^2}^T(Y_i + Y_{i+N} - 2\boldsymbol{a}).$$

Based on $\{\boldsymbol{R}_i(\boldsymbol{a})\}_{i=1}^N$, we define the empirical likelihood ratio function for $\boldsymbol{a}$ as

$$L_1(\boldsymbol{a}) = \sup\{\prod_{i=1}^N (Np_i) : \sum_{i=1}^N p_i = 1, \sum_{i=1}^N p_i \boldsymbol{R}_i(\boldsymbol{a}) = 0, p_1 \geq 0, \ldots, p_N \geq 0\}. \quad (2.77)$$

When μ is unknown, instead of using $\{\boldsymbol{R}_i(\boldsymbol{a})\}_{i=1}^N$, we use $\{\boldsymbol{R}_i^*(\boldsymbol{a})\}_{i=1}^N$ where μ is replaced by the sample means. That is, put $\overline{X_j^1} = \frac{1}{N}\sum_{i=1}^N X_{ij}$ and $\overline{X_j^2} = \frac{1}{N}\sum_{i=N+1}^{2N} X_{ij}$ for $j = 1, \ldots, p$, and define

$$Y_i^* = ((X_{i1} - \overline{X_1^1})^2, \ldots, (X_{i1} - \overline{X_1^1})(X_{ip} - \overline{X_p^1}), (X_{i2} - \overline{X_2^1})(X_{i1} - \overline{X_1^1}), \ldots, (X_{ip} - \overline{X_p^1})^2)^T$$

for $i = 1, \ldots, N$, and

$$Y_i^* = ((X_{i1} - \overline{X_1^2})^2, \ldots, (X_{i1} - \overline{X_1^2})(X_{ip} - \overline{X_p^2}), (X_{i2} - \overline{X_2^2})(X_{i1} - \overline{X_1^2}), \ldots, (X_{ip} - \overline{X_p^2})^2)^T$$

for $i = N+1, \ldots, 2N$. Put $\boldsymbol{R}_i^*(\boldsymbol{a}) = (e_i^*(\boldsymbol{a}), v_i^*(\boldsymbol{a}))^T$, where

$$e_i^*(\boldsymbol{a}) = (Y_i^* - \boldsymbol{a})^T(Y_{i+N}^* - \boldsymbol{a}) \text{ and } v_i^*(\boldsymbol{a}) = \mathbf{1}_{p^2}^T(Y_i^* + Y_{i+N}^* - 2\boldsymbol{a}).$$

Similar to (2.77), define the empirical likelihood ratio function for $\boldsymbol{a}$ as

$$L_2(\boldsymbol{a}) = \sup\{\prod_{i=1}^N (Np_i) : \sum_{i=1}^N p_i = 1, \sum_{i=1}^N p_i \boldsymbol{R}_i^*(\boldsymbol{a}) = 0, p_1 \geq 0, \ldots, p_N \geq 0\}. \quad (2.78)$$

Let $q = p^2$ and $\Theta = (\theta_{ij})_{q\times q}$ be the covariance matrix of Y_1, i.e., $\Theta = \mathrm{E}[(Y_1 - \boldsymbol{a})(Y_1 - \boldsymbol{a})^T]$. Then $\mathbb{E}(e_1^2(\boldsymbol{a})) = \sum_{i=1}^q \sum_{j=1}^q \theta_{ij}^2$ and $\mathbb{E}(v_1^2(\boldsymbol{a})) = \sum_{i=1}^q \sum_{j=1}^q \theta_{ij}$. First we show that Wilks' Theorem holds for the above empirical likelihood methods without imposing any special structure. Note that in the following theorems, the condition $\sum_{i=1}^p \sum_{j=1}^p \sigma_{ij} = \mathbb{E}(\sum_{j=1}^p X_{1j})^2 > 0$ simply means that $e_1(\boldsymbol{a})$ and $v_1(\boldsymbol{a})$ are not constants.

Theorem 2.4.1. *Suppose that* $\sum_{i=1}^p \sum_{j=1}^p \sigma_{ij} > 0$ *and for some* $\delta > 0$,

$$\max\left\{ \mathrm{E}|e_1(\boldsymbol{a})|^{2+\delta} / \Big(\sum_{i=1}^q \sum_{j=1}^q \theta_{ij}^2 \Big)^{\frac{2+\delta}{2}}, \mathrm{E}|v_1(\boldsymbol{a})|^{2+\delta} / \Big(\sum_{i=1}^q \sum_{j=1}^q \theta_{ij} \Big)^{\frac{2+\delta}{2}} \right\} = o(N^{\frac{\delta+\min\{2,\delta\}}{4}}) \quad (2.79)$$

Then both $-2\log L_1(\boldsymbol{a}_0)$ *and* $-2\log L_2(\boldsymbol{a}_0)$ *converge in distribution to a chi-square distribution with two degrees of freedom as* $n \to \infty$.

Remark 2.4.1. (2.79) can be interpreted as $e_1(\boldsymbol{a})$ and $v_1(\boldsymbol{a})$ satisfy condition (P).

Using Theorem 2.4.1, one can test $H_0 : \Sigma = \Sigma_0$ against $\Sigma \neq \Sigma_0$. Condition (2.79) ensures that the central limit theorem holds for $\frac{1}{\sqrt{N}} \sum_{i=1}^N e_i(\boldsymbol{a}_0)$ and $\frac{1}{\sqrt{N}} \sum_{i=1}^N v_i(\boldsymbol{a}_0)$, Similar to Section 2.2 and 2.3, the conditions in Theorem 2.4.1 can be simplified by imposing some conditions on the moments and dimension of X_1. Note that here the sample $\{Y_i\}$ is of size p^2.

Corollary 2.4.2. *Suppose* Y_1 *satisfy (A), then both* $-2\log L_1(\boldsymbol{a}_0)$ *and* $-2\log L_2(\boldsymbol{a}_0)$ *converge in distribution to a chi-square distribution with two degrees of freedom as* $n \to \infty$.

Theorem 2.4.3. *Suppose* $\{X\}$ *satisfies condition (B') with* $\sum_{i=1}^p \sum_{j=1}^p \sigma_{ij} > 0$. *Then both* $-2\log L_1(\boldsymbol{a}_0)$ *and* $-2\log L_2(\boldsymbol{a}_0)$ *converge in distribution to a chi-square distribution with two degrees of freedom as* $n \to \infty$.

Remark 2.4.2. For testing $H_0 : \Sigma = I_p$, where I_p denotes the $p \times p$ identity matrix, Chen, Zhang and Zhong [18] proposed a test based on the above model and required $p \to \infty$ as $n \to \infty$. In comparison, the proposed empirical likelihood method works for both fixed and divergent p.

Testing bandedness. Suppose $\{X_i\}$ is a sequence of i.i.d. normal random vectors with mean zero and covariance $\Sigma = (\sigma_{ij})_{1 \le i,j \le p}$. Cai and Jiang [12] proposed to use the maximum of the absolute values of the sample correlations (called the coherence) to test a banded structure

$$H_0 : \sigma_{ij} = 0 \text{ for all } |i-j| \ge \tau, \quad (2.80)$$

where $\tau < p$. It is known that the rate of convergence of coherence to a Gumble distribution is very slow in general, which results in a not powerful test. Here we modify the proposed empirical likelihood method to test the above banded structure as follows.

Define

$$Y_l' = (X_{l1}X_{l(1+\tau)},\ \ldots, X_{l1}X_{lp},\ X_{l2}X_{l(2+\tau)}, \ldots, X_{l(p-\tau)}X_{lp})^T,\ l = 1, \ldots, n,$$

$\boldsymbol{a}' = (\sigma_{1(1+\tau)}, \ldots, \sigma_{1p}, \sigma_{2(2+\tau)}, \ldots, \sigma_{(p-\tau)p})^T$ and $\boldsymbol{a}_0'^T = (0, 0, \ldots, 0) \in \mathbb{R}^{(p-\tau)(p+1-\tau)/2}$, then testing (2.80) is equivalent to testing $H_0 : \boldsymbol{a}' = \boldsymbol{a}_0'$. As before, define

$$e_i'(\boldsymbol{a}') = (Y_i' - \boldsymbol{a}')^T (Y_{N+i}' - \boldsymbol{a}'), \tag{2.81}$$

and

$$v_i'(\boldsymbol{a}') = \mathbf{1}^T_{(p-\tau)(p+1-\tau)/2}(Y_i' + Y_{N+i}' - 2\boldsymbol{a}'). \tag{2.82}$$

Based on $\boldsymbol{R}_i'(\boldsymbol{a}') = (e_i'(\boldsymbol{a}'), v_i'(\boldsymbol{a}'))^T$, we define the empirical likelihood function for $\boldsymbol{a}'$ as

$$L_3(\boldsymbol{a}') = \sup\{\prod_{i=1}^{N}(Np_i) : \sum_{i=1}^{N} p_i = 1, \sum_{i=1}^{N} p_i \boldsymbol{R}_i'(\boldsymbol{a}') = 0, p_i \geq 0, i = 1, \ldots, N\}. \tag{2.83}$$

Theorem 2.4.4. *Suppose that X_i follows the model of Cai and Jiang [12], i.e., $X_i \sim N(0, \Sigma)$. Assume $C_1 \leq \liminf_{n\to\infty} \min_{1\leq i\leq p} \sigma_{ii} \leq \limsup_{n\to\infty} \max_{1\leq i\leq p} \sigma_{ii} \leq C_2$ for some constants $0 < C_1 \leq C_2 < \infty$, and $\tau = o(\min\{\frac{(\sum_{1\leq i,j\leq p}\sigma_{ij})}{(\sum_{1\leq i,j\leq p}|\sigma_{ij}|)^{1/2}}, (\sum_{1\leq i,j\leq p}\sigma_{ij}^2)^{1/2}\})$. Then under H_0 in (2.80), $-2\log L_3(\boldsymbol{a}_0')$ converges in distribution to a chi-square distribution with two degrees of freedom as $n \to \infty$.*

Remark 2.4.3. The test in Cai and Jiang [12] requires that $\tau = o(p^s)$ for all $s > 0$, $\log p = o(n^{1/3})$. However, the new test only imposes conditions between τ and p. Note that with more information on the higher-order moments of X_i, one can impose conditions such as (2.79) in Theorem 2.4.1 for e_i' and v_i' to test the bandedness so that the normality assumption is not required.

The condition $\tau = o(\sum_{1\le i,j\le p}\sigma_{ij}/(\sum_{1\le i,j\le p}|\sigma_{ij}|)^{1/2})$ is sometimes difficult to check. Next we remove this condition in the above theorem by choosing a different linear equation in defining v_i' in (2.82). More specifically, define

$$\tilde{v}_i'(\boldsymbol{a}') = \sum_{k=1}^{t}\sum_{j=t+\tau}^{p}(X_{ik}X_{ij}-\sigma_{kj}) + \sum_{k=1}^{t}\sum_{j=t+\tau}^{p}(X_{N+i,k}X_{N+i,j}-\sigma_{kj}), \tag{2.84}$$

where $t = [(p-\tau)/2]$. Based on $\tilde{\boldsymbol{R}}_i'(\boldsymbol{a}') = (e_i'(\boldsymbol{a}'), \tilde{v}_i'(\boldsymbol{a}'))^T$, we define the empirical likelihood function for $\boldsymbol{a}'$ as

$$L_4(\boldsymbol{a}') = \sup\{\prod_{i=1}^{N}(Np_i) : \sum_{i=1}^{N}p_i = 1, \sum_{i=1}^{N}p_i\tilde{\boldsymbol{R}}_i'(\boldsymbol{a}') = 0, p_i \ge 0, i = 1,\ldots,N\}. \tag{2.85}$$

Theorem 2.4.5. *Suppose that X_i follows the model of Cai and Jiang [12], i.e., $X_i \sim N(0,\Sigma)$. Assume $C_1 \le \liminf_{n\to\infty}\min_{1\le i\le p}\sigma_{ii} \le \limsup_{n\to\infty}\max_{1\le i\le p}\sigma_{ii} \le C_2$ for some constants $0 < C_1 \le C_2 < \infty$, and $\tau = o((\sum_{1\le i,j\le p}\sigma_{ij}^2)^{1/2})$. Then under H_0 in (2.80), $-2\log L_4(\boldsymbol{a}_0')$ converges in distribution to a chi-square distribution with two degrees of freedom as $n\to\infty$.*

Remark 2.4.4. From the proof we can see that the above theorem holds for any choice of t. Different t can be chosen to improve the power of the proposed test, based on some prior information. Since $\tau = o(p^{1/2})$ implies $\tau = o((\sum_{1\le i,j\le p}\sigma_{ij}^2)^{1/2})$, the proposed test imposes much weaker conditions on τ and p than those in Cai and Jiang [12].

Power analysis. In the following we study the power analysis of our new tests. Denote $\pi_{11} = \sum_{i=1}^{q}\sum_{i=1}^{q}\theta_{ij}^2 = \mathbb{E}(e_1^2(\boldsymbol{a}))$, $\pi_{22} = 2\sum_{i=1}^{q}\sum_{i=1}^{q}\theta_{ij} = \mathbb{E}(v_1^2(\boldsymbol{a}))$,

$$\zeta_{n1} = (\boldsymbol{a}-\boldsymbol{a}_0)^T(\boldsymbol{a}-\boldsymbol{a}_0)/\sqrt{\pi_{11}} = \mathrm{tr}((\Sigma-\Sigma_0)^2)/\sqrt{\pi_{11}}$$

and

$$\zeta_{n2} = 2\mathbf{1}_q^T(\boldsymbol{a}-\boldsymbol{a}_0)/\sqrt{\pi_{22}} = 2\mathbf{1}_p^T(\Sigma-\Sigma_0)\mathbf{1}_p/\sqrt{\pi_{22}}.$$

Theorem 2.4.6. *In addition to the conditions of Theorem 2.1, if* $H_1 : \boldsymbol{a} \neq \boldsymbol{a}_0$ *holds with*

$$\zeta_{n1} = o(1), \tag{2.86}$$

then

$$P\{-2\log L_1(\boldsymbol{a}_0) > \xi_{1-\alpha}\} = P\{\chi^2_{2,\nu} > \xi_{1-\alpha}\} + o(1) \tag{2.87}$$

to as $n \to \infty$*, where* $\chi^2_{2,\nu}$ *is a noncentral chi-square distribution with two degrees of freedom and noncentrality parameter* $\nu = N(\zeta^2_{n1} + \zeta^2_{n2})$*,*

Remark 2.4.5. From the above power analysis, the new test rejects the null hypothesis with probability tending to one when $\sqrt{n}\zeta_{n1}$ or $\sqrt{n}|\zeta_{n2}|$ goes to infinity. Note that the test given in Chen, Zhang and Zhong [14] for the identity hypothesis $H_0 : \Sigma = I_p$ against $H_1 : \Sigma \neq I_p$ requires $n\rho_{1,n} \to \infty$ which is equivalent to $\rho_{2,n} \to 0$ where

$$\rho_{1,n} = \frac{1}{p}\mathrm{tr}((\Sigma - I_p)^2),$$

$$\rho_{2,n} = \frac{\mathrm{tr}(\Sigma^2)}{n\mathrm{tr}((\Sigma - I_p)^2)}.$$

See (3.4) and the proof of Theorem 4 in Chen, Zhang and Zhong (2010). When model (B) holds, $\sqrt{\pi_{11}} = O(\mathrm{tr}(\Sigma^2))$ (similar to the proof of Lemma **??**), therefore the condition $\rho_{2,n} \to 0$ is exactly $n\zeta_{n1} \to \infty$. Our test requires $\max(\sqrt{n}\zeta_{n1}, \sqrt{n}|\zeta_{n2}|) \to \infty$. Thus, our test may have a better power or a worse power in different settings.

Remark 2.4.6. For the tests for the banded structure introduced in Theorems 2.4.4 and 2.4.5, we have similar power results.

2.4.3 Simulation study

In this section we investigate the finite sample behavior of the proposed empirical likelihood test in terms of both size and power, and compare it with the test based on maximum in Cai and Jiang [12] for testing a banded structure.

Draw 2,000 random samples with sample size $n = 200$ or 500 from the random variable $W_1 + (\delta/\sqrt{n})^{0.5} W_2$, where $W_1 \sim N(0, (\sigma_{ij})_{1 \le i,j \le p})$, $\sigma_{ij} = 0.5^{|i-j|} I(|i-j| < \tau)$, $W_2 \sim N(0, 1_{p \times p})$ where $1_{p \times p}$ is a $p \times p$ matrix with all entries being 1, and W_1 is independent of W_2. We are interested in testing the banded structure $H_0 : \sigma_{ij} = 0$ for $|i-j| \ge \tau$. We consider $\tau = 5$ and increase p with a step 5 from 10 till 200. We also take $\tau = 20$ and start with $p = 25$ since $p > \tau$ is required. We plot the sizes ($\delta = 0$) and the powers ($\delta = 0.1, 0.5$) against p for the proposed empirical likelihood tests based on both Theorems 2.4.4 and 2.4.5, and the test based on maximum in Cai and Jiang [12] in Figures 2.5–2.8. In each figure, the solid line, dashed line and dotted line represent the proposed empirical likelihood tests based on Theorems 2.4.4 and 2.4.5, and the test based on maximum in Cai and Jiang [12], respectively.

From these figures, we observe that i) panels in the first row of each figure show that the proposed empirical likelihood tests have a more accurate size than the test in Cai and Jiang [12], and the size for these three tests becomes accurate when the sample size increases; ii) panels in the second and third rows of each figure show that the proposed empirical likelihood tests are much more powerful than the test in Cai and Jiang [12].

2.4.4 Proofs

Without loss of generality, we assume $\mu_0 = 0$ throughout. For simplicity, we use $||\cdot||$ to denote the L_2 norm of a vector or matrix and write $e_i(\boldsymbol{a}_0) = e_i$, $v_i(\boldsymbol{a}_0) = v_i$ and $e_i^*(\boldsymbol{a}_0) = e_i^*$, $v_i^*(\boldsymbol{a}_0) = v_i^*$. We first show some lemmas.

Lemma 2.4.7. *Under the conditions of Theorem 2.4.1, we have*

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Big(\frac{e_i}{\sqrt{\pi_{11}}}, \frac{v_i}{\sqrt{\pi_{22}}} \Big)^T \xrightarrow{d} N(0, I_2) \tag{2.88}$$

and

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} \Big(\frac{e_i^*}{\sqrt{\pi_{11}}}, \frac{v_i^*}{\sqrt{\pi_{22}}} \Big)^T \xrightarrow{d} N(0, I_2), \tag{2.89}$$

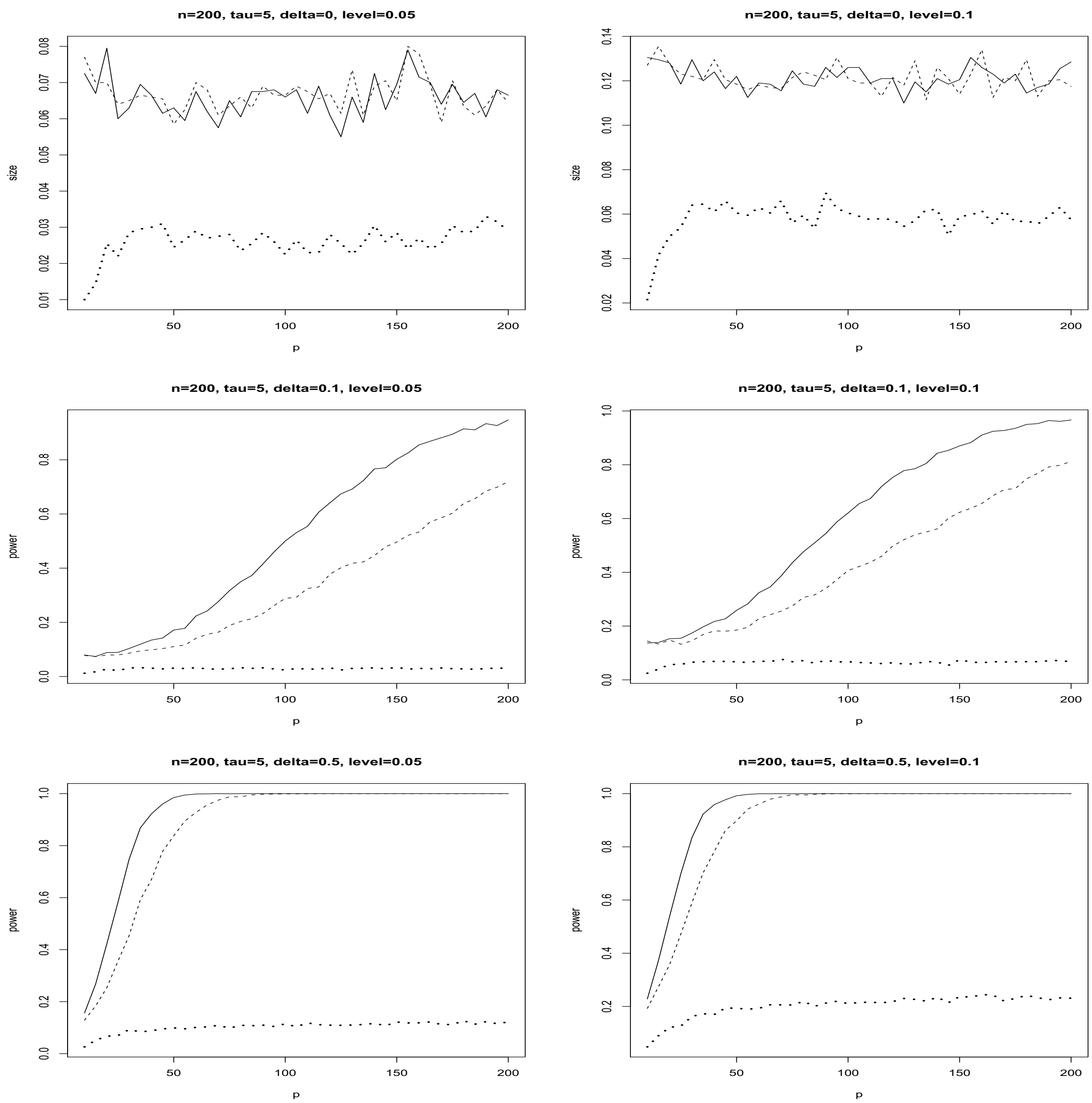

Figure 2.5: Powers of tests are plotted against $p = 10, 15, \ldots, 200$ with levels 0.05 and 0.1 for $n = 200$ and $\tau = 5$.

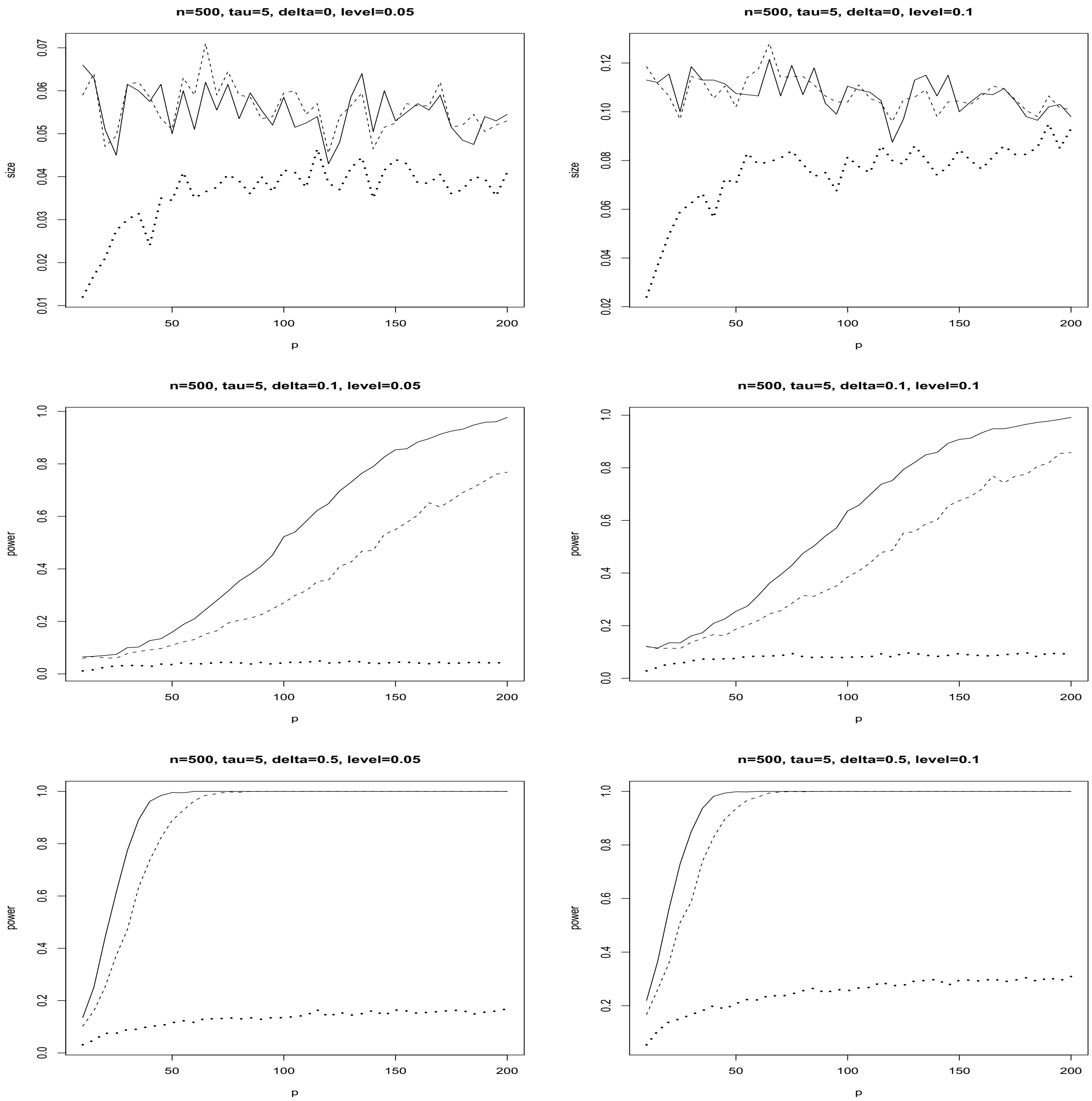

Figure 2.6: Powers of tests are plotted against $p = 10, 15, \ldots, 200$ with levels 0.05 and 0.1 for $n = 500$ and $\tau = 5$.

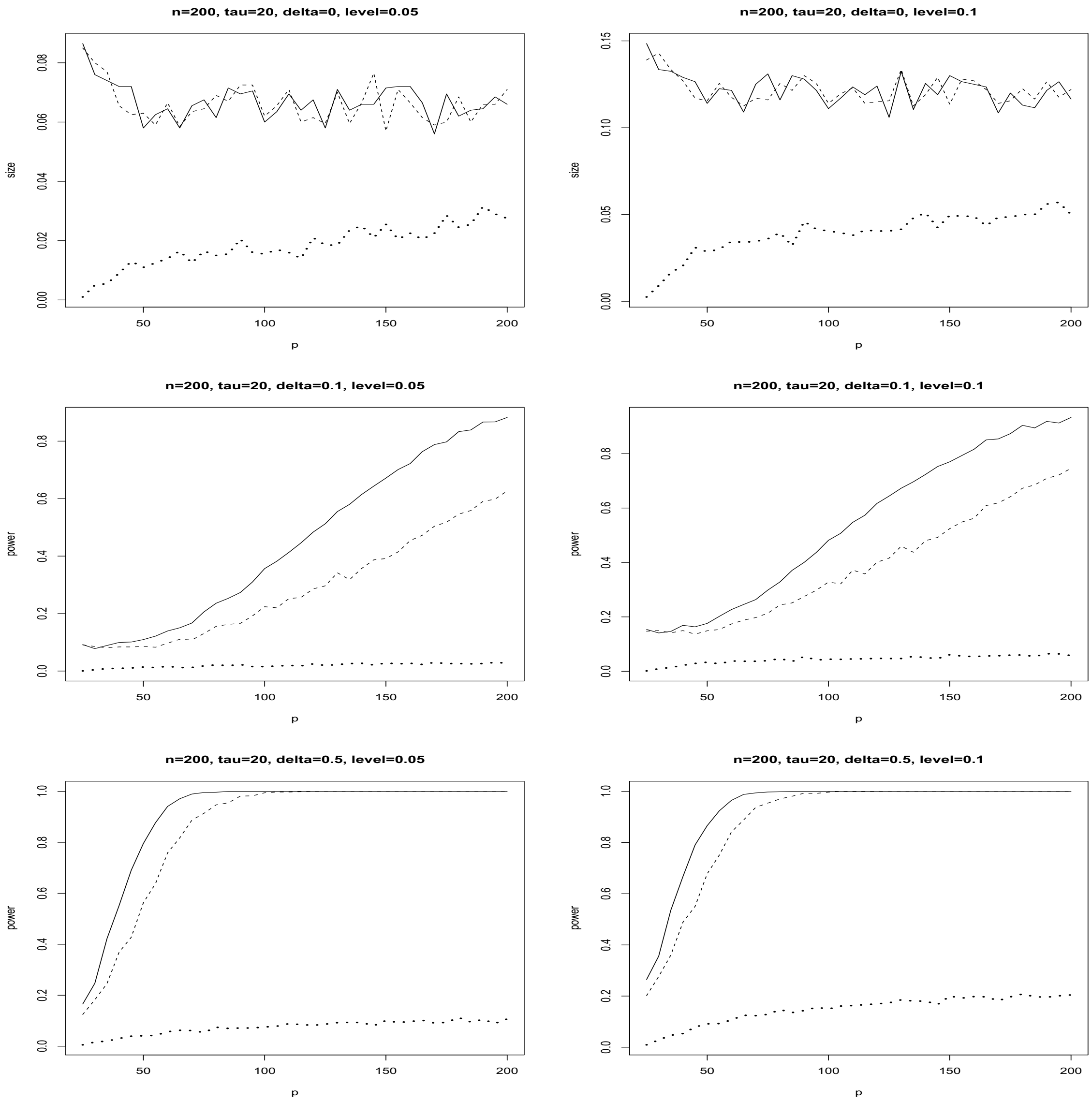

Figure 2.7: Powers of tests are plotted against $p = 25, 30, \ldots, 200$ with levels 0.05 and 0.1 for $n = 200$ and $\tau = 20$.

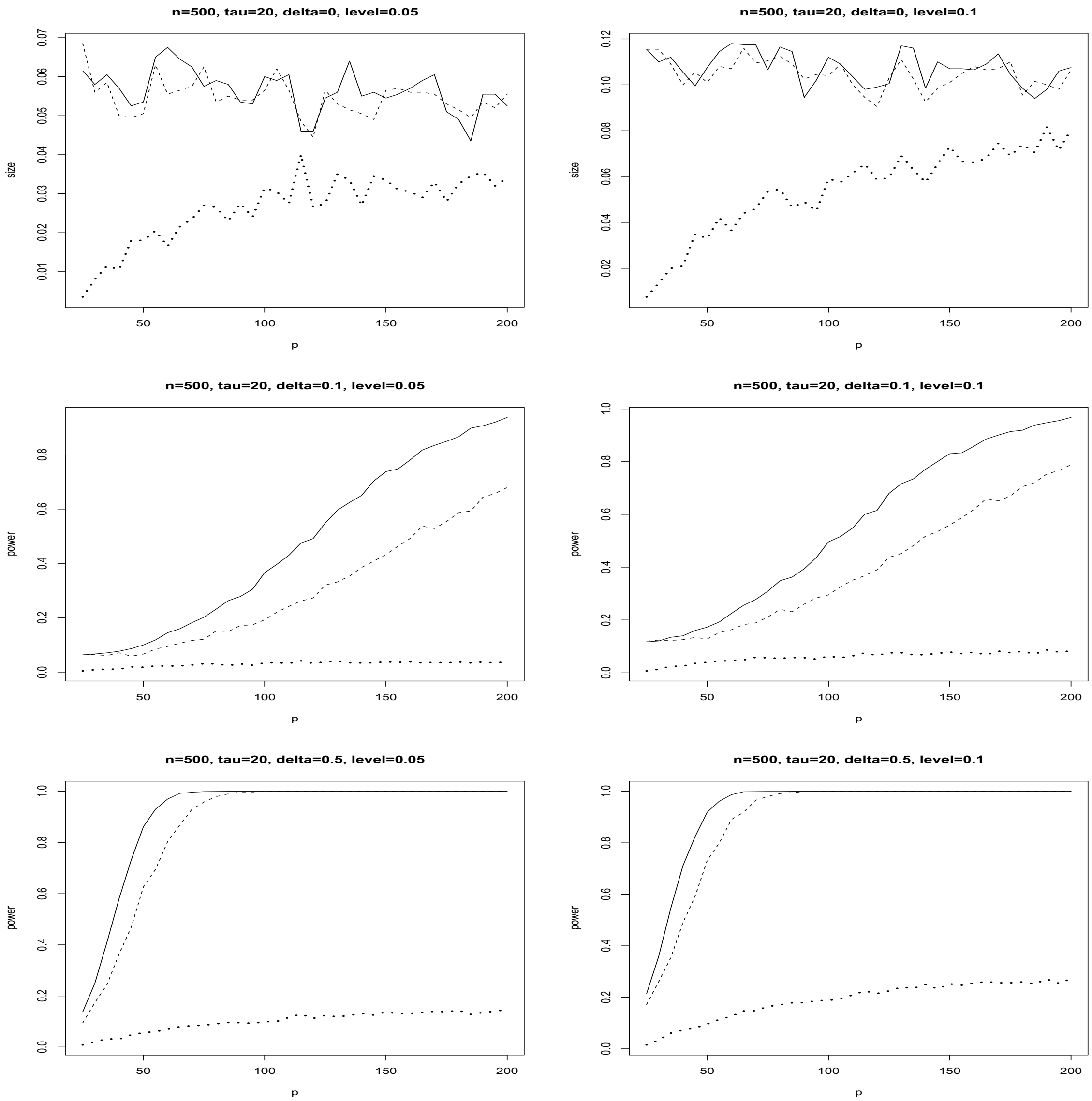

Figure 2.8: Powers of tests are plotted against $p = 25, 30, \ldots, 200$ with levels 0.05 and 0.1 for $n = 500$ and $\tau = 20$.

where I_2 *is the* 2×2 *identity matrix. Further,*

$$\frac{\sum_{i=1}^N e_i^2}{N\pi_{11}} - 1 \xrightarrow{p} 0, \qquad \frac{\sum_{i=1}^N e_i^{*2}}{N\pi_{11}} - 1 \xrightarrow{p} 0, \tag{2.90}$$

$$\frac{\sum_{i=1}^N v_i^2}{N\pi_{22}} - 1 \xrightarrow{p} 0, \qquad \frac{\sum_{i=1}^N v_i^{*2}}{N\pi_{22}} - 1 \xrightarrow{p} 0, \tag{2.91}$$

$$\frac{\sum_{i=1}^N e_i v_i}{N\sqrt{\pi_{11}\pi_{22}}} \xrightarrow{p} 0, \qquad \frac{\sum_{i=1}^N e_i^* v_i^*}{N\sqrt{\pi_{11}\pi_{22}}} \xrightarrow{p} 0, \tag{2.92}$$

$$\max_{1\le i\le N} |e_i/\sqrt{\pi_{11}}| = o_p(N^{1/2}), \qquad \max_{1\le i\le N} |v_i/\sqrt{\pi_{22}}| = o_p(N^{1/2}), \tag{2.93}$$

$$\max_{1\le i\le N} |e_i^*/\sqrt{\pi_{11}}| = o_p(N^{1/2}), \qquad \max_{1\le i\le N} |v_i^*/\sqrt{\pi_{22}}| = o_p(N^{1/2}). \tag{2.94}$$

Proof. Since the proofs for the sequence $\{(e_i^*, v_i^*)^T\}$ are similar to those for $\{(e_i, v_i)^T\}$, we only prove the cases of $\{(e_i, v_i)^T\}$. It is easily seen that $\mathrm{E}e_1 = \mathrm{E}v_1 = \mathrm{E}[e_1 v_1] = 0$ and

$$\mathrm{E}e_1^2 = \sum_{i=1}^q \sum_{i=1}^q \theta_{ij}^2 = \pi_{11}, \quad \mathrm{E}v_1^2 = 2\sum_{i=1}^q \sum_{i=1}^q \theta_{ij} = \pi_{22}.$$

Thus, by the Cramer-Wold device, for proving (2.88), it suffices to show that for any constants c, d,

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \Big(c\frac{e_i}{\sqrt{\pi_{11}}} + d\frac{v_i}{\sqrt{\pi_{22}}} \Big) \xrightarrow{d} N(0, c^2 + d^2). \tag{2.95}$$

Since $\{ce_i/\sqrt{\pi_{11}} + dv_i/\sqrt{\pi_{22}}\}$ is an i.i.d random sequence with mean zero and

$$\begin{aligned} & \frac{1}{N^{(2+\delta)/2}} \sum_{i=1}^N \mathrm{E}\Big| c\frac{e_i}{\sqrt{\pi_{11}}} + d\frac{v_i}{\sqrt{\pi_{22}}} \Big|^{2+\delta} \\ = \;& N^{-\delta/2}\mathrm{E}\Big| c\frac{e_1}{\sqrt{\pi_{11}}} + d\frac{v_1}{\sqrt{\pi_{22}}} \Big|^{2+\delta} \\ \le \;& 2^{1+\delta} N^{-\delta/2} [|c|^{2+\delta}\mathrm{E}|e_1/\sqrt{\pi_{11}}|^{2+\delta} + |d|^{2+\delta}\mathrm{E}|v_1/\sqrt{\pi_{22}}|^{2+\delta}] \\ = \;& 2^{2+\delta} N^{-\delta/2} o(N^{\frac{\delta+\min\{2,\delta\}}{4}}) = o(1), \end{aligned} \tag{2.96}$$

(2.95) follows from Lyapunov central limit theorem.

Next, we show the first equation in (2.90). Since $\mathrm{E}e_1^2 = \pi_{11}$, by von Bahr-Esseen's inequality [98], we have for any $0 < \delta \leq 2$,

$$\mathrm{E}|\sum_{i=1}^{N}(e_i^2 - \pi_{11})|^{(2+\delta)/2} \leq 2N\mathrm{E}|e_1^2 - \pi_{11}|^{(2+\delta)/2} = O(N\mathrm{E}|e_1|^{2+\delta}) \tag{2.97}$$

and when $\delta > 2$, it follows from Dharmadhikari, Fabian and Jogdeo [27] that

$$\mathrm{E}|\sum_{i=1}^{N}(e_i^2 - \pi_{11})|^{(2+\delta)/2} \leq CN^{(2+\delta)/4}\mathrm{E}|e_1^2 - \pi_{11}|^{(2+\delta)/2} = O(N^{(2+\delta)/4}\mathrm{E}|e_1|^{2+\delta}). \tag{2.98}$$

By (2.97), (2.98) and (2.79) we get

$$\mathrm{E}|\sum_{i=1}^{N}(e_i^2 - \pi_{11})|^{(2+\delta)/2} = O(N^{(2+\max\{2,\delta\})/4}\mathrm{E}|e_1|^{2+\delta}) = o((N\pi_{11})^{(2+\delta)/2}). \tag{2.99}$$

It follows from (2.99) and Chebyshev's inequality that for any $\varepsilon > 0$,

$$P\Big(\Big|\sum_{i=1}^{N}\frac{e_i^2}{N\pi_{11}} - 1\Big| > \varepsilon\Big) \leq (N\pi_{11}\varepsilon)^{-(2+\delta)/2}\mathrm{E}|\sum_{i=1}^{N}(e_i^2 - \pi_{11})|^{(2+\delta)/2} = o(1), \tag{2.100}$$

which implies the first equation of (2.90).

Since $\mathrm{E}(e_i v_i) = 0$, similar to (2.97) and (2.98), we have

$$\begin{aligned}\mathrm{E}|\sum_{i=1}^{N}e_i v_i|^{(2+\delta)/2} &= O(N^{(2+\max\{2,\delta\})/4}\mathrm{E}|e_1 v_1|^{(2+\delta)/2}) \\ &= O\left(N^{(2+\max\{2,\delta\})/4}(\mathrm{E}|e_1|^{2+\delta})^{1/2}(\mathrm{E}|e_1|^{2+\delta})^{1/2}\right),\end{aligned}$$

which implies the first equation of (2.92) by using the Chebyshev's inequality. Equation (2.93) follows from (2.96) by letting $c = 0, d = 1$ or $c = 1, d = 0$. □

Lemma 2.4.8. *Under conditions of Theorem 2.4.3, we have*

$$\mathrm{E}e_1^4/(\mathrm{E}e_1^2)^2 = O(1) \quad \textit{and} \quad \mathrm{E}v_1^4/(\mathrm{E}v_1^2)^2 = O(1).$$

Proof. Without loss of generality we assume $\mu = 0$. Since $\sigma_{ij} = \mathbb{E}(X_{1i}X_{1j})$, we have

$$\begin{aligned}\mathrm{E}e_1^4 &= \mathrm{E}\Big\{\sum_{i=1}^{p}\sum_{j=1}^{p}(X_{1i}X_{1j} - \sigma_{ij})(X_{N+1,i}X_{N+1,j} - \sigma_{ij})\Big\}^4 \\ &\leq 256\mathrm{E}\Big\{\sum_{i=1}^{p}\sum_{j=1}^{p}X_{1i}X_{1j}X_{N+1,i}X_{N+1,j}\Big\}^4 \\ &= 256\sum_{1\leq i_1,\ldots,i_8\leq p}\Big\{\mathrm{E}(\prod_{l=1}^{8}X_{1i_l})\Big\}^2 =: 256\sum_{1\leq i_1,\ldots,i_8\leq p}[T(i_1,\ldots,i_8)]^2.\end{aligned}$$

Write $\Gamma = (\gamma_{ij})_{p\times m}$ and $U_i = Z_{1i}$ for $i = 1, \ldots, m$. Then $\mathrm{var}(X_1) = \Sigma = (\sigma_{ij})_{p\times p} = \Gamma\Gamma^T$ and it follows from (B') that

$$X_{1i} = \sum_{j=1}^{m} \gamma_{ij} U_j, \quad i = 1, \ldots, p,$$

and

$$T(i_1, \ldots, i_8) = \mathbb{E}\left(\prod_{l=1}^{8}\left(\sum_{j=1}^{m} \gamma_{i_l j} U_j\right)\right).$$

By (B'), we know that

$$\mathbb{E}(\prod_{i=1}^{8} U_{a_i}) \neq 0$$

implies each value of a_i appears at least twice in the sequence $a_1, \ldots, a_8$. Denote

$$B_l = \{(a_1, \ldots, a_l) : 1 \leq a_1, \ldots, a_l \leq m\},$$

and let S_k be the set of k-permutations. Then

$$\begin{aligned} & T(i_1, \ldots, i_8) \\ = & \sum_{(a,b,c,d)\in B_4} \sum_{\substack{(k_1,\ldots,k_8)=\sigma(i_1,\ldots,i_8),\\ \sigma\in S_8}} \gamma_{k_1 a}\gamma_{k_2 a}\gamma_{k_3 b}\gamma_{k_4 b}\gamma_{k_5 c}\gamma_{k_6 c}\gamma_{k_7 d}\gamma_{k_8 d}\mathrm{E}(U_a^2 U_b^2 U_c^2 U_d^2) \\ & + \sum_{\substack{(a,b,c)\in B_3,\\ a\neq b}} \sum_{\substack{(k_1,\ldots,k_8)=\sigma(i_1,\ldots,i_8),\\ \sigma\in S_8}} \gamma_{k_1 a}\gamma_{k_2 a}\gamma_{k_3 a}\gamma_{k_4 b}\gamma_{k_5 b}\gamma_{k_6 b}\gamma_{k_7 c}\gamma_{k_8 c}\mathrm{E}(U_a^3 U_b^3 U_c^2) \\ = & T_1(i_1, \ldots, i_8) + T_2(i_1, \ldots, i_8). \end{aligned}$$

In the following we denote $\Lambda = \Gamma^T\Gamma = (\lambda_{ij})_{m\times m}$ and let L be the uniform bound of $\mathbb{E}(U_1^8)$. Note that the summation $\sum_{(k_1,\ldots,k_8)=\sigma(i_1,\ldots,i_8),\ \sigma\in S_8}$ consists of at most $8!$ terms, and for each choice of $\{k_1, \ldots, k_8\}$ (for example, $k_1 = i_1, \ldots, k_8 = i_8$), we get the same value of

$$\sum_{1\leq i_1,\ldots,i_8\leq p}\left(\sum_{(a,b,c,d)\in B_4} \gamma_{k_1 a}\gamma_{k_2 a}\gamma_{k_3 b}\gamma_{k_4 b}\gamma_{k_5 c}\gamma_{k_6 c}\gamma_{k_7 d}\gamma_{k_8 d}\mathrm{E}(U_a^2 U_b^2 U_c^2 U_d^2)\right)^2.$$

Hence

$$
\begin{aligned}
& \sum_{1\le i_1,\dots,i_8\le p} T_1^2(i_1,\dots,i_8) \\
\le\ & (8!)^2 \sum_{1\le i_1,\dots,i_8\le p} \sum_{(a,b,c,d)\in B_4} \sum_{(a',b',c',d')\in B_4} \gamma_{i_1a}\gamma_{i_1a'}\gamma_{i_2a}\gamma_{i_2a'}\cdots\gamma_{i_7d}\gamma_{i_7d'}\gamma_{i_8d}\gamma_{i_8d'} \\
& \times \mathrm{E}(U_a^2U_b^2U_c^2U_d^2)\mathrm{E}(U_{a'}^2U_{b'}^2U_{c'}^2U_{d'}^2) \\
=\ & (8!)^2 \sum_{(a,b,c,d)\in B_4} \sum_{(a',b',c',d')\in B_4} (\lambda_{aa'}\lambda_{bb'}\lambda_{cc'}\lambda_{dd'})^2 \mathrm{E}(U_a^2U_b^2U_c^2U_d^2)\mathrm{E}(U_{a'}^2U_{b'}^2U_{c'}^2U_{d'}^2) \\
\le\ & (8!)^2L^2 \sum_{(a,b,c,d)\in B_4} \sum_{(a',b',c',d')\in B_4} (\lambda_{aa'}\lambda_{bb'}\lambda_{cc'}\lambda_{dd'})^2 \\
=\ & O(1)\Big(\sum_{1\le a,a'\le m} \lambda_{aa'}^2\Big)^4 \\
=\ & O((\mathrm{tr}(\Lambda^2))^4) = O((\mathrm{tr}(\Sigma^2))^4). \qquad (2.101)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
& \sum_{1\le i_1,\dots,i_8\le p} T_2^2(i_1,\dots,i_8) \\
\le\ & O(1) \sum_{1\le i_1,\dots,i_8\le p} \sum_{(a,b,c)\in B_3} \sum_{(a',b',c')\in B_3} \gamma_{i_1a}\gamma_{i_1a'}\gamma_{i_2a}\gamma_{i_2a'}\gamma_{i_3a}\gamma_{i_3a'}\cdots\gamma_{i_8c}\gamma_{i_8c'} \\
=\ & O(1) \sum_{(a,b,c)\in B_3} \sum_{(a',b',c')\in B_3} \lambda_{aa'}^3\lambda_{bb'}^3\lambda_{cc'}^2 \\
=\ & O\Big(\Big(\sum_{1\le c,c'\le m} \lambda_{cc'}^2\Big)\Big(\sum_{1\le a,a',b,b'\le m} |\lambda_{aa'}\lambda_{bb'}|^3\Big)\Big) \\
=\ & O(\mathrm{tr}(\Lambda^2))O\Big(\Big(\sum_{1\le a,a',b,b'\le m} \lambda_{aa'}^2\lambda_{bb'}^2(\lambda_{aa'}^2+\lambda_{bb'}^2)\Big)\Big) \\
=\ & O(\mathrm{tr}(\Lambda^2))O\Big(\Big(\sum_{1\le a,a'\le m} \lambda_{aa'}^4\Big)\Big(\sum_{1\le b,b'\le m} \lambda_{bb'}^2\Big)\Big) \\
=\ & O(\mathrm{tr}(\Lambda^2)(\mathrm{tr}(\Lambda^2))^3) = O((\mathrm{tr}(\Sigma^2))^4). \qquad (2.102)
\end{aligned}
$$

Thus by (2.101) and (2.102),

$$
\mathbb{E}e_1^4 \le 256 \sum_{1\le i_1,\dots,i_8\le p} T^2(i_1,\dots,i_8) = O((\mathrm{tr}(\Sigma^2))^4). \qquad (2.103)
$$

On the other hand, let $V_i = Z_{N+1,i}$ for $i = 1,\dots,m$ and then

$$
\mathrm{E}e_1^2 \ = \ \mathrm{E}\left(\sum_{a=1}^m\sum_{b=1}^m\sum_{c=1}^m\sum_{d=1}^m\sum_{i=1}^p\sum_{j=1}^p \gamma_{ia}\gamma_{jb}\gamma_{ic}\gamma_{jd}(U_aU_b-\delta_{ab})(V_cV_d-\delta_{cd})\right)^2
$$

Denote $C = \min\{\mathbb{E}(U_1^2 - 1)^2, 1\} > 0$. Note that if $c \neq a$ or $d \neq b$,

$$\mathbb{E}((U_aU_b - \delta_{ab})(U_cU_d - \delta_{cd})) = \mathbb{E}((V_aV_b - \delta_{ab})(V_cV_d - \delta_{cd})) = 0.$$

Thus,

$$\begin{aligned}
\mathrm{E}e_1^2 &= \mathrm{E}\left[\sum_{a,b,c,d=1}^{m}\left(\sum_{1\le i,j\le p}\gamma_{ia}\gamma_{jb}\gamma_{ic}\gamma_{jd}\right)^2(U_aU_b-\delta_{ab})^2(V_cV_d-\delta_{cd})^2\right] \\
&\geq \sum_{a,b,c,d=1}^{m}\left(\sum_{1\le i,j\le p}\gamma_{ia}\gamma_{jb}\gamma_{ic}\gamma_{jd}\right)^2 C^2 \\
&= C^2\sum_{i_1=1}^{p}\sum_{j_1=1}^{p}\sum_{i_2=1}^{p}\sum_{j_2=1}^{p}(\sigma_{i_1i_2}\sigma_{j_1j_2})^2 \\
&= C^2(\mathrm{tr}(\Sigma^2))^2. \qquad (2.104)
\end{aligned}$$

Therefore,

$$\mathrm{E}e_1^4/(\mathrm{E}e_1^2)^2 = O(1).$$

Next, we show that $\mathrm{E}v_1^4 = O((\mathrm{E}v_1^2)^2)$. It is easy to verify that

$$\begin{aligned}
\mathrm{E}v_1^4 &= \mathrm{E}\Big(\sum_{i=1}^{p}\sum_{j=1}^{p}(X_{1i}X_{1j} + X_{N+1,i}X_{N+1,j} - 2\sigma_{ij})\Big)^4 \\
&\leq 16\mathrm{E}\left(\sum_{i=1}^{p}\sum_{j=1}^{p}(X_{1i}X_{1j}-\sigma_{ij})\right)^4 \\
&\leq 256\mathbb{E}\left(\sum_{i=1}^{p}\sum_{j=1}^{p}X_{1i}X_{1j}\right)^4 \\
&\leq 256\sum_{1\le i_1,\ldots,i_8\le p}T(i_1,\ldots,i_8).
\end{aligned}$$

Similar to (2.101), we can show that

$$\begin{aligned}
&\sum_{1\le i_1,\ldots,i_8\le p} T_1(i_1,\ldots,i_8)\\
\le\ & 8!\sum_{1\le i_1,\ldots,i_8\le p}\ \sum_{(a,b,c,d)\in B_4}\gamma_{i_1a}\gamma_{i_2a}\cdots\gamma_{i_7d}\gamma_{i_8d}\mathrm{E}(U_a^2U_b^2U_c^2U_d^2)\\
=\ & 8!\sum_{(a,b,c,d)\in B_4}\left(\sum_{i=1}^p\lambda_{ia}\right)^2\left(\sum_{i=1}^p\lambda_{ib}\right)^2\left(\sum_{i=1}^p\lambda_{ic}\right)^2\left(\sum_{i=1}^p\lambda_{id}\right)^2\mathrm{E}(U_a^2U_b^2U_c^2U_d^2)\\
\le\ & (8!)L\left(\sum_{a=1}^m\left(\sum_{i=1}^p\lambda_{ia}\right)^2\right)^4\\
=\ & O(1)\Big(\sum_{1\le i,j\le p}\sigma_{ij}\Big)^4
\end{aligned}$$

For $T_2(i_1,\ldots,i_8)$, we have

$$\begin{aligned}
&\sum_{1\le i_1,\ldots,i_8\le p} T_2(i_1,\ldots,i_8)\\
=\ & O(1)\sum_{\substack{(a,b,c)\in B_3,\\ a\ne b}}(\sum_{i=1}^p\gamma_{ia})^3(\sum_{j=1}^p\gamma_{jb})^3(\sum_{k=1}^p\gamma_{kc})^2\\
\le\ & O(1)\Big[\sum_{a=1}^m\sum_{b=1}^m(\sum_{i=1}^p\gamma_{ia})^2(\sum_{j=1}^p\gamma_{jb})^2\Big((\sum_{i=1}^p\gamma_{ia})^2+(\sum_{j=1}^p\gamma_{jb})^2\Big)\Big]\Big(\sum_{c=1}^m(\sum_{k=1}^p\gamma_{kc})^2\Big)\\
=\ & O\Big[\Big(\sum_{a=1}^m(\sum_{i=1}^p\gamma_{ia})^2\Big)^4\Big]=O(1)\Big(\sum_{1\le i,j\le p}\sigma_{ij}\Big)^4.
\end{aligned}$$

Therefore,

$$\mathrm{E}v_1^4=O\left(\Big(\sum_{1\le i,j\le p}\sigma_{ij}\Big)^4\right). \tag{2.105}$$

On the other hand, similar to (2.104), we have

$$\begin{aligned}
\mathbb{E}v_1^2\ \ge\ & \sum_{1\le a,b\le m}\Big(\sum_{i=1}^p\sum_{j=1}^p\gamma_{ia}\gamma_{jb}\Big)^2\mathrm{E}(U_aU_b-\delta_{ab})^2\\
\ge\ & C\Big(\sum_{a=1}^m\sum_{i=1}^p\sum_{j=1}^p\gamma_{ia}\gamma_{ja}\Big)^2\\
=\ & C\Big(\sum_{i=1}^p\sum_{j=1}^p\sigma_{ij}\Big)^2. \qquad (2.106)
\end{aligned}$$

Thus, by the condition $\sum_{1\le i,j\le p}\sigma_{ij}>0$, (2.105) and (2.106), we have

$$\mathbb{E}v_1^4=O((\mathbb{E}v_1^2)^2).$$

This completes the proof of Lemma 4.3. □

Lemma 2.4.9. *Under the assumptions of Theorem 2.4.4, we have*

$$\mathrm{E}e_1'^4/(\mathrm{E}e_1'^2)^2=O(1)\quad and\quad \mathrm{E}v_1'^4/(\mathrm{E}v_1'^2)^2=O(1), \tag{2.107}$$

where $e_1'=e_1'(\boldsymbol{a}_0)$ *and* $v_1'=v_1'(\boldsymbol{a}_0)$ *as defined in* (2.81) *and* (2.82).

Proof. By Isserlis' theorem, we have

$$\begin{aligned}
\mathrm{E}e_1'^4 &= \mathrm{E}\Big\{\sum_{i=1}^{p-\tau}\sum_{j=i+\tau}^{p}X_{1i}X_{1j}X_{(N+1)i}X_{(N+1)j}\Big\}^4\\
&= \sum_{1\le i_1,\dots,i_8\le p}\Big\{\mathrm{E}(\prod_{l=1}^{8}X_{1i_l})\Big\}^2\\
&= \frac{1}{8*6*4*2}\sum_{1\le i_1,\dots,i_8\le p}\sum_{\substack{(k_1,\dots,k_8)=\sigma(i_1,\dots,i_8),\\ \sigma\in S_8}}(\sigma_{k_1k_2}\sigma_{k_3k_4}\sigma_{k_5k_6}\sigma_{k_7k_8})^2\\
&= 7*5*3\sum_{1\le k_1,\dots,k_8\le p}\sigma_{k_1k_2}^2\sigma_{k_3k_4}^2\sigma_{k_5k_6}^2\sigma_{k_7k_8}^2=O((\mathrm{tr}(\Sigma^2))^4).
\end{aligned}$$

On the other hand, by Isserlis' theorem again, we have

$$\begin{aligned}
\mathrm{E}e_1'^2 &= \sum_{i_1=1}^{p-\tau}\sum_{j_1=i_1+\tau}^{p}\cdots\sum_{i_4=1}^{p}\sum_{j_4=i_4+\tau}^{p}\left\{\mathbb{E}(\prod_{l=1}^{4}X_{1i_l}X_{1j_l})\right\}^2\\
&= \sum_{i_1=1}^{p-\tau}\sum_{j_1=i_1+\tau}^{p}\sum_{i_2=1}^{p-\tau}\sum_{j_2=i_2+\tau}^{p}(\sigma_{i_1i_2}\sigma_{j_1j_2}+\sigma_{i_1j_2}\sigma_{i_2j_1}+\sigma_{i_1j_1}\sigma_{i_2j_2})^2\\
&= \sum_{i_1=1}^{p-\tau}\sum_{j_1=i_1+\tau}^{p}\sum_{i_2=1}^{p-\tau}\sum_{j_2=i_2+\tau}^{p}(\sigma_{i_1i_2}\sigma_{j_1j_2})^2\\
&= \frac{1}{4}\sum_{i_1=1}^{p}\sum_{j_1:|j_1-i_1|\ge\tau}\sum_{i_2=1}^{p}\sum_{j_2:|j_2-i_2|\ge\tau}(\sigma_{i_1i_2}\sigma_{j_1j_2})^2\\
&= \frac{1}{4}\sum_{i_1,i_2:|i_1-i_2|\le\tau}\left(\sigma_{i_1i_2}^2\sum_{j_1,j_2:|j_1-i_1|\ge\tau,|j_2-i_2|\ge\tau}\sigma_{j_1j_2}^2\right).
\end{aligned}$$

Here, note that when $|i_1 - i_2| \le \tau$,

$$\begin{aligned}\{(j_1, j_2) : |j_1 - j_2| \le \tau\} \quad \subset \quad & \{(j_1, j_2) : |j_1 - i_1| \ge \tau, |j_2 - i_2| \ge \tau\} \\ & \cup\{(j_1, j_2) : |j_1 - i_1| \le 3\tau, |j_2 - i_2| \le 3\tau\}.\end{aligned}$$

Thus

$$\begin{aligned}\mathrm{E}e_1'^2 &\ge \frac{1}{4}\sum_{i_1,i_2:|i_1-i_2|\le\tau}\left(\sigma_{i_1i_2}^2\left(\sum_{j_1,j_2:|j_1-j_2|\le\tau}\sigma_{j_1j_2}^2 - \sum_{j_1,j_2:|j_1-i_1|\le3\tau,|j_2-i_2|\le3\tau}\sigma_{j_1j_2}^2\right)\right) \\ &\ge \frac{1}{4}\sum_{i_1,i_2=1}^{p}\left(\sigma_{i_1i_2}^2\left(\sum_{j_1,j_2=1}^{p}\sigma_{j_1j_2}^2 - 36\tau^2C_2^2\right)\right) \\ &= \frac{1}{4}\mathrm{tr}(\Sigma^2)(\mathrm{tr}(\Sigma^2) - 36\tau^2C_2^2). \qquad (2.108)\end{aligned}$$

Since $\tau = o((\sum_{1\le i,j\le p}\sigma_{ij}^2)^{1/2})$, we have $(\mathrm{E}e_1'^2)^{-1} = O((\mathrm{tr}(\Sigma^2))^{-2})$, which implies that $\mathrm{E}e_1'^4/(\mathrm{E}e_1'^2)^2 = O(1)$, i.e., the first equality in (2.107) holds.

Next we prove the second half of (2.107). Note that

$$\begin{aligned}\mathrm{E}v_1'^4 &= \mathrm{E}\Big(\sum_{i=1}^{p}\sum_{j=i+\tau}^{p}(X_{1i}X_{1j} + X_{N+1,i}X_{N+1,j})\Big)^4 \\ &\le 16\mathrm{E}\left(\sum_{i=1}^{p}\sum_{j=1}^{p}X_{1i}X_{1j} - \sum_{i=1}^{p}\sum_{|j-i|\le\tau}X_{1i}X_{1j}\right)^4 \\ &\le 256\mathrm{E}\left(\sum_{i=1}^{p}\sum_{j=1}^{p}X_{1i}X_{1j}\right)^4 + 256\mathrm{E}\left(\sum_{i=1}^{p}\sum_{|j-i|\le\tau}X_{1i}X_{1j}\right)^4 \\ &=: 256(T_1' + T_2'). \qquad (2.109)\end{aligned}$$

By (2.105), we have

$$T_1' = O\left(\left(\sum_{i_1=1}^{p}\sum_{i_2=1}^{p}\sigma_{i_1i_2}\right)^4\right). \qquad (2.110)$$

On the other hand, it follows from Isserlis' theorem that

$$
\begin{aligned}
& T_2' \\
= & \sum_{1\le i_1,i_2,i_3,i_4\le p} \sum_{|j_l-i_l|\le\tau, l=1,\ldots,4} \mathrm{E}\left(\prod_{l=1}^{4} X_{1i_l}X_{1j_l}\right) \\
= & \frac{1}{8*6*4*2} \sum_{1\le i_1,i_2,i_3,i_4\le p} \sum_{|j_l-i_l|\le\tau, l=1,\ldots,4} \sum_{\substack{(p_1,q_1,\ldots,p_4,q_4)=\sigma(i_1,j_1,\ldots,i_4,j_4), \\ \sigma\in S_8,\, |p_l-q_l|\le\tau, l=1,\ldots,4}} \sigma_{p_1q_1}\sigma_{p_2q_2}\sigma_{p_3q_3}\sigma_{p_4q_4}
\end{aligned}
\tag{2.111}
$$

For a given permutation $\{p_1, q_1, \ldots, p_4, q_4\}$ of $\{i_1, j_1, \ldots, i_4, j_4\}$, the number of common elements in the two sets $\{|p_1 - q_1|, \ldots, |p_4 - q_4|\}$ and $\{|i_1 - j_1|, \ldots, |i_4 - j_4|\}$ can be four, two, one and zero. Next we analyze each case.

(i) When there are four common elements, we have

$$
\begin{aligned}
\sum_{1\le i_1,i_2,i_3,i_4\le p} \sum_{|j_l-i_l|\le\tau, l=1,2,3,4} \sigma_{p_1q_1}\sigma_{p_2q_2}\sigma_{p_3q_3}\sigma_{p_4q_4} &= \left(\sum_{1\le i\le p}\sum_{|i-j|\le\tau}\sigma_{ij}\right)^4 \\
&= \left(\sum_{i=1}^{p}\sum_{j=1}^{p}\sigma_{ij}\right)^4 .
\end{aligned}
$$

(ii) When there are two common elements, without loss of generality, we assume $|p_1 - q_1| = |i_1 - j_1|, |p_2 - q_2| = |i_2 - j_2|, p_3 = i_3, q_3 = j_4, p_4 = i_4, q_4 = j_3$. Other possibilities can be shown in the same way. Use the fact that $|\sigma_{j_1j_2}| \le$

$\sqrt{\sigma_{j_1j_1}\sigma_{j_2j_2}} \le C_2$ for all $1 \le j_1, j_2 \le p$, we have

$$\begin{aligned}
&|\sum_{1\le i_1,i_2,i_3,i_4\le p}\sum_{|j_l-i_l|\le\tau,l=1,2,3,4}\sigma_{p_1q_1}\sigma_{p_2q_2}\sigma_{p_3q_3}\sigma_{p_4q_4}| \\
=& \left(\sum_{i=1}^{p}\sum_{j=1}^{p}\sigma_{ij}\right)^2 |\sum_{1\le i_3,i_4\le p}\sum_{|i_3-j_3|\le\tau,|i_4-j_4|\le\tau}\sigma_{p_3,q_3}\sigma_{p_4,q_4}| \\
=& \left(\sum_{i=1}^{p}\sum_{j=1}^{p}\sigma_{ij}\right)^2 |\sum_{1\le i_3,i_4\le p}\sum_{|i_3-j_3|\le\tau,|i_4-j_4|\le\tau,|i_3-j_4|\le\tau,|i_4-j_3|\le\tau}\sigma_{i_3,j_4}\sigma_{i_4,j_3}| \\
\le& \left(\sum_{i=1}^{p}\sum_{j=1}^{p}\sigma_{ij}\right)^2 \sum_{i_3=1}^{p}\sum_{j_4=1}^{p}|\sigma_{i_3,j_4}|\sum_{|i_4-i_3|\le 4\tau,|j_3-i_3|\le 4\tau}|\sigma_{i_4,j_3}| \\
\le& \left(\sum_{i=1}^{p}\sum_{j=1}^{p}\sigma_{ij}\right)^2 \sum_{i_3=1}^{p}\sum_{j_4=1}^{p}|\sigma_{i_3,j_4}|(8\tau)^2C_2 \\
=& o\left((\sum_{i=1}^{p}\sum_{j=1}^{p}\sigma_{ij})^4\right).
\end{aligned}$$

(iii) When there is one common element, using similar arguments as above, we can show that

$$\begin{aligned}
\sum_{1\le i_1,i_2,i_3,i_4\le p}\sum_{|j_l-i_l|\le\tau,l=1,2,3,4}\sigma_{p_1q_1}\sigma_{p_2q_2}\sigma_{p_3q_3}\sigma_{p_4q_4} &= O\left((\sum_{i=1}^{p}\sum_{j=1}^{p}\sigma_{ij})(\sum_{i=1}^{p}\sum_{j=1}^{p}|\sigma_{ij}|)\tau^4\right) \\
&= o\left((\sum_{i=1}^{p}\sum_{j=1}^{p}\sigma_{ij})^4\right).
\end{aligned}$$

(iv) When there is no common element, similarly we can show that

$$\begin{aligned}
\sum_{1\le i_1,i_2,i_3,i_4\le p}\sum_{|j_l-i_l|\le\tau,l=1,2,3,4}\sigma_{p_1q_1}\sigma_{p_2q_2}\sigma_{p_3q_3}\sigma_{p_4q_4} &= O\left((\tau^2\sum_{i=1}^{p}\sum_{j=1}^{p}|\sigma_{ij}|)^2\right) \\
&= o\left((\sum_{i=1}^{p}\sum_{j=1}^{p}\sigma_{ij})^4\right).
\end{aligned}$$

Hence, it follows from the above results that

$$T_2' = O\left(\left(\sum_{i=1}^{p}\sum_{j=1}^{p}\sigma_{ij}\right)^4\right). \tag{2.112}$$

On the other hand, we have

$$
\begin{aligned}
& \mathrm{E}v_1'^2 \\
= \;& \mathrm{E}\Big(\sum_{i_1=1}^{p-\tau}\sum_{j_1=i_1+\tau}^{p}\sum_{i_2=1}^{p-\tau}\sum_{j_2=i_2+\tau}^{p}(X_{1i}X_{1j}+X_{N+1,i}X_{N+1,j})\Big)^2 \\
= \;& 2\sum_{i_1=1}^{p-\tau}\sum_{j_1=i_1+\tau}^{p}\sum_{i_2=1}^{p-\tau}\sum_{j_2=i_2+\tau}^{p}\mathbb{E}\left(X_{1i_1}X_{1j_1}X_{1i_2}X_{1j_2}\right) \\
= \;& 2\sum_{i_1=1}^{p-\tau}\sum_{j_1=i_1+\tau}^{p}\sum_{i_2=1}^{p-\tau}\sum_{j_2=i_2+\tau}^{p}\left(\sigma_{i_1i_2}\sigma_{j_1j_2}+\sigma_{i_1j_1}\sigma_{i_2j_2}+\sigma_{i_1j_2}\sigma_{i_2j_1}\right) \\
= \;& \frac{1}{2}\sum_{i_1=1}^{p}\sum_{|j_1-i_1|\ge\tau}\sum_{i_2=1}^{p}\sum_{|j_2-i_2|\ge\tau}\sigma_{i_1i_2}\sigma_{j_1j_2} \\
= \;& \frac{1}{2}\sum_{i_1=1}^{p}\sum_{i_2=1}^{p}\sigma_{i_1i_2}\left[\left(\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}-\sum_{j_2=1}^{p}\sum_{|j_1-i_1|<\tau}-\sum_{j_1=1}^{p}\sum_{|j_2-i_2|<\tau}-\sum_{|j_1-i_1|\le\tau}\sum_{|j_2-i_2|<\tau}\right)\sigma_{j_1j_2}\right] \\
= \;& \frac{1}{2}\sum_{i_1=1}^{p}\sum_{i_2=1}^{p}\sigma_{i_1i_2}\left[\left(\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}-\sum_{|j_1-j_2|\le\tau}\sum_{|j_1-i_1|<\tau}-\sum_{|j_1-j_2|<\tau}\sum_{|j_2-i_2|<\tau}-\sum_{|j_1-i_1|\le\tau}\sum_{|j_2-i_2|<\tau}\right)\sigma_{j_1j_2}\right] \\
\ge \;& \frac{1}{2}\left(\sum_{i_1=1}^{p}\sum_{i_2=1}^{p}\sigma_{i_1i_2}I(\sigma_{i_1i_2}\ge 0)\right)\left(\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}\sigma_{j_1j_2}-20\tau^2C_2\right) \\
& +\frac{1}{2}\left(\sum_{i_1=1}^{p}\sum_{i_2=1}^{p}\sigma_{i_1i_2}I(\sigma_{i_1i_2}<0)\right)\left(\sum_{j_1=1}^{p}\sum_{j_2=1}^{p}\sigma_{j_1j_2}+20\tau^2C_2\right) \\
= \;& \frac{1}{2}\left(\sum_{i_1=1}^{p}\sum_{i_2=1}^{p}\sigma_{i_1i_2}\right)^2-20\tau^2C_2\sum_{i_1=1}^{p}\sum_{i_2=1}^{p}|\sigma_{i_1i_2}|.
\end{aligned}
$$

Since $\tau = o(\min\{(\sum_{1\le i,j\le p}\sigma_{ij})/(\sum_{1\le i,j\le p}|\sigma_{ij}|)^{1/2}\})$, it follows that

$$
(\mathrm{E}v_1'^2)^{-1} = O\left(\left(\sum_{i_1=1}^{p}\sum_{i_2=1}^{p}\sigma_{i_1i_2}\right)^{-2}\right). \tag{2.113}
$$

Therefore, the second equality in (2.107) follows from (2.109)–(2.113). □

Lemma 2.4.10. *Under the assumptions of Theorem 2.4.5, we have*

$$
\mathrm{E}e_1'^4/(\mathrm{E}e_1'^2)^2 = O(1) \quad \textit{and} \quad \mathrm{E}\tilde{v}_1'^4/(\mathrm{E}\tilde{v}_1'^2)^2 = O(1), \tag{2.114}
$$

where $e_1' = e_1'(\boldsymbol{a}_0)$ *and* $\tilde{v}_1' = \tilde{v}_1'(\boldsymbol{a}_0)$ *are defined in* (2.81) *and* (2.84).

Proof. The first equality follows from the proof of Lemma 2.4.9. To show the second half of (2.107), let $g_1 = \sum_{i=1}^{t} X_{1i} \sum_{j=t+\tau}^{p} X_{1j}$ and $g_2 = \sum_{i=1}^{t} X_{N+1,i} \sum_{j=t+\tau}^{p} X_{N+1,j}$. Then we have that $\tilde{v}_1' = g_1 + g_2$, $\mathbb{E}(\tilde{v}_1'^4) \le 16\mathbb{E}(g_1^4)$ and $\mathbb{E}(\tilde{v}_1'^2) = 2\mathbb{E}(g_1^2)$.

Note that g_1 is the product of two Gaussian random variables from a multivariate Gaussian vector with mean zero. Hence we can write $g_1 = A(A+B)$ where A, B are independent Gaussian random variables with mean zero, variance a^2 and b^2 respectively. It follows that

$$\begin{aligned}
\mathbb{E}g_1^4 &\le 16(\mathbb{E}(A^8) + \mathbb{E}(A^4B^4)) \\
&= 16(105a^8 + 9a^4b^4) \\
&\le 200(9a^8 + a^4b^4) \\
&\le 200(3a^4 + a^2b^2)^2 = 200(\mathbb{E}g_1^2)^2 = 100(\mathbb{E}\tilde{v}_1'^2)^2.
\end{aligned}$$

This completes the proof of the lemma. □

Proof of Theorem 2.4.1. It follows from Lemma 2.4.7 and the standard arguments of empirical likelihood method. □

Proof of Corollary 2.4.2. It follows from the argument in the proof of Corollary 2.2.2. □

Proof of Theorem 2.4.3. It follows from Lemma 2.4.8 that (2.79) in Theorem 2.4.1 holds with $\delta = 2$. Hence Theorem 2.4.3 follows from Theorem 2.4.1. □

Proof of Theorem 2.4.4. It follows from Lemma 2.4.9 that (2.79) holds for random sequence $\{e_i'\}$ and $\{v_i'\}$ with $\delta = 2$. Hence Theorem 2.4.4 follows from the same arguments as in the proof of Theorem 2.4.1. □

Proof of Theorem 2.4.5. It follows from Lemma 2.4.10 that (2.79) holds for random sequence $\{e_i'\}$ and $\{\tilde{v}_i'\}$ with $\delta = 2$. Hence Theorem 2.4.5 follows from the same arguments as in the proof of Theorem 2.4.1. □

Proof of Theorem 2.4.6. Note that under the alternative hypothesis H_1, $\mathrm{E}Y_1 = \boldsymbol{a}$ and for $1 \le i \le N$,

$$\begin{aligned} e_i(\boldsymbol{a}_0) &= e_i(\boldsymbol{a}) + (\boldsymbol{a}-\boldsymbol{a}_0)^T(\boldsymbol{a}-\boldsymbol{a}_0) + (\boldsymbol{a}-\boldsymbol{a}_0)^T(Y_i + Y_{N+i} - 2\boldsymbol{a}) \\ &= e_i(\boldsymbol{a}) + (\boldsymbol{a}-\boldsymbol{a}_0)^T(\boldsymbol{a}-\boldsymbol{a}_0) + (\boldsymbol{a}-\boldsymbol{a}_0)^T(Y_i + Y_{N+i} - 2\boldsymbol{a}) \\ v_i(\boldsymbol{a}_0) &= v_i(\boldsymbol{a}) + 2\boldsymbol{1}_q^T(\boldsymbol{a}-\boldsymbol{a}_0) = v_i(\boldsymbol{a}) + 2\boldsymbol{1}_q^T(\boldsymbol{a}-\boldsymbol{a}_0), \end{aligned}$$

where $q = p^2$. As a result, we have

$$\Big(\frac{e_i(\boldsymbol{a}_0)}{\sqrt{\pi_{11}}}, \frac{v_i(\boldsymbol{a}_0)}{\sqrt{\pi_{22}}}\Big)^T = \Big(\frac{e_i(\boldsymbol{a})}{\sqrt{\pi_{11}}}, \frac{v_i(\boldsymbol{a})}{\sqrt{\pi_{22}}}\Big)^T + \Big(\zeta_{n1}, \zeta_{n2}\Big)^T + \Big(\eta_i(\boldsymbol{a}), 0\Big)^T,$$

where $\eta_i(\boldsymbol{a}) = (\boldsymbol{a}-\boldsymbol{a}_0)^T(Y_i + Y_{N+i} - 2\boldsymbol{a})/\sqrt{\pi_{11}}$. Since

$$\begin{aligned} \mathrm{E}[\sum_{i=1}^N \eta_i(\boldsymbol{a})]^2 &= 4N(\boldsymbol{a}-\boldsymbol{a}_0)^T\Theta(\boldsymbol{a}-\boldsymbol{a}_0)/\pi_{11} = 4N(\boldsymbol{a}-\boldsymbol{a}_0)^T\Theta(\boldsymbol{a}-\boldsymbol{a}_0)/\mathrm{tr}(\Theta^2) \\ &= O[N(\boldsymbol{a}-\boldsymbol{a}_0)^T(\boldsymbol{a}-\boldsymbol{a}_0)/\sqrt{\pi_{11}}], \end{aligned}$$

it follows from condition (2.86) that

$$\mathrm{E}[\sum_{i=1}^N \eta_i(\boldsymbol{a})]^2 = o(N).$$

Similar to the argument in proving Theorem 3 of Peng, Qi and Wang [78], we see that to prove Theorem 2.4.6, it suffices to show

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N \Big(\frac{e_i(\boldsymbol{a})}{\sqrt{\pi_{11}}}, \frac{v_i(\boldsymbol{a})}{\sqrt{\pi_{22}}}\Big)^T \xrightarrow{d} N(0, I_2). \tag{2.115}$$

which follows directly from Lemma 2.4.7. This completes the proof of Theorem 2.4.6. □

CHAPTER III

JACKKNIFE EMPIRICAL LIKELIHOOD METHODS IN RISK MANAGEMENT

Risk-distortion measures, Spearman's rho and parametric copulas are important quantities in the research of risk management, and the interval estimation for those quantities is known to be challenging. In this chapter, we construct interval estimation for important quantities in these fields: using the jackknife empirical likelihood methods. The content in this chapter is mainly based on the following papers.

1. Peng, L., Qi, Y., Wang, R. and Yang, J. (2012). Jackknife empirical likelihood methods for risk measures and related quantities. *Insurance: Mathematics and Economics*, to appear.
2. Wang, R., Peng, L. and Yang, J. (2012). Jackknife empirical likelihood for parametric copulas. *Scandinavian Actuarial Journal*, to appear.
3. Wang, R. and Peng, L. (2011). Jackknife empirical likelihood intervals for Spearman's rho. *North American Actuarial Journal*, **15**(4), 475-486.

3.1 Introduction

Statistical inference plays an important role in the modern research of actuarial science and risk management. In this chapter, we consider new methods of interval estimation for three different quantities of importance in risk management. We refer to Jones and Zitikis [51], McNeil, Frey and Embrechts [65] and Genest, Ghoudi and Rivest [42] for summary of the statistical inference on risk-distortion measures, Spearman's rho and parametric copulas, respectively. See also the introduction in each of the following sections for more references.

Quantifying risk is always an important topic in actuarial science and risk management. For a given non-negative function ψ, the risk-distortion measure $R(F) = \int_0^1 F^{-}(t)\psi(t)dt$ is used to measure the corresponding risk with a loss distribution F. It is known that the asymptotical variance of the estimation of $R(F)$ is very complicated; see Jones and Zitikis [51] for more details about $R(F)$. Under some regularity conditions (same as in [51]), we find an interval estimation for $R(F)$ in Section 3.2. The functional $R(F)$ is also known as the L-statistics (see Chapter 2 of Shao and Tu [93]). The results also contribute to the study of the asymptotical behavior of the L-statistics.

For dependent risks X and Y with marginal distributions F and G respectively, Spearman's rho $\rho^s = 12\mathbb{E}[(F(X) - 1/2)(G(Y) - 1/2)]$ is one of the most commonly-used non-parametric measures of dependence between risks X and Y. As a measure of dependence, ρ^s is determined by the copula of X and Y. Although ρ^s can be estimated non-parametrically by a natural estimator $\hat{\rho}^s$, the asymptotical variance of $\hat{\rho}^s$ depends on the underlining dependence structure of X and Y and is hard to estimate. Using the jackknife empirical likelihood method, we construct an interval estimation for $R(F)$ without calculating the asymptotic variance in Section 3.3.

As introduced in Chapter I, the analysis of multivariate dependence structures is often dealt with by using copulas. To fit a parametric copula to multivariate data, a popular way is to employ the so-called pseudo maximum likelihood estimation proposed in Genest, Ghoudi and Rivest [42]. However, the asymptotical variance of the above estimator is unavailable except for a few classes of copulas. Under some regularity conditions, we gave a region estimation for the parameter of the copula family in Section 3.4 based on the score equations.

In each section, there are separate subsections of an introduction, the main results, the simulation and proofs.

3.2 Interval Estimation for Risk-distortion Measures

Quantifying risks is of importance in insurance. In this section, we employ the jackknife empirical likelihood method to construct confidence intervals for some risk measures and related quantities studied by Jones and Zitikis [51]. A simulation study shows the advantages of the new method over the normal approximation method and the naive bootstrap method.

3.2.1 Introduction

In life insurance and finance, quantifying risks is a very important task for pricing an insurance product or managing a financial portfolio. Generally speaking, a risk measure is constructed to be a mapping from a set of risks to the set of real numbers. Some well-known risk measures include coherent risk measures (Yaari [111], Artzner [4]), distortion risk measures, Wang's premium principle and proportional hazards transform risk measures; see Wang, Young and Panjer [104]; Wang [100, 101, 102]; Wirch and Hardy [110] and Necir and Meraghni [68] for references.

For a risk variable X with distribution function F, Jones and Zitikis [51] defined a large class of risk measures associated with X as

$$R(F) = \int_0^1 F^-(t)\psi(t)dt, \tag{3.1}$$

where F^- denotes the generalized inverse function of F, and ψ is a nonnegative function chosen for showing the objective opinion about the risk loading. Different choices of ψ result in different risk measures. For example, Tail Value-at-Risk has $\psi(t) = I(t > \alpha)/(1-\alpha)$ with $0 < \alpha < 1$, the proportional hazards transform risk measure has $\psi(t) = r(1-t)^{r-1}$ and Wang's premium principle has $\psi(t) = g'(1-t)$, where g is an increasing convex function with derivatives over $[0, 1]$; see Jones and Zitikis [51] for details. Other choices of the function ψ can be found in Jones and Zitikis [53]. Jones and Zitikis [51] also introduced a related quantity to illustrate the right-tail, left-tail and two-sided deviations, which is defined as

$$r(F) = \frac{R(F)}{\mathbb{E}(X)}. \tag{3.2}$$

Note that the general definition of distortion measures as mentioned in Wang and Young [103] and Wirth and Hardy [110] includes the two widely used risk measures: Value-at-Risk (VaR) and Tail Value-at-Risk (T-VaR). However the class defined by (3.1) excludes the VaR. In this section, we focus on the statistical inference of the risk measure and its related quantity defined in (3.1) and (3.2), respectively.

Statistical inference for $R(F)$ and $r(F)$ plays an important role in the applications of risk measures. Jones and Zitikis [51] proposed nonparametric estimation by replacing F^{-} and $\mathbb{E}(X)$ by the sample quantile function and sample mean respectively, and derived the asymptotic normality. Therefore, confidence intervals for $R(F)$ and $r(F)$ can be constructed via estimating the asymptotic variance. For comparing two risk measures, we refer to Jones and Zitikis [52]. Jones and Zitikis [53] investigated the nonparametric estimation of the parameter associated with distortion-based risk measures.

Because of the complexity of the asymptotic variance of $R(F)$ and $r(F)$, constructing non-parametric confidence intervals via estimating the asymptotic variance is usually inaccurate. In order to construct confidence intervals for $R(F)$ and $r(F)$ without estimating the asymptotic variance, we investigate the possibility of applying an empirical likelihood method in this section so as to improve the inference.

The empirical likelihood method, as introduced in Chapter I, is a nonparametric likelihood approach for statistical inference, which has been shown to be powerful in interval estimation and hypothesis testing. Since the risk measure $R(F)$ and its related quantity $r(F)$ are non-linear functionals, we propose to employ the jackknife empirical likelihood method to obtain interval estimation for these two quantities. Note that for some special risk measures such as VaR and T-VaR one can simply linearized them so that the profile empirical likelihood method can be employed; see

Baysal and Staum [7] for the study of VaR and T-VaR.

The whole section is organized as follows. In Section 3.2.2, the methodologies and main results are presented. A simulation study is given in Section 3.2.3. All proofs are put in Section 3.2.4.

3.2.2 Methodology

Throughout we assume that the observations $X_1, \ldots, X_n$ are independent non-negative random variables with continuous distribution function $F(x)$. Put $\Psi(t) = \int_0^t \psi(s)ds$. When $R(F) < \infty$, we have $t\{\Psi(1) - \Psi(F(t))\} \to 0$ as $t \to \infty$. Thus the risk measure defined in (3.1) can be written as

$$R = R(F) = \int_0^\infty \{\Psi(1) - \Psi(F(t))\}dt.$$

Define the empirical distribution function as $F_n(x) = \frac{1}{n}\sum_{j=1}^n I(X_j \le x)$. Then Jones and Zitikis [51] proposed to estimate $R(F)$ and $r(F)$ by

$$\hat{R}_n = \int_0^\infty (\Psi(1) - \Psi(F_n(t)))dt, \quad \text{and} \quad \hat{r}_n = \frac{n\int_0^\infty (\Psi(1) - \Psi(F_n(t)))dt}{\sum_{j=1}^n X_j},$$

respectively, and showed that

$$\sqrt{n}\{\hat{R}_n - R\} \xrightarrow{d} N(0, \sigma_1^2) \quad \text{and} \quad \sqrt{n}\{\hat{r}_n - r(F)\} \xrightarrow{d} N(0, \sigma_2^2) \tag{3.3}$$

under some regularity conditions, where

$$\sigma_1^2 = Q_F(\Psi, \Psi), \quad \sigma_2^2 = \frac{1}{\mu^2}\left(Q_F(\Psi, \Psi) - 2r(F)Q_F(\Psi, 1) + (r(F))^2 Q_F(1, 1)\right) \tag{3.4}$$

and

$$Q_F(a, b) = \int_0^\infty \int_0^\infty (F(x \wedge y) - F(x)F(y))a(F(x))b(F(y))dxdy,$$

where $a(\cdot), b(\cdot)$ are two functions on $[0, 1]$. Based on (3.3), confidence intervals for $R(F)$ and $r(F)$ can be obtained via estimating σ_1^2 and σ_2^2.

An alternative way to construct confidence intervals is to employ the empirical likelihood method. Since the risk measure R is non-linear, a common technique is to linearize the functional by introducing some link variables before applying the profile empirical likelihood method; see the study for ROC curve (Claeskens, Jing, Peng and Zhou [21]) and copulas (Chen, Peng and Zhao [19]). Unfortunately it remains unknown on how to linearize R by introducing some link variables. Here we propose to apply the jackknife empirical likelihood method developed by Jing, Yuan and Zhou [47]. This procedure is easy to implement and is described as follows.

Define $F_{n,i} = \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} I(X_j \le x)$ and $\hat{R}_{n,i} = \int_0^\infty (\Psi(1) - \Psi(F_{n,i}(t)))dt$ for $i = 1, \ldots, n$. Then the jackknife sample is defined as

$$Y_i = n\hat{R}_n - (n-1)\hat{R}_{n,i}, \qquad i = 1, \ldots, n.$$

Now we apply the empirical likelihood method to the above jackknife sample. That is, define the jackknife empirical likelihood function for $\theta = R(F)$ as

$$L_1(\theta) = \sup\{\prod_{i=1}^{n}(np_i) : \ p_i \ge 0, \text{ for } i = 1, \ldots, n; \ \sum_{i=1}^{n} p_i = 1; \ \sum_{i=1}^{n} p_i Y_i = \theta\}.$$

By Lagrange multiplier technique, we have $p_i = n^{-1}\{1+\lambda(Y_i-\theta)\}^{-1}$ and $-2\log L_1(\theta) = 2\sum_{i=1}^{n}\log\{1+\lambda(Y_i-\theta)\}$, where $\lambda = \lambda(\theta)$ satisfies

$$\sum_{i=1}^{n} \frac{Y_i - \theta}{1+\lambda(Y_i-\theta)} = 0. \tag{3.5}$$

The following theorem shows that Wilks' Theorem holds for the proposed jackknife empirical likelihood method.

Theorem 3.2.1. *Assume that $|\psi(x)| \le cx^{\alpha-1}(1-x)^{\beta-1}$, $\psi'(x)$ exists and $|\psi'(x)| \le cx^{\alpha-2}(1-x)^{\beta-2}$ for all $0 < x < 1$ and some constants $\alpha > 1/2$, $\beta > 1/2$ and $c > 0$. Further assume $\mathbb{E}(|X_i|^\gamma) < \infty$ for some γ such that $\gamma > 1/(\alpha - 1/2)$ and $\gamma > 1/(\beta - 1/2)$. Then we have*

$$-2\log L_1(R_0) \xrightarrow{d} \chi_1^2 \qquad \text{as } n \to \infty,$$

where R_0 denotes the true value of R and χ_1^2 denotes a chi-square distribution with one degree of freedom.

Remark 3.2.1. Some well-known risk measures, such as proportional hazards transform risk measure, Wang's right-tail deviation and Wang's left-tail deviation satisfy the assumptions of Theorem 3.2.1; see Jones and Zitikis [51]. Although the definition of (3.1) includes the widely employed risk measure T-VaR, the assumptions in the Theorem 3.2.1 exclude it.

Remark 3.2.2. Note that when X_i is a real-valued random variable, $t\Psi(F(t)) \to 0$ as $t \to -\infty$ and $t\{\Psi(1) - \Psi(F(t))\} \to 0$ as $t \to \infty$, one can write

$$R = R(F) = \int_0^\infty \{\Psi(1) - \Psi(F(t))\}\,dt + \int_{-\infty}^0 \Psi(F(t))\,dt.$$

Hence a similar jackknife empirical likelihood method can be applied.

Based on the above theorem, a confidence interval for R_0 with level b can be obtained as

$$I_b^R = \{R : -2\log L_1(R) \le \chi^2_{1,b}\},$$

where $\chi^2_{1,b}$ is the b-th quantile of χ^2_1.

Next we consider the related quantity $r(F) = R(F)/\mu$ where $\mu = \mathbb{E}(X_1)$. Alternatively, we consider the quantity $R - \theta\mu$ with $\theta = r(F)$. Then one can estimate this quantity by

$$\hat{R}_n - \theta n^{-1}\sum_{i=1}^n X_i = \hat{R}_n - \theta\int_0^\infty x\,dF_n(x) = \hat{R}_n - \theta\int_0^\infty (1 - F_n(x))\,dx.$$

As before, we define the jackknife sample as

$$n\left(\hat{R}_n - \theta\int_0^\infty x\,dF_n(x)\right) - (n-1)\left(\hat{R}_{n,i} - \theta\int_0^\infty x\,dF_{n,i}(x)\right) = Y_i - \theta X_i$$

for $i = 1, \ldots, n$, where $Y_i's$ are defined as above. So the jackknife empirical likelihood function for $\theta = r(F)$ is defined as

$$L_2(\theta) = \sup\{\prod_{i=1}^n (np_i) : \ p_i \ge 0, \ \text{for } i = 1, \ldots, n; \ \sum_{i=1}^n p_i = 1; \ \sum_{i=1}^n p_i(Y_i - \theta X_i) = 0\}.$$

The following theorem shows that Wilks' Theorem holds for the proposed jackknife empirical likelihood method for $r(F)$.

Theorem 3.2.2. *Assume the conditions in Theorem 3.2.1 hold. Further assume* $\mathbb{E}(X_1^2) < \infty$*. Then*

$$-2 \log L_2(r_0) \xrightarrow{d} \chi_1^2 \qquad as\ n \to \infty,$$

where r_0 *denotes the true value of* $r(F)$*.*

Based on the above theorem, a confidence interval for r_0 with level b can be obtained as

$$I_b^r = \{r : -2 \log L_2(r) \leq \chi_{1,b}^2\}.$$

Remark 3.2.3. The intervals given after Theorems 3.2.1 and are two sided. Constructing one-sided intervals may be useful in risk management and similar jackknife empirical likelihood confidence intervals can be obtained.

3.2.3 Simulation study

In this section we examine the finite sample behavior of the proposed jackknife empirical likelihood method in terms of coverage accuracy and interval length, and compare it with the normal approximation method and the naive bootstrap method. Interval estimation for contaminated data is studied by Kaiser and Brazauskas [56]. We focus on the proportional hazards transform risk measure with $\psi(s) = a(1-s)^{a-1}$ and choose $a = 0.55$ and 0.85 for simulation. Since the Pareto distribution, log-normal distribution, Weibull distribution and Gamma distribution are widely used in fitting the losses data in insurance (see Klugman, Panjer and Willmot [57]), our simulation study is based on these four distributions.

We draw $5,000$ random samples of sizes $n = 300$ and 1000 from the following distributions:

1. Pareto distribution $F_1(x; \theta) = 1 - x^{-\theta}$ for $x \geq 1$;

2. Log-normal distribution $F_2(x; \theta_1, \theta_2) = \Phi((\log x - \theta_1)/\theta_2)$ for $x > 0$, where $\Phi(x)$ denotes the standard normal distribution function;

3. Weibull distribution $F_3(x;\theta_1,\theta_2) = 1 - \exp\{-(x/\theta_2)^{\theta_1}\}$ for $x > 0$;

4. Gamma distribution

$$F_4(x;\theta_1,\theta_2) = \int_0^x \frac{\theta_2^{\theta_1}}{\Gamma(\theta_1)} s^{\theta_1 - 1} \exp\{-\theta_2 s\} ds \quad \text{for} \quad x > 0.$$

For calculating the proposed jackknife empirical likelihood intervals (JELCI) for both $R(F)$ and $r(F)$, we use the R package `'emplik'`. For calculating the confidence intervals for $R(F)$ based on the normal approximation method (NACI), we use the variance estimation in Jones and Zitikis [51]. For computing the naive bootstrap confidence intervals for $r(F)$ (NBCI), we draw $5,000$ bootstrap samples with replacement from each random sample $X_1, \ldots, X_n$. Empirical coverage probabilities are reported in Tables 3.1 and 3.2 for these three confidence intervals with levels 0.9, 0.95 and 0.99. Tables 3.3 and 3.4 report the average interval lengths for these intervals. From these tables, we conclude that the proposed jackknife empirical likelihood method gives more accurate coverage probability than the other two methods especially for the case of $n = 300$. On the other hand, the new method has a bigger interval length than the other methods for most cases.

3.2.4 Proofs

Throughout we put $U_i = F(X_i)$ for $i = 1, \ldots, n$, $G_n(t) = n^{-1}\sum_{i=1}^n I(U_i \le t)$ and $G_{n,i} = (n-1)^{-1}\sum_{j=1, j\neq i}^n I(U_j \le t)$ for $i = 1, \ldots, n$. Since F is continuous, $U_1, \ldots, U_n$ are independent and uniformly distributed over $(0, 1)$. Without loss of generality we assume no ties in $U_1, \ldots, U_n$, and let $U_{n,1} < \cdots < U_{n,n}$ denote the order statistics of $U_1, \ldots, U_n$. We also use C to denote a generic constant which may be different in different places.

Under the conditions of Theorem 3.2.1, we first list some facts which will be employed in the proofs. We assume $\beta \le \alpha$ throughout since proofs for the case of $\beta > \alpha$ are exactly the same. Therefore we have $|\psi(x)| \le cx^{\beta-1}(1-x)^{\beta-1}$ and

Table 3.1: Coverage probabilities for $R(F)$ are reported for the intervals based on the proposed jackknife empirical likelihood method (JELCI) and the normal approximation method (NACI).

(n, a, F)	JELCI level 0.9	NACI level 0.9	JELCI level 0.95	NACI level 0.95	JELCI level 0.99	NACI level 0.99
$(300, 0.55, F_1(;4))$	0.6316	0.4408	0.7096	0.4978	0.8348	0.6082
$(300, 0.85, F_1(;4))$	0.8618	0.8500	0.9202	0.9020	0.9768	0.9512
$(1000, 0.55, F_1(;4))$	0.6160	0.4438	0.7084	0.5032	0.8402	0.6108
$(1000, 0.85, F_1(;4))$	0.8702	0.8642	0.9330	0.9240	0.9870	0.9738
$(300, 0.55, F_2(;0,1))$	0.6906	0.5376	0.7692	0.6020	0.8808	0.7012
$(300, 0.85, F_2(;0,1))$	0.8664	0.8560	0.9270	0.9104	0.9802	0.9590
$(1000, 0.55, F_2(;0,1))$	0.7206	0.5870	0.7968	0.6522	0.8972	0.7556
$(1000, 0.85, F_2(;0,1))$	0.8810	0.8698	0.9332	0.9236	0.9828	0.9750
$(300, 0.55, F_3(;4,1))$	0.8998	0.8798	0.9496	0.9344	0.9872	0.9802
$(300, 0.85, F_3(;4,1))$	0.9080	0.9066	0.9556	0.9534	0.9890	0.9884
$(1000, 0.55, F_3(;4,1))$	0.9032	0.8918	0.9530	0.9462	0.9912	0.9876
$(1000, 0.85, F_3(;4,1))$	0.9094	0.9068	0.9558	0.9560	0.9926	0.9932
$(300, 0.55, F_4(;4,1))$	0.8568	0.8024	0.9152	0.8718	0.9774	0.9460
$(300, 0.85, F_4(;4,1))$	0.8934	0.8842	0.9458	0.9402	0.9898	0.9870
$(1000, 0.55, F_4(;4,1))$	0.8728	0.8430	0.9336	0.9060	0.9844	0.9696
$(1000, 0.85, F_4(;4,1))$	0.9010	0.8988	0.9514	0.9490	0.9904	0.9900

Table 3.2: Coverage probabilities for $r(F)$ are reported for the intervals based on the proposed jackknife empirical likelihood method (JELCI) and the naive bootstrap method (NBCI).

(n, a, F)	JELCI level 0.9	NBCI level 0.9	JELCI level 0.95	NBCI level 0.95	JELCI level 0.99	NBCI level 0.99
$(300, 0.55, F_1(;4))$	0.5002	0.3682	0.5802	0.4060	0.6990	0.4858
$(300, 0.85, F_1(;4))$	0.7310	0.6782	0.8026	0.7366	0.8980	0.8128
$(1000, 0.55, F_1(;4))$	0.5550	0.4342	0.6344	0.4840	0.7610	0.5600
$(1000, 0.85, F_1(;4))$	0.7924	0.7536	0.8646	0.8124	0.9482	0.8830
$(300, 0.55, F_2(;0,1))$	0.5432	0.4242	0.6098	0.4744	0.7184	0.5628
$(300, 0.85, F_2(;0,1))$	0.7116	0.6546	0.7770	0.7168	0.8762	0.8084
$(1000, 0.55, F_2(;0,1))$	0.6102	0.5296	0.6850	0.5854	0.7908	0.6698
$(1000, 0.85, F_2(;0,1))$	0.7670	0.7290	0.8384	0.7928	0.9202	0.8726
$(300, 0.55, F_3(;4,1))$	0.8554	0.8380	0.9118	0.8936	0.9736	0.9608
$(300, 0.85, F_3(;4,1))$	0.8922	0.8798	0.9444	0.9320	0.9850	0.9802
$(1000, 0.55, F_3(;4,1))$	0.8646	0.8538	0.9192	0.9130	0.9776	0.9762
$(1000, 0.85, F_3(;4,1))$	0.8850	0.8796	0.9390	0.9330	0.9886	0.9842
$(300, 0.55, F_4(;4,1))$	0.7740	0.7200	0.8452	0.7924	0.9282	0.8820
$(300, 0.85, F_4(;4,1))$	0.8560	0.8346	0.9180	0.8960	0.9738	0.9598
$(1000, 0.55, F_4(;4,1))$	0.8200	0.7944	0.8876	0.8584	0.9538	0.9326
$(1000, 0.85, F_4(;4,1))$	0.8828	0.8758	0.9342	0.9254	0.9844	0.9780

Table 3.3: Average interval lengths for $R(F)$ are reported for the intervals based on the proposed jackknife empirical likelihood method (JELCI) and the normal approximation method (NACI).

(n, a, F)	JELCI level 0.9	NACI level 0.9	JELCI level 0.95	NACI level 0.95	JELCI level 0.99	NACI level 0.99
$(300, 0.55, F_1(;4))$	0.3336	0.2416	0.4038	0.2879	0.5409	0.3784
$(300, 0.85, F_1(;4))$	0.1217	0.1170	0.1485	0.1394	0.2041	0.1832
$(1000, 0.55, F_1(;4))$	0.2405	0.1762	0.2939	0.2100	0.4028	0.2760
$(1000, 0.85, F_1(;4))$	0.0678	0.0684	0.0830	0.0815	0.1142	0.1071
$(300, 0.55, F_2(;0,1))$	1.1940	1.1396	1.3265	1.3580	1.5084	1.7847
$(300, 0.85, F_2(;0,1))$	0.5835	0.5447	0.7034	0.6490	0.9342	0.8530
$(1000, 0.55, F_2(;0,1))$	0.9583	0.8167	1.0952	0.9731	1.3048	1.2789
$(1000, 0.85, F_2(;0,1))$	0.3319	0.3165	0.4016	0.3771	0.5446	0.4956
$(300, 0.55, F_3(;4,1))$	0.0996	0.0968	0.1209	0.1154	0.1643	0.1516
$(300, 0.85, F_3(;4,1))$	0.0911	0.0956	0.1097	0.1139	0.1461	0.1497
$(1000, 0.55, F_3(;4,1))$	0.0520	0.0545	0.0633	0.0649	0.0862	0.0853
$(1000, 0.85, F_3(;4,1))$	0.0498	0.0525	0.0596	0.0626	0.0788	0.0822
$(300, 0.55, F_4(;4,1))$	0.3132	0.2689	0.3809	0.3204	0.5221	0.4211
$(300, 0.85, F_4(;4,1))$	0.2043	0.2058	0.2454	0.2452	0.3273	0.3223
$(1000, 0.55, F_4(;4,1))$	0.1756	0.1582	0.2134	0.1885	0.2921	0.2477
$(1000, 0.85, F_4(;4,1))$	0.1092	0.1135	0.1314	0.1353	0.1750	0.1778

Table 3.4: Average interval lengths for $r(F)$ are reported for the intervals based on the proposed jackknife empirical likelihood method (JELCI) and the naive bootstrap method (NBCI).

(n, a, F)	JELCI level 0.9	NBCI level 0.9	JELCI level 0.95	NBCI level 0.95	JELCI level 0.99	NBCI level 0.99
$(300, 0.55, F_1(;4))$	0.1342	0.1273	0.1504	0.1445	0.1739	0.1761
$(300, 0.85, F_1(;4))$	0.0268	0.0226	0.0326	0.0262	0.0445	0.0330
$(1000, 0.55, F_1(;4))$	0.1218	0.1084	0.1387	0.1242	0.1650	0.1539
$(1000, 0.85, F_1(;4))$	0.0182	0.0160	0.0220	0.0187	0.0307	0.0239
$(300, 0.55, F_2(;0,1))$	0.4298	0.3964	0.4838	0.4509	0.5661	0.5488
$(300, 0.85, F_2(;0,1))$	0.0743	0.0634	0.0881	0.0732	0.1134	0.0910
$(1000, 0.55, F_2(;0,1))$	0.3922	0.3423	0.4468	0.3923	0.5342	0.4827
$(1000, 0.85, F_2(;0,1))$	0.0535	0.0461	0.0646	0.0538	0.0864	0.0682
$(300, 0.55, F_3(;4,1))$	0.0277	0.0249	0.0337	0.0296	0.0460	0.0387
$(300, 0.85, F_3(;4,1))$	0.0059	0.0061	0.0072	0.0073	0.0097	0.0096
$(1000, 0.55, F_3(;4,1))$	0.0154	0.0144	0.0187	0.0171	0.0256	0.0224
$(1000, 0.85, F_3(;4,1))$	0.0030	0.0034	0.0036	0.0041	0.0049	0.0053
$(300, 0.55, F_4(;4,1))$	0.0851	0.0689	0.1019	0.0810	0.1322	0.1038
$(300, 0.85, F_4(;4,1))$	0.0152	0.0141	0.0185	0.0167	0.0253	0.0217
$(1000, 0.55, F_4(;4,1))$	0.0532	0.0442	0.0649	0.0521	0.0890	0.0673
$(1000, 0.85, F_4(;4,1))$	0.0084	0.0083	0.0102	0.0098	0.0140	0.0128

$|\psi'(x)| \le cx^{\beta-2}(1-x)^{\beta-2}$ for all $0 < x < 1$. Since $\mathbb{E}|X_1|^{\gamma} < \infty$ with $\frac{1}{\gamma} + 1 - \beta < \frac{1}{2}$, we have

$$P(|X_1| > x) = o(x^{-\gamma}) \quad \text{as} \quad x \to \infty, \tag{3.6}$$

which implies

$$\int_0^{\infty} (F(x))^{\beta-1+\delta}(1-F(x))^{\beta-1+\delta}dx \le 2 + C\int_1^{\infty} x^{-(\beta-1+\delta)\gamma}dx < \infty \tag{3.7}$$

whenever $\delta \in (\frac{1}{\gamma} + 1 - \beta, \frac{1}{2})$, and

$$\max_{1\le j\le n} |X_j| = \max_{1\le j\le n} |F^{-}(U_j)| = o_p(n^{1/\gamma}). \tag{3.8}$$

It follows from the given conditions on ψ that

$$\Psi(\frac{1}{n}) = O(n^{-\beta}) \quad \text{and} \quad \Psi(\frac{1}{n-1}) - \Psi(\frac{1}{n}) = O(n^{-\beta-1}). \tag{3.9}$$

Lemma 3.2.3. *Under the conditions of Theorem 3.2.1, we have*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Y_i - R_0) \xrightarrow{d} N(0, \sigma_1^2), \tag{3.10}$$

where σ_1^2 is given in (3.4).

Proof. Write

$$
\begin{aligned}
Y_i &= (n-1)\int_0^\infty \{\Psi(F_{n,i}(t)) - \Psi(F_n(t))\}dt + \hat{R}_n \\
&= (n-1)\int_0^1 \{\Psi(G_{n,i}(t)) - \Psi(G_n(t))\}dF^-(t) + \hat{R}_n \\
&= (n-1)\int_0^1 \{\Psi(G_{n,i}(t)) - \Psi(G_n(t))\}I(U_{n,1} \le t < U_{n,n})dF^-(t) + \hat{R}_n \\
&= (n-1)\int_0^1 \{\Psi(G_{n,i}(t)) - \Psi(G_n(t))\}I(U_{n,1} \le t < U_{n,2})dF^-(t) \\
&\quad + (n-1)\int_0^1 \{\Psi(G_{n,i}(t)) - \Psi(G_n(t))\}I(U_{n,2} \le t < U_{n,n-1})dF^-(t) \\
&\quad + (n-1)\int_0^1 \{\Psi(G_{n,i}(t)) - \Psi(G_n(t))\}I(U_{n,n-1} \le t < U_{n,n})dF^-(t) + \hat{R}_n \\
&= \underbrace{(n-1)\int_0^1 \{\Psi(G_{n,i}(t)) - \Psi(G_n(t))\}I(U_{n,1} \le t < U_{n,2})dF^-(t)}_{Z_{i,1}} \\
&\quad + \underbrace{(n-1)\int_0^1 \psi(G_n(t))\{G_{n,i}(t) - G_n(t)\}I(U_{n,2} \le t < U_{n,n-1})dF^-(t)}_{Z_{i,2}} \\
&\quad + \underbrace{\frac{n-1}{2}\int_0^1 \psi'(\xi_{n,i}(t))\{G_{n,i}(t) - G_n(t)\}^2 I(U_{n,2} \le t < U_{n,n-1})dF^-(t)}_{Z_{i,3}} \\
&\quad + \underbrace{(n-1)\int_0^1 \{\Psi(G_{n,i}(t)) - \Psi(G_n(t))\}I(U_{n,n-1} \le t < U_{n,n})dF^-(t)}_{Z_{i,4}} + \hat{R}_n \\
&= Z_{i,1} + Z_{i,2} + Z_{i,3} + Z_{i,4} + \hat{R}_n,
\end{aligned}
$$

where

$$
\xi_{n,i}(t) = G_n(t) + \theta_i(t)\{G_{n,i}(t) - G_n(t)\} = G_n(t) + \frac{\theta_i(t)}{n-1}\{G_n(t) - I(U_i \le t)\}
$$

for some $\theta_i(t) \in [0,1]$.

When $U_{n,1} \le t < U_{n,2}$, we have

$$
G_n(t) = \frac{1}{n} \quad \text{and} \quad G_{n,i}(t) = \begin{cases} 0 & \text{if} \quad U_i = U_{n,1} \\ \frac{1}{n-1} & \text{else.} \end{cases} \tag{3.11}
$$

Hence, it follows from (3.8) and (3.9) that

$$\begin{aligned}\sum_{i=1}^{n} Z_{i,1} &= (n-1)\int_0^1 \{\Psi(0)-\Psi(\frac{1}{n})\}I(U_{n,1}\le t<U_{n,2})dF^-(t)\\ &\quad + (n-1)^2\int_0^1 \{\Psi(\frac{1}{n-1})-\Psi(\frac{1}{n})\}I(U_{n,1}\le t<U_{n,2})dF^-(t)\\ &= -(n-1)\Psi(\frac{1}{n})\{F^-(U_{n,2})-F^-(U_{n,1})\}\\ &\quad + (n-1)^2\{\Psi(\frac{1}{n-1})-\Psi(\frac{1}{n})\}\{F^-(U_{n,2})-F^-(U_{n,1})\}\\ &= O((n-1)n^{-\beta})o_p(n^{1/\gamma}) + O((n-1)^2 n^{-1-\beta})o_p(n^{1/\gamma})\\ &= o_p(n^{1/2-\beta+1/\gamma})\sqrt{n}\\ &= o_p(\sqrt{n})\end{aligned} \tag{3.12}$$

since $\frac{1}{2}-\beta+\frac{1}{\gamma}<0$. Similarly, we can show that

$$\sum_{i=1}^{n} Z_{i,4} = o_p(\sqrt{n}). \tag{3.13}$$

Since $\sum_{i=1}^{n}\{G_{n,i}(t)-G_n(t)\}=0$, we have

$$\sum_{i=1}^{n} Z_{i,2} = 0. \tag{3.14}$$

When $t \ge U_{n,2}$, we have

$$\frac{(n-1)^{-1}I(U_i\le t)}{G_n(t)} \le \frac{1/(n-1)}{2/n} = \frac{n}{2(n-1)},$$

i.e.,

$$\xi_{n,i}(t) \ge G_n(t)\{1-\frac{n}{2(n-1)}\}$$

uniformly in $t \ge U_{n,2}$. In the same manner, we can show that

$$1-\xi_{n,i}(t) \ge (1-G_n(t))\{1-\frac{n}{2(n-1)}\}$$

holds uniformly in $t < U_{n,n-1}$. Hence, for n large enough,

$$(\xi_{n,i}(t), 1-\xi_{n,i}(t)) \ge \frac{1}{3}(G_n(t), 1-G_n(t)) \tag{3.15}$$

uniformly for $U_{n,2} \leq t < U_{n,n-1}$ and $1 \leq i \leq n$.

Note that

$$\sup_{U_{n,2}\leq t\leq U_{n,n-1}} \frac{G_n(t)}{t} = O_p(1) \quad \text{and} \quad \sup_{U_{n,2}\leq t\leq U_{n,n-1}} \frac{1-G_n(t)}{1-t} = O_p(1) \tag{3.16}$$

(see Page 404 of Shorack and Wellner [94]). It follows from (3.15) and (3.16) that

$$|Z_{i,3}| = O_p\left(n\int_0^1 t^{\beta-2}(1-t)^{\beta-2}\{G_{n,i}(t)-G_n(t)\}^2 I(U_{n,2}\leq t<U_{n,n-1})dF^-(t)\right),$$

which coupled with (3.7) and (3.16) yields

$$\begin{aligned}
\sum_{i=1}^n Z_{i,3} &= O_p\left(n\int_0^1 t^{\beta-2}(1-t)^{\beta-2}\sum_{i=1}^n\{G_{n,i}(t)-G_n(t)\}^2 I(U_{n,2}\leq t<U_{n,n-1})dF^-(t)\right)\\
&= O_p\left(n\int_0^1 t^{\beta-2}(1-t)^{\beta-2}\frac{n}{(n-1)^2}G_n(t)\{1-G_n(t)\}I(U_{n,2}\leq t<U_{n,n-1})dF^-(t)\right)\\
&= O_p\left(\int_0^1 t^{\beta-1}(1-t)^{\beta-1}I(U_{n,2}\leq t<U_{n,n-1})dF^-(t)\right)\\
&= O_p\left(\int_{n^{-1}}^{1-n^{-1}} t^{\beta-1}(1-t)^{\beta-1}dF^-(t)\right)\\
&= O_p\left(n^\delta\int_{n^{-1}}^{1-n^{-1}} t^{\beta-1+\delta}(1-t)^{\beta-1+\delta}dF^-(t)\right)\\
&= O_p\left(n^\delta\int_0^\infty (F(x))^{\beta-1+\delta}(1-F(x))^{\beta-1+\delta}dx\right)\\
&= O_p(n^\delta)
\end{aligned} \tag{3.17}$$

for any $\delta \in (\frac{1}{\gamma}+1-\beta, \frac{1}{2})$. By Jones and Zitikis [51], we have

$$\sqrt{n}\{\hat{R}_n - R\} \xrightarrow{d} N(0, \sigma_1^2). \tag{3.18}$$

Hence, the lemma follows from (3.12), (3.14), (3.17), (3.13) and (3.18). □

Lemma 3.2.4. *Under the conditions of Theorem 3.2.1, we have*

$$\frac{1}{n}\sum_{i=1}^n (Y_i - R)^2 \xrightarrow{p} \sigma_1^2 \quad as \quad n\to\infty.$$

Proof. We use the same notations $Z_{i,j}$ as in the proof of Lemma 3.2.3. Then, it follows from (3.11) and (3.9) that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} Z_{i,1}^2 =& \frac{(n-1)^2}{n}\int_0^1\int_0^1 \{\Psi(0)-\Psi(\frac{1}{n})\}^2 I(U_{n,1}\le t_1, t_2 < U_{n,2})dF^-(t_1)dF^-(t_2)\\ &+\frac{(n-1)^3}{n}\int_0^1\int_0^1 \{\Psi(\frac{1}{n-1})-\Psi(\frac{1}{n})\}^2 I(U_{n,1}\le t_1, t_2 < U_{n,2})dF^-(t_1)dF^-(t_2)\\ &= O(\frac{(n-1)^2}{n}n^{-2\beta})o_p(n^{2/\gamma}) + O(\frac{(n-1)^3}{n}n^{-2-2\beta})o_p(n^{2/\gamma})\\ &= o_p(1).\end{aligned} \tag{3.19}$$

Similarly,

$$\frac{1}{n}\sum_{i=1}^{n} Z_{i,4}^2 = o_p(1). \tag{3.20}$$

It is easy to check that

$$\begin{aligned}&\frac{1}{n}\sum_{i=1}^{n} Z_{i,2}^2\\ =&\frac{(n-1)^2}{n}\int_0^1\int_0^1 \psi(G_n(t_1))\psi(G_n(t_2))\sum_{i=1}^{n}\{G_{n,i}(t_1)-G_n(t_1)\}\{G_{n,i}(t_2)-G_n(t_2)\}\\ &\times I(U_{n,2}\le t_1, t_2 < U_{n,n-1})dF^-(t_2)dF^-(t_1)\\ =&\int_0^1\int_0^1 \psi(G_n(t_1))\psi(G_n(t_2))\{G_n(t_1\wedge t_2)-G_n(t_1)G_n(t_2)\}\\ &I(U_{n,2}\le t_1, t_2 < U_{n,n-1})dF^-(t_2)dF^-(t_1)\\ =&2\underbrace{\int_0^1\int_0^{t_1} \psi(G_n(t_1))\psi(G_n(t_2))G_n(t_2)\{1-G_n(t_1)\}I(U_{n,1}\le t_1, t_2 < U_{n,n-1})dF^-(t_2)dF^-(t_1)}_{I_0}.\end{aligned}$$

By (3.16), we have

$$\begin{aligned}&\sup_{U_{n,2}\le t_1, t_2 < U_{n,n-1}} \psi(G_n(t_1))\psi(G_n(t_2))G_n(t_2)\{1-G_n(t_1)\}\\ =& O_p\left(t_1^{\beta-1}(1-t_1)^{\beta-1}t_2^{\beta-1}(1-t_2)^{\beta-1}t_2(1-t_1)\right).\end{aligned}$$

Similar to the proof of (3.7), we can show that

$$\begin{aligned}
&\int_0^1 \int_0^{t_1} t_1^{\beta-1}(1-t_1)^{\beta-1} t_2^{\beta-1}(1-t_2)^{\beta-1} t_2(1-t_1) dF^-(t_2) dF^-(t_1) \\
= &\int_0^{\infty} \int_0^{F^-(t_1)} F(x)^{\beta-1}(1-F(x))^{\beta-1} F(y)^{\beta-1}(1-F(y))^{\beta-1} F(y)(1-F(x)) dy dx \\
< &\ \infty.
\end{aligned}$$

By the Glivenko-Cantelli theorem, $\sup_{0<t<1} |G_n(t) - t| \to 0$ almost surely. It then follows from the dominated convergence theorem that

$$I_0 \xrightarrow{p} \int_0^1 \int_0^{t_1} \psi(t_1)\psi(t_2) t_2(1-t_1) dF^-(t_2) dF^-(t_1).$$

Hence

$$\frac{1}{n} \sum_{i=1}^{n} Z_{i,2}^2 \xrightarrow{p} \int_0^1 \int_0^1 \psi(t_1)\psi(t_2)\{t_1 \wedge t_2 - t_1 t_2\} dF^-(t_2) dF^-(t_1) = \sigma_1^2. \tag{3.21}$$

Note that

$$\begin{aligned}
&\sum_{i=1}^{n} \{G_{n,i}(t_1) - G_n(t_1)\}^2 \{G_{n,i}(t_2) - G_n(t_2)\}^2 \\
=&\sum_{i=1}^{n} \{\frac{G_n(t_1)}{n-1} - \frac{I(U_i \leq t_1)}{n-1}\}^2 \{\frac{G_n(t_2)}{n-1} - \frac{I(U_i \leq t_2)}{n-1}\}^2 \\
=&\frac{n}{(n-1)^4}\{-3G_n^2(t_1)G_n^2(t_2) + G_n^2(t_1)G_n(t_2) + G_n(t_1)G_n^2(t_2) + 4G_n(t_1)G_n(t_2)G_n(t_1 \wedge t_2) \\
&- 2G_n(t_1)G_n(t_1 \wedge t_2) - 2G_n(t_2)G_n(t_1 \wedge t_2) + G_n(t_1 \wedge t_2)\} \\
=&\frac{n}{(n-1)^4}\{\underbrace{3G_n(t_1)G_n(t_2)\left(G_n(t_1 \wedge t_2) - G_n(t_1)G_n(t_2)\right)}_{I_1} - \underbrace{G_n(t_1)\left(G_n(t_1 \wedge t_2) - G_n(t_1)G_n(t_2)\right)}_{I_2} \\
&- \underbrace{G_n(t_2)\left(G_n(t_1 \wedge t_2) - G_n(t_1)G_n(t_2)\right)}_{I_3} + \underbrace{(1 - G_n(t_1))(1 - G_n(t_2))G_n(t_1 \wedge t_2)}_{I_4}\} \\
=&\frac{n}{(n-1)^4}\{I_1 - I_2 - I_3 + I_4\}.
\end{aligned}$$

It follows from (3.16) that

$$\begin{aligned}
\sup_{U_{n,2} \leq t_1, t_2 \leq U_{n,n-1}} \frac{|G_n(t_1 \wedge t_2) - G_n(t_1)G_n(t_2)|}{t_1 \wedge t_2 - t_1 t_2} &= O_p(1), \\
\sup_{U_{n,2} \leq t_1, t_2 \leq U_{n,n-1}} \frac{|G_n(t_1 \wedge t_2)(1 - G_n(t_1 \vee t_2)|}{t_1 \wedge t_2(1 - t_1 \vee t_2)} &= O_p(1).
\end{aligned}$$

This coupled with (3.15) and (3.16), yields that

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} Z_{i,3}^2\\
=&O_p\Big(\frac{(n-1)^2}{4n}\int_0^1\int_0^1 t_1^{\beta-2}(1-t_1)^{\beta-2}t_2^{\beta-2}(1-t_2)^{\beta-2}\\
&\times\sum_{i=1}^{n}\{G_{n,i}(t_1)-G_n(t_1)\}^2\{G_{n,i}(t_2)-G_n(t_2)\}^2 I(U_{n,2}\le t_1,t_2<U_{n,n-1})dF^-(t_2)dF^-(t_1)\Big)\\
=&O_p\big(n^{-2}\int_0^1\int_0^1 t_1^{\beta-2}(1-t_1)^{\beta-2}t_2^{\beta-2}(1-t_2)^{\beta-2}\\
&\times\{I_1-I_2-I_3+I_4\}I(U_{n,2}\le t_1,t_2<U_{n,n-1})dF^-(t_2)dF^-(t_1)\big).
\end{aligned}
$$

From the above equation we can get that

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} Z_{i,3}^2\\
=&O_p\big(\underbrace{n^{-2}\int_{U_{n,2}}^{U_{n,n-1}}\int_{U_{n,2}}^{U_{n,n-1}} t_1^{\beta-2}(1-t_1)^{\beta-2}t_2^{\beta-2}(1-t_2)^{\beta-2}t_1t_2(t_1\wedge t_2-t_1t_2)dF^-(t_2)dF^-(t_1)}_{J_1}\big)\\
&+O_p\big(\underbrace{n^{-2}\int_{U_{n,2}}^{U_{n,n-1}}\int_{U_{n,2}}^{U_{n,n-1}} t_1^{\beta-2}(1-t_1)^{\beta-2}t_2^{\beta-2}(1-t_2)^{\beta-2}t_1(t_1\wedge t_2-t_1t_2)dF^-(t_2)dF^-(t_1)}_{J_2}\big)\\
&+O_p\big(\underbrace{n^{-2}\int_{U_{n,2}}^{U_{n,n-1}}\int_{U_{n,2}}^{U_{n,n-1}} t_1^{\beta-2}(1-t_1)^{\beta-2}t_2^{\beta-2}(1-t_2)^{\beta-2}t_2(t_1\wedge t_2-t_1t_2)dF^-(t_2)dF^-(t_1)}_{J_3}\big)\\
&+O_p\big(\underbrace{n^{-2}\int_{U_{n,2}}^{U_{n,n-1}}\int_{U_{n,2}}^{U_{n,n-1}} t_1^{\beta-2}(1-t_1)^{\beta-2}t_2^{\beta-2}(1-t_2)^{\beta-2}(1-t_1)(1-t_2)(t_1\wedge t_2)dF^-(t_2)dF^-(t_1)}_{J_4}\big)\\
=&O_p(J_1)+O_p(J_2)+O_p(J_3)+O_p(J_4).
\end{aligned}
$$

It is easy to check from (3.7) that for every $\delta\in(\frac{1}{\gamma}+1-\beta,\frac{1}{2})$

$$
\begin{aligned}
J_2+J_3=&2n^{-2}\int_{U_{n,2}}^{U_{n,n-1}}\int_{U_{n,2}}^{t_1} t_1^{\beta-2}(1-t_1)^{\beta-2}t_2^{\beta-2}(1-t_2)^{\beta-2}(t_1+t_2)t_2(1-t_1)dF^-(t_2)dF^-(t_1)\\
\le&4n^{-2}\int_{U_{n,2}}^{U_{n,n-1}} t_1^{\beta-1}(1-t_1)^{\beta-1}\int_{U_{n,2}}^{t_1} t_2^{\beta-1}(1-t_2)^{\beta-2}dF^-(t_2)dF^-(t_1)\\
=&n^{-2}\int_0^1 t_1^{\beta-1}(1-t_1)^{\beta-1}O(U_{n,2}^{-\delta}+(1-t_1)^{-1-\delta})dF^-(t_1)\\
=&O(n^{-2}U_{n,2}^{-\delta})\int_0^1 t_1^{\beta-1}(1-t_1)^{\beta-1}dF^-(t_1)+O(n^{-2})\int_0^1 t_1^{\beta-1}(1-t_1)^{\beta-2-\delta}dF^-(t_1)\\
=&O(n^{-2}U_{n,2}^{-\delta})(U_{n,2}^{-\delta}+(1-U_{n,n-1})^{-\delta})+O(n^{-2})(U_{n,2}^{-\delta}+(1-U_{n,n-1})^{-1-2\delta})\\
=&O_p(n^{-2+2\delta}+n^{-1+2\delta})\\
=&o_p(1).
\end{aligned}
$$

Similarly, we can show that

$$J_1 = o_p(1) \quad \text{and} \quad J_4 = o_p(1).$$

Hence,

$$\frac{1}{n}\sum_{i=1}^{n} Z_{i,3}^2 = o_p(1). \tag{3.22}$$

Since $\hat{R}_n \xrightarrow{p} R$, we have

$$\frac{1}{n}\sum_{i=1}^{n} (\hat{R}_n - R)^2 = o_p(1). \tag{3.23}$$

It follows from (3.19), (3.20), (3.22) and (3.23) that

$$\frac{1}{n}\sum_{i=1}^{n} \{Z_{i,1}+Z_{i,3}+Z_{i,4}+\hat{R}_n-R\}^2 = O(\frac{1}{n}\sum_{i=1}^{n}\{Z_{i,1}^2+Z_{i,3}^2+Z_{i,4}^2+(\hat{R}_n-R)^2\}) = o_p(1). \tag{3.24}$$

Note that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} Z_{i,2}\{Z_{i,1}+Z_{i,3}+Z_{i,4}+\hat{R}_n-R\} \\
&\le \sqrt{\frac{1}{n}\sum_{i=1}^{n} Z_{i,2}^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\{Z_{i,1}+Z_{i,3}+Z_{i,4}+\hat{R}_n-R\}^2} \\
&= o_p(1).
\end{aligned} \tag{3.25}$$

Therefore, the lemma follows from (3.19)–(3.25). □

Proof of Theorem 3.2.1. First we observe by using (3.7) that for any $\delta \in (\frac{1}{\gamma}+1-\beta, \frac{1}{2})$,

$$\begin{aligned}
\max_{1\le i\le n} |Z_{i,2}| &\le \int_0^1 \psi(G_n(t))I(U_{n,2} \le t < U_{n,n-1})dF^-(t) \\
&= O_p\left(\int_{U_{n,2}}^{U_{n,n-1}} t^{\beta-1}(1-t)^{\beta-1}dF^-(t)\right) \\
&= O_p(U_{n,2}^{-\delta} + (1-U_{n,n-1})^{-\delta}) \\
&= o_p(n^{1/2}).
\end{aligned} \tag{3.26}$$

Similarly we can show that

$$\max_{1\le i\le n} |Z_{i,j}| = o_p(n^{1/2}) \quad \text{for} \quad j = 1, 3, 4.$$

Hence, $\max_{1\le i\le n}|Y_i| = o_p(n^{1/2})$. By the standard arguments in the empirical likelihood method (see Chapter 11 of Owen [73]), it follows from Lemmas 3.2.3 and 3.2.4 that

$$-2\log L_1(R) = \frac{\{\sum_{i=1}^n (Y_i - R)\}^2}{\sum_{i=1}^n (Y_i - R)^2} + o_p(1) \xrightarrow{d} \chi^2(1).$$

□

In order to prove Theorem 3.2.3, we need the following lemmas.

Lemma 3.2.5. *Under the conditions of Theorem 3.2.3, we have*

$$\sqrt{n}\left(\hat{R}_n - \frac{R(F)}{\mu}\frac{1}{n}\sum_{i=1}^n X_i\right) \xrightarrow{d} N(0,\bar{\sigma}^2) \quad \text{as} \quad n\to\infty,$$

where

$$\bar{\sigma}^2 = \int_0^1\int_0^1 \psi(t_1)\psi(t_2)(t_1\wedge t_2 - t_1t_2)dF^-(t_1)dF^-(t_2) + \frac{R^2(F)}{\mu^2}\mathbb{E}(X_1-\mu)^2$$
$$+ 2\frac{R(F)}{\mu}\int_0^1\int_0^1 \psi(t_1)(t_1\wedge t_2 - t_1t_2)dF^-(t_1)dF^-(t_2).$$

Proof. It is known that there exists a Brownian bridge W such that

$$\sup_{0\le t\le 1}\frac{\sqrt{n}(G_n(t)-t)-W(t)}{t^{\delta_0}(1-t)^{\delta_0}} = o_p(1) \tag{3.27}$$

for any $\delta_0\in(0,1/2)$ (see Chapter 4 of Csorgo and Horvath [22]). It follows from (3.8) and (3.9) that

$$\begin{aligned}
&\sqrt{n}\textstyle\int_0^1\{\Psi(t)-\Psi(G_n(t))\}I(t<U_{n,1})\,dF^-(t)\\
=\;&\sqrt{n}\textstyle\int_0^{U_{n,1}}\Psi(t)\,dF^-(t)\\
\le\;&\sqrt{n}\Psi(U_{n,1})F^-(U_{n,1}) \qquad\qquad (3.28)\\
=\;&o_p(\sqrt{n}n^{-\beta}n^{1/\gamma})\\
=\;&o_p(1).
\end{aligned}$$

Similarly we can show that

$$\begin{cases}
\sqrt{n}\int_0^1\{t-G_n(t)\}\psi(t)I(t<U_{n,1})\,dF^-(t) = o_p(1)\\
\sqrt{n}\int_0^1\{\Psi(t)-\Psi(G_n(t))\}I(t>U_{n,n-1})\,dF^-(t) = o_p(1)\\
\sqrt{n}\int_0^1\{t-G_n(t)\}\psi(t)I(t<U_{n,n-1})\,dF^-(t) = o_p(1).
\end{cases} \tag{3.29}$$

Note that (3.16) holds with $U_{n,2}$ replaced by $U_{n,1}$ and we assume $\beta \le \alpha$ in the beginning of Section 3.2.4. Hence, by the Taylor expansion, (3.27), (3.7) and choosing δ_0 close to 1/2 enough such that $\delta + 1/2 - 2\delta_0 < 0$ with $\delta \in (1/\gamma + 1 - \beta, 1/2)$, we have

$$\begin{aligned}
& \sqrt{n}\int_0^1 \{\Psi(t) - \Psi(G_n(t)) - (t - G_n(t))\psi(t)\} I(U_{n,1} \le t \le U_{n,n-1})\, dF^-(t) \\
= & \sqrt{n}\int_{U_{n,1}}^{U_{n,n-1}} \tfrac{1}{2}\psi'(\xi)\{t - G_n(t)\}^2\, dF^-(t) \\
= & O_p\left(\tfrac{1}{\sqrt{n}}\int_{n^{-1}}^{1-n^{-1}} t^{\beta-2}(1-t)^{\beta-2} t^{2\delta_0}(1-t)^{2\delta_0}\, dF^-(t)\right) \\
= & O_p\left(n^{-1/2+\delta+1-2\delta_0}\int_{n^{-1}}^{1-n^{-1}} t^{\beta-1+\delta}(1-t)^{\beta-1+\delta}\, dF^-(t)\right) \\
= & O_p\left(n^{\delta+1/2-2\delta_0}\int_0^\infty (F(x))^{\beta-1+\delta}(1-F(x))^{\beta-1+\delta}\, dx\right) \\
= & o_p(1),
\end{aligned} \tag{3.30}$$

where ξ depends on t and lies between t and $G_n(t)$. It follows from (3.27)–(3.30) that

$$\sqrt{n}\int_0^1 \{\Psi(t) - \Psi(G_n(t))\}\, dF^-(t) + \int_0^1 \psi(t)W(t)\, dF^-(t) = o_p(1).$$

Therefore

$$\begin{aligned}
& \sqrt{n}\{\hat{R}_n - \frac{R(F)}{\mu}\frac{1}{n}\sum_{i=1}^n X_i\} \\
= & \sqrt{n}\int_0^1 \{\Psi(t) - \Psi(G_n(t))\} dF^-(t) + \frac{R(F)}{\mu}\sqrt{n}\int_0^1 \{t - G_n(t)\} dF^-(t) \\
\xrightarrow{d} & -\int_0^1 \psi(t)W(t) dF^-(t) - \frac{R(F)}{\mu}\int_0^1 W(t) dF^-(t).
\end{aligned}$$

□

Lemma 3.2.6. *Under the conditions of Theorem 3.2.3, we have*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \left(Y_i - \frac{R(F)}{\mu}X_i\right) \xrightarrow{d} N(0, \bar{\sigma}^2) \quad \text{as} \quad n \to \infty.$$

Proof. It can be shown in a way similar to the proof of Lemma 3.2.3. □

Lemma 3.2.7. *Under the conditions of Theorem 3.2.3, we have*

$$\frac{1}{n}\sum_{i=1}^n \left(Y_i - \frac{R(F)}{\mu}X_i\right)^2 \xrightarrow{p} \bar{\sigma}^2 \quad \text{as} \quad n \to \infty.$$

Proof. It can be proved in a similar way to the proof of Lemma 3.2.4. □

Proof of Theorem 3.2.3. This can be done in a way similar to the proof of Theorem 3.2.1. □

3.3 Interval Estimation for Spearman's Rho

In connection with copulas, rank correlation such as Kendall's tau and Spearman's rho has been employed in risk management for summarizing dependence among two variables and estimating some parameters in bivariate copulas and elliptical models. In this paper, a jackknife empirical likelihood method is proposed to construct confidence intervals for Spearman's rho without estimating the asymptotic variance. A simulation study confirms the advantages of the proposed method.

3.3.1 Introduction

Correlation has been used to summarize dependence among variables for a long history and plays an important role in modern finance such as Capital Asset Pricing Model and portfolio selection. Given the fact that copula and elliptical distributions have been heavily employed in risk management, copula-based dependence measures such as Kendall's tau and Spearman's rho are receiving more and more attention. Some pitfalls on using the linear correlation measure in elliptical models are given in Embrechts, McNeil and Straumann [34]. Advantages of using Kendall's tau and Spearman's rho include estimating some parameters in copulas. For example, if (X, Y) is a bivariate meta-Gaussian distribution with copula

$$C_\rho^{Ga}(u_1, u_2) = \int_{-\infty}^{\Phi^{-1}(u_1)} \int_{-\infty}^{\Phi^{-1}(u_2)} \frac{1}{2\pi(1-\rho^2)^{1/2}} \exp\{-\frac{s_1^2 - 2\rho s_1 s_2 + s_2^2}{2(1-\rho^2)}\} ds_1 ds_2$$

and continuous marginals, where Φ^{-1} denotes the inverse function of the standard normal distribution function, then the Kendall's tau and Spearman's rho can be written as

$$\tau = \frac{2}{\pi} \arcsin \rho \quad \text{and} \quad \rho^s = \frac{6}{\pi} \arcsin \frac{\rho}{2}.$$

Therefore ρ can be estimated via estimating τ and ρ^s. More details can be found in Chapter 5.3 of McNeil, Frey and Embrechts [65].

Let $(X_1, Y_1), \cdots, (X_n, Y_n)$ be independent random vectors with distribution function H and continuous marginals $F(x) = H(x, \infty)$ and $G(y) = H(\infty, y)$. Then the

Kendall's tau and Spearman's rho are defined as

$$\tau = \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) > 0] - \mathbb{P}[(X_1 - X_2)(Y_1 - Y_2) < 0]$$

and

$$\rho^s = 12\mathbb{E}[(F(X_1) - 1/2)(G(Y_1) - 1/2)],$$

respectively. Define

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq x) \quad \text{and} \quad G_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(Y_i \leq x).$$

Then the simple nonparametric estimators for τ and ρ^s are

$$\hat{\tau}_n = \frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} \{I((X_i - X_j)(Y_i - Y_j) > 0) - I((X_i - X_j)(Y_i - Y_j) < 0)\}$$

and

$$\hat{\rho}_n^s = \frac{12}{n}\sum_{i=1}^{n}\{F_n(X_i) - 1/2\}\{G_n(Y_i) - 1/2\},$$

respectively.

In order to construct confidence intervals for τ and ρ^s, one can simply use the asymptotic limits of $\sqrt{n}\{\hat{\tau}_n - \tau\}$ and $\sqrt{n}\{\hat{\rho}_n^s - \rho^s\}$. However, this method requires to estimate the asymptotic variances. As shown in the next section, the asymptotic variance of $\hat{\rho}_n^s$ is quite complicated and it is hard to estimate it explicitly. Most likely, it involves density estimation and numerical integration. Therefore, bootstrap method is a common way to construct a confidence interval for the Spearman's rho. As an alternative way of constructing confidence intervals, empirical likelihood method introduced in Chapter I is powerful in dealing with linear functionals without estimating any extra quantities such as asymptotic variance. Since the Kendall's tau and Spearman's rho are non-linear functionals, a direct application of empirical likelihood method fails to obtaining a chi-square limit, and the jackknife empirical likelihood method is required. Noting that $\hat{\tau}_n$ is a U-statistic, one can directly employ the method in Jing, Yuan and Zhou [47] to construct confidence intervals for the

Kendall's tau without estimating the asymptotic variance. In this section, we employ the jackknife empirical likelihood method to construct confidence intervals for the Spearman's rho and investigate the finite sample behavior of the proposed method.

We organize the rest of this section as follows. Section 3.3.2 presents the methodology and asymptotic results. A simulation study and a real data analysis are given in Section 3.3.3. All proofs are put in Section 3.3.4.

3.3.2 Methodology

Define the copula and empirical copula of (X_i, Y_i) as

$$C(x,y) = \mathbb{P}(F(X_1) \le x, G(Y_1) \le y)$$

and

$$C_n(x,y) = \frac{1}{n}\sum_{i=1}^{n} I(F_n(X_i) \le x, G_n(Y_i) \le y),$$

respectively. Put

$$C_1(x,y) = \frac{\partial}{\partial x}C(x,y) \quad \text{and} \quad C_2(x,y) = \frac{\partial}{\partial y}C(x,y).$$

Assume that

$$\begin{cases} C_1(x,y) & \text{exists and is continuous on the set} \quad \{(x,y): 0 < x < 1, 0 \le y \le 1\}, \\ C_2(x,y) & \text{exists and is continuous on the set} \quad \{(x,y): 0 \le x \le 1, 0 < y < 1\}. \end{cases} \tag{3.31}$$

Then it follows from Proposition 3.1 of Segers [90] that

$$\sup_{0\le x,y\le 1} |\sqrt{n}\{C_n(x,y) - C(x,y)\} - W(x,y) + C_1(x,y)W(x,1) + C_2(x,y)W(1,y)| = o_p(1), \tag{3.32}$$

where $W(x,y)$ is a Gaussian process with mean zero and covariance

$$\mathbb{E}[W(x_1,y_1)W(x_2,y_2)] = C(x_1 \wedge x_2, y_1 \wedge y_2) - C(x_1,y_1)C(x_2,y_2). \tag{3.33}$$

Note that (3.50) holds via the Skorohod construction. By (3.50), we have

$$\begin{aligned} \sqrt{n}\{\hat{\rho}_n^s - \rho^s\} &= 12\int_0^1\int_0^1 \sqrt{n}\{C_n(x,y) - C(x,y)\}dxdy \\ &\xrightarrow{d} 12\int_0^1\int_0^1 \{W(x,y) - C_1(x,y)W(x,1) - C_2(x,y)W(1,y)\}dxdy. \end{aligned}$$

Hence, the asymptotic limit depends on the copula $C(x, y)$ and its partial derivatives. In order to avoid estimating the complicated asymptotic variance for constructing confidence intervals for ρ^s, we employ the following jackknife empirical likelihood method.

Define

$$\begin{cases} F_{n,i}(x) = \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} I(X_j \leq x), \\ G_{n,i}(x) = \frac{1}{n-1}\sum_{j=1,j\neq i}^{n} I(Y_j \leq x), \\ \hat{\rho}_{n,i}^s = \frac{12}{n-1}\sum_{j=1,j\neq i}^{n}\{F_{n,i}(X_j) - 1/2\}\{G_{n,i}(Y_j) - 1/2\} \\ Z_i = n\hat{\rho}_n^s - (n-1)\hat{\rho}_{n,i}^s \end{cases}$$

for $i = 1, \cdots, n$. As in Jing, Yuan and Zhou [47], a jackknife empirical likelihood function for $\theta = \rho^s$ is defined as

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

By the Lagrange multiplier technique, we obtain that $p_i = n^{-1}\{1 + \lambda(Z_i - \theta)\}^{-1}$ and $-2\log L(\theta) = 2\sum_{i=1}^{n}\log\{1 + \lambda(Z_i - \theta)\}$, where $\lambda = \lambda(\theta)$ satisfies

$$\frac{1}{n}\sum_{i=1}^{n}\frac{Z_i - \theta}{1 + \lambda(Z_i - \theta)} = 0. \tag{3.34}$$

The following theorem shows that Wilks' Theorem holds for the proposed jackknife empirical likelihood method.

Theorem 3.3.1. *Assume condition (3.31) holds. Then $-2\log L(\rho^s)$ converges in distribution to a chi-square distribution with one degree of freedom as $n \to \infty$.*

Based on the above theorem, a jackknife empirical likelihood confidence interval for ρ^s with level α can be obtained as

$$I_\alpha = \{\theta : -2\log L(\theta) \leq \chi^2_{1,\alpha}\},$$

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

3.3.3 Simulation study and data analysis

Simulation study. We investigate the finite sample behavior of the proposed jackknife empirical likelihood method and compare it with the normal approximation method in terms of coverage accuracy.

We draw $10,000$ random samples of sample size $n = 100, 300$ from a bivariate normal distribution with correlation ρ and marginals being the standard normal distribution. In this case, the Spearman's rho equals $\frac{6}{\pi}\arcsin(\rho/2)$. We calculate the jackknife empirical likelihood interval I_α at levels $\alpha = 0.9, 0.95, 0.99$ for $\rho = 0, \pm 0.2, \pm 0.8$, which correspond to $\rho^s = 0, \pm 0.1913, \pm 0.7859$, respectively. For constructing a confidence interval based on the asymptotic limit of $\hat{\rho}_n^s$, we employ the percentile bootstrap confidence interval. More specifically, we draw $1,000$ bootstrap samples of size n from each original sample. Based on each bootstrap sample, we calculate the Spearman's rho estimator. Therefore we obtained $1,000$ bootstrapped Spearman's rho estimators denoted by $\hat{\rho}_{n,1}^{s*}, \cdots, \hat{\rho}_{n,1000}^{s*}$. Let c_1 and c_2 denote the $[1000(1-\alpha)/2] - th$ and $[1000(1+\alpha)/2] - th$ largest order statistics of $\{\hat{\rho}_{n,i}^{s*} - \hat{\rho}_n^s\}_{i=1}^{1000}$. Hence, the percentile bootstrap confidence interval for ρ^s with level α is

$$I_\alpha^B = (\hat{\rho}_n^s - c_2, \quad \hat{\rho}_n^s - c_1).$$

The empirical coverage probabilities and average interval lengths for both I_α and I_α^B are reported in Tables 3.5 and 3.6, which show that i) the proposed jackknife empirical likelihood method produces much more accurate confidence intervals than the percentile bootstrap method in most cases, specially for $n = 100$; ii) the interval lengths of the jackknife empirical likelihood method are slightly longer.

Data analysis. Next, we apply the proposed method to the Danish fire insurance claims. This data set is available at *www.ma.hw.ac.uk/∼mcneil/*, which consists of loss to buildings, loss to contents and loss to profits. As described there, the data were collected at the Copenhagen Reinsurance Company and comprise 2167

Table 3.5: Coverage probabilities for the intervals I_α and I_α^B at levels $\alpha = 0.9, 0.95, 0.99$ are reported for $n = 100, 300$ and $\rho = 0, \pm 0.2, \pm 0.8$.

(n, ρ)	$I_{0.9}$	$I_{0.9}^B$	$I_{0.95}$	$I_{0.95}^B$	$I_{0.99}$	$I_{0.99}^B$
$(100, 0)$	0.9024	0.8874	0.9524	0.9352	0.9898	0.9794
$(100, 0.2)$	0.9016	0.8867	0.9524	0.9349	0.9900	0.9791
$(100, -0.2)$	0.9003	0.8858	0.9513	0.9347	0.9896	0.9773
$(100, 0.8)$	0.9013	0.8876	0.9473	0.9264	0.9850	0.9624
$(100, -0.8)$	0.8926	0.8691	0.9390	0.9105	0.9818	0.9509
$(300, 0)$	0.9055	0.8999	0.9530	0.9476	0.9915	0.9864
$(300, 0.2)$	0.9035	0.8996	0.9513	0.9440	0.9906	0.9852
$(300, -0.2)$	0.9073	0.9017	0.9529	0.9467	0.9908	0.9860
$(300, 0.8)$	0.9037	0.8957	0.9529	0.9393	0.9900	0.9776
$(300, -0.8)$	0.9008	0.8920	0.9505	0.9377	0.9899	0.9782

Table 3.6: Average interval lengths for I_α and I_α^B at levels $\alpha = 0.9, 0.95, 0.99$ are reported for $n = 100, 300$ and $\rho = 0, \pm 0.2, \pm 0.8$.

(n, ρ)	$I_{0.9}$	$I_{0.9}^B$	$I_{0.95}$	$I_{0.95}^B$	$I_{0.99}$	$I_{0.99}^B$
$(100, 0)$	0.337	0.332	0.403	0.394	0.529	0.515
$(100, 0.2)$	0.327	0.322	0.391	0.383	0.515	0.501
$(100, -0.2)$	0.327	0.322	0.390	0.382	0.515	0.499
$(100, 0.8)$	0.148	0.148	0.177	0.177	0.235	0.236
$(100, -0.8)$	0.147	0.147	0.176	0.175	0.234	0.231
$(300, 0)$	0.192	0.190	0.229	0.227	0.302	0.298
$(300, 0.2)$	0.186	0.185	0.222	0.220	0.293	0.289
$(300, -0.2)$	0.186	0.185	0.222	0.220	0.293	0.288
$(300, 0.8)$	0.083	0.082	0.099	0.098	0.130	0.129
$(300, -0.8)$	0.083	0.082	0.099	0.098	0.130	0.128

fire losses over the period 1980 to 1990. They have been adjusted for inflation to reflect 1985 values and are expressed in millions of Danish Kroner. Here we consider the first two variables: loss to building and loss to contents; see Figure 1 below. For computing I_α^B, we draw 1,000 bootstrap samples as before. We find that $\hat{\rho}_n^s = 0.1411$, $I_{0.9}^B = (0.0959, \quad 0.1866)$, $I_{0.95}^B = (0.0897, \quad 0.1942)$, $I_{0.9} = (0.0962, \quad 0.1862)$ and $I_{0.95} = (0.0882, \quad 0.1952)$, which show that the proposed jackknife empirical likelihood method produces similar interval length as the bootstrap method. Both intervals indicate the Spearman's rho is positive, which means that the loss to contents is positively correlated with the loss to profits.

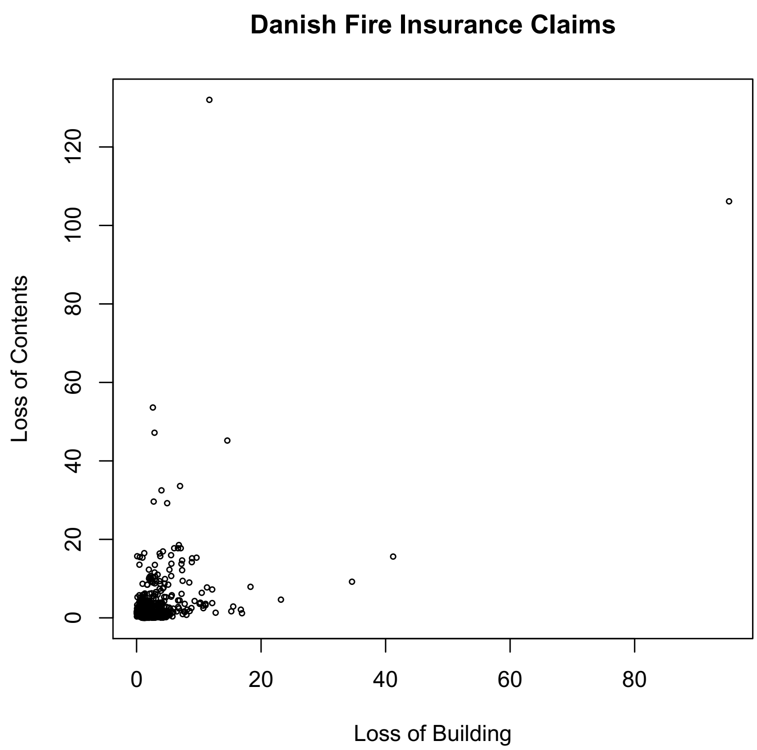

Figure 3.1: Scatterplot of the Danish fire insurance data.

3.3.4 Proofs

Before proving Theorem 3.3.1, we show the following two lemmas.

Lemma 3.3.2. *Under conditions of Theorem 3.3.1, we have*

$$\sqrt{n}\{\tfrac{1}{n}\textstyle\sum_{i=1}^{n} Z_i - \rho^s\} \quad \xrightarrow{d} 12\int_0^1\int_0^1\{W(x,y) - C_1(x,y)W(x,1) - C_2(x,y)W(1,y)\}dxdy$$
$$\stackrel{d}{=} N(0,\sigma^2)$$

as $n \to \infty$, where

$$\sigma^2 = \mathbb{E}[(12\int_0^1\int_0^1\{W(x,y) - C_1(x,y)W(x,1) - C_2(x,y)W(1,y)\}dxdy)^2].$$

Proof. For $i = 1, \cdots, n$, write

$$\begin{aligned}
& Z_i - \rho^s \\
= \;& n\hat{\rho}_n^s - (n-1)\hat{\rho}_{n,i}^s - \rho^s \\
= \;& 12\sum_{j=1}^{n}(F_n(X_j) - 1/2)(G_n(Y_j) - 1/2) - 12\sum_{j\neq i}(F_{n,i}(X_j) - 1/2)(G_{n,i}(Y_j) - 1/2) - \rho^s \\
= \;& 12\sum_{j\neq i}[(F_n(X_j) - 1/2)(G_n(Y_j) - 1/2) - (F_{n,i}(X_j) - 1/2)(G_{n,i}(Y_j) - 1/2)] \\
& +(12(F_n(X_i) - 1/2)(G_n(Y_i) - 1/2) - \rho^s) \\
= \;& 12\left(\sum_{j\neq i}(F_n(X_j) - F_{n,i}(X_j))(G_n(Y_j) - 1/2) + \sum_{j\neq i}(G_n(Y_j) - G_{n,i}(Y_j))(F_n(X_j) - 1/2)\right. \\
& \left. + \sum_{j\neq i}(F_n(X_j) - F_{n,i}(X_j))(G_n(Y_j) - G_{n,i}(Y_j))\right) + (12(F_n(X_i) - 1/2)(G_n(Y_i) - 1/2) - \rho^s) \\
= \;& 12\left(\sum_{j\neq i}(F_n(X_j) - F_{n,i}(X_j))(G_n(Y_j) - 1/2) + \sum_{j\neq i}(G_n(Y_j) - G_{n,i}(Y_j))(F_n(X_j) - 1/2)\right) \\
& +O(1/n) + (12(F_n(X_i) - 1/2)(G_n(Y_i) - 1/2) - \rho^s) \\
= \;& 12\{\underbrace{\sum_{j=1}^{n}(F_n(X_j) - F_{n,i}(X_j))(G_n(Y_j) - 1/2)}_{V_{1,i}} + \underbrace{\sum_{j=1}^{n}(G_n(Y_j) - G_{n,i}(Y_j))(F_n(X_j) - 1/2)}_{V_{2,i}} \\
& + \underbrace{((F_n(X_i) - 1/2)(G_n(Y_i) - 1/2) - \rho^s/12)}_{V_{3,i}}\} + O(1/n).
\end{aligned}$$

Thus

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_i - \rho^s) = \frac{12}{\sqrt{n}}\sum_{i=1}^{n}(V_{i,1} + V_{i,2} + V_{i,3}) + O(1/\sqrt{n}).$$

We already have

$$\begin{aligned}\frac{12}{\sqrt{n}}\sum_{i=1}^{n}V_{3,i} &= \sqrt{n}(\hat{\rho}_n^s-\rho^s)\\ &\xrightarrow{d} 12\int_0^1\int_0^1\{W(x,y)-C_1(x,y)W(x,1)-C_2(x,y)W(1,y)\}dxdy.\end{aligned}$$

It is easy to check that

$$\begin{aligned}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_{1,i} &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{n}(F_n(X_j)-F_{n,i}(X_j))(G_n(Y_j)-1/2)\\ &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(G_n(Y_j)-1/2)\sum_{i=1}^{n}(F_n(X_j)-F_{n,i}(X_j))\\ &= \frac{1}{\sqrt{n}}\sum_{j=1}^{n}(G_n(Y_j)-1/2)\times 0\\ &= 0.\end{aligned}$$

Similarly

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_{2,i}=0.$$

Thus it follows from the above equations that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_i-\rho^s)\xrightarrow{d}12\int_0^1\int_0^1\{W(x,y)-C_1(x,y)W(x,1)-C_2(x,y)W(1,y)\}dxdy.$$

□

Lemma 3.3.3. *Under conditions of Theorem 3.3.1, we have*

$$\frac{1}{n}\sum_{i=1}^{n}(Z_i-\rho^s)^2\xrightarrow{p}\sigma^2 \quad as \quad n\to\infty.$$

Proof. Write

$$\begin{aligned}&\sigma^2\\ =\ & \mathbb{E}[(12\int_0^1\int_0^1\{W(x,y)-C_1(x,y)W(x,1)-C_2(x,y)W(1,y)\}dxdy)^2]\\ =\ & 144\mathbb{E}[(\underbrace{\iint W(x,y)dxdy}_{A_1}-\underbrace{\iint C_1(x,y)W(x,1)dxdy}_{A_2}-\underbrace{\iint C_2(x,y)W(1,y)dxdy}_{A_3})^2]\\ =\ & 144\mathbb{E}[A_1^2+A_2^2+A_3^2-2A_1A_2-2A_1A_3+2A_2A_3],\end{aligned}$$

where by convention we use $\int = \int_0^1$. Using (3.33) we have

$$\begin{aligned}\mathbb{E}(A_1^2) &= \iiiint (C(x_1 \wedge x_2, y_1 \wedge y_2) - C(x_1, y_1)C(x_2, y_2))dx_1 dx_2 dy_1 dy_2 \\ &= 4 \iint C(x,y)(1-x)(1-y)dxdy - \left(\iint C(x,y)dxdy\right)^2,\end{aligned} \tag{3.35}$$

$$\mathbb{E}(A_2^2) = \iiiint C_1(x_1, y_1)C_1(x_2, y_2)(x_1 \wedge x_2 - x_1 x_2)dx_1 dx_2 dy_1 dy_2, \tag{3.36}$$

and

$$\mathbb{E}(A_3^2) = \iiiint C_2(x_1, y_1)C_2(x_2, y_2)(y_1 \wedge y_2 - y_1 y_2)dx_1 dx_2 dy_1 dy_2. \tag{3.37}$$

By integration by parts, we have

$$\begin{aligned}&\iiint C_1(x_1, y_1)C(x_1, y_2)dx_1 dy_1 dy_2 \\ =\ &\iint C(1, y_1)C(1, y_2)dy_1 dy_2 - \iiint C_1(x_1, y_2)C(x_1, y_1)dx_1 dy_1 dy_2 \\ =\ &\tfrac{1}{4} - \iiint C_1(x_1, y_2)C(x_1, y_1)dx_1 dy_1 dy_2,\end{aligned}$$

which implies that

$$\iiint C_1(x_1, y_1)C(x_1, y_2)dx_1 dy_1 dy_2 = \frac{1}{8}. \tag{3.38}$$

Similarly,

$$\begin{aligned}&\iiint x_1 C_1(x_1, y_1)C(x_1, y_2)dx_1 dy_1 dy_2 \\ =\ &\iint C(1, y_1)C(1, y_2)dy_1 dy_2 \\ &- \iiint (x_1 C_1(x_1, y_1)C(x_1, y_2) - C(x_1, y_1)C(x_1, y_2))dx_1 dy_1 dy_2,\end{aligned}$$

which implies that

$$2 \iiint x_1 C_1(x_1, y_1)C(x_1, y_2)dx_1 dy_1 dy_2 + \iiint C(x_1, y_1)C(x_1, y_2)dx_1 dy_1 dy_2 = \frac{1}{4}. \tag{3.39}$$

It follows from (3.38), (3.39) and (3.33) that

$$\begin{aligned}
& \mathbb{E}(A_1A_2) \\
= & \iiint C_1(x_1,y_1)(C(x_1\wedge x_2,y_2)-x_1C(x_2,y_2))dx_1dx_2dy_1dy_2 \\
= & \iiint C_1(x_1,y_1)C(x_1,y_2)dx_1dy_1dy_2 - \iiint x_1C_1(x_1,y_1)C(x_1,y_2)dx_1dy_1dy_2 \\
& - \iiint C(x_1,y_1)C(x_1,y_2)dx_1dy_1dy_2 + \left(\iint C(x,y)dxdy\right)^2 \\
= & -\tfrac{1}{8} + \iiint x_1C_1(x_1,y_1)C(x_1,y_2)dx_1dy_1dy_2 + \left(\iint C(x,y)dxdy\right)^2 .
\end{aligned} \tag{3.40}$$

Using the same arguments, we can show that

$$\mathbb{E}(A_1A_3) = -\frac{1}{8} + \iiint y_1C_2(x_1,y_1)C(x_2,y_1)dx_1dx_2dy_1 + \left(\iint C(x,y)dxdy\right)^2, \tag{3.41}$$

and

$$\mathbb{E}(A_2A_3) = \iiiint C_1(x_1,y_1)C_2(x_2,y_2)(C(x_1,y_2)-x_1y_2)dx_1dx_2dy_1dy_2. \tag{3.42}$$

Note that

$$\begin{aligned}
\tfrac{1}{n}\textstyle\sum_{i=1}^n (Z_i-\rho^s)^2 &= \tfrac{144}{n}\textstyle\sum_{i=1}^n (V_{1,i}+V_{2,i}+V_{3,i}+O(1/n))^2 \\
&= \tfrac{144}{n}\textstyle\sum_{i=1}^n (V_{1,i}+V_{2,i}+V_{3,i})^2 + O(1/n)
\end{aligned} \tag{3.43}$$

since $V_{1,i}$, $V_{2,i}$ and $V_{3,i}$ are uniformly bounded for $i=1,\cdots,n$. A straightforward calculation shows that

$$\begin{aligned}
& \tfrac{1}{n}\textstyle\sum_{i=1}^n V_{1,i}^2 \\
= & \tfrac{1}{n}\textstyle\sum_{i=1}^n \left(\sum_{j=1}^n (F_n(X_j)-F_{n,i}(X_j))(G_n(Y_j)-1/2)\right)^2 \\
= & \tfrac{1}{n}\textstyle\sum_{i=1}^n\sum_{j=1}^n\sum_{k=1}^n (F_n(X_j)-F_{n,i}(X_j))(G_n(Y_j)-1/2)(F_n(X_k)-F_{n,i}(X_k))(G_n(Y_k)-1/2) \\
= & \tfrac{1}{n}\textstyle\sum_{j=1}^n\sum_{k=1}^n (G_n(Y_j)-1/2)(G_n(Y_k)-1/2)\sum_{i=1}^n (F_n(X_j)-F_{n,i}(X_j))(F_n(X_k)-F_{n,i}(X_k)) \\
= & \tfrac{1}{(n-1)^2}\textstyle\sum_{j=1}^n\sum_{k=1}^n (G_n(Y_j)-1/2)(G_n(Y_k)-1/2)(F_n(X_j\wedge X_k)-F_n(X_j)F_n(X_k)) \\
\xrightarrow{p} & \iiiint (y_1-\tfrac{1}{2})(y_2-\tfrac{1}{2})(x_1\wedge x_2 - x_1x_2)C(dx_1,dy_1)C(dx_2,dy_2) \\
= & \iint \tfrac{1}{4}(x_1\wedge x_2-x_1x_2)dx_1dx_2 - \iiint (x_1\wedge x_2-x_1x_2)C_1(x_1,y_1)dx_1dx_2dy_1 \\
& + \iiiint C_1(x_1,y_1)C_1(x_2,y_2)(x_1\wedge x_2-x_1x_2)dx_1dx_2dy_1dy_2 \\
= & \tfrac{1}{48} + \tfrac{1}{2}\iint C(x,y)dxdy - \iint xC(x,y)dxdy + \mathbb{E}(A_2^2)
\end{aligned} \tag{3.44}$$

as $n \to \infty$. Similarly, we can show that

$$\frac{1}{n}\sum_{i=1}^{n} V_{2,i}^2 \xrightarrow{p} \frac{1}{48} + \frac{1}{2}\iint C(x,y)dxdy - \iint yC(x,y)dxdy + \mathbb{E}(A_3^2), \tag{3.45}$$

$$\begin{aligned}
& \tfrac{1}{n}\textstyle\sum_{i=1}^{n} V_{3,i}^2 \\
\xrightarrow{p} \quad & \textstyle\iint (x-\frac{1}{2})^2(y-\frac{1}{2})^2 C(dx,dy) - (\rho^s/12)^2 \\
= \quad & \textstyle\iint (x-\frac{1}{2})(y-\frac{1}{2})C(x,y)dxdy - \frac{1}{48} - \left(\iint C(x,y)dxdy - \frac{1}{4}\right)^2 \\
= \quad & \textstyle 4\iint (x-1)(y-1)C(x,y)dxdy + \iint 2(x+y)C(x,y)dxdy - 3\iint C(x,y)dxdy \\
& \textstyle -\frac{1}{48} - \left(\iint C(x,y)dxdy\right)^2 + \frac{1}{2}\iint C(x,y)dxdy - \frac{1}{16} \\
= \quad & \textstyle \mathbb{E}(A_1^2) + \iint 2(x+y)C(x,y)dxdy - \frac{5}{2}\iint C(x,y)dxdy - \frac{1}{12},
\end{aligned} \tag{3.46}$$

$$\begin{aligned}
& \tfrac{1}{n}\textstyle\sum_{i=1}^{n} V_{1,i}V_{2,i} \\
\xrightarrow{p} \quad & \textstyle\iiiint (x_2-\frac{1}{2})(y_1-\frac{1}{2})(C(x_1,y_2)-x_1y_2)C(dx_1,dy_1)C(dx_2,dy_2) \\
= \quad & \textstyle\iint \frac{1}{4}(C(x_1,y_2)-x_1y_2)dx_1dy_2 - \iiint \frac{1}{2}C_2(x_2,y_2)(C(x_1,y_2)-x_1y_2)dx_1dx_2dy_2 \\
& \textstyle -\iiint \frac{1}{2}C_1(x_1,y_1)(C(x_2,y_1)-x_2y_1)dx_1dy_1dy_2 \\
& \textstyle +\iiiint C_1(x_1,y_1)C_2(x_2,y_2)(C(x_1,y_2)-x_1y_2)dx_1dx_2dy_1dy_2 \\
= \quad & \textstyle \frac{3}{16} - \frac{1}{4}\iint C(x,y)dxdy - \frac{1}{2}\iint C_2(x_2,y_2)C(x_1,y_2)dx_1dx_2dy_2 \\
& \textstyle -\frac{1}{2}\iint C_1(x_1,y_1)C(x_2,y_1)dx_1dy_1dy_2 + \mathbb{E}(A_2A_3),
\end{aligned} \tag{3.47}$$

$$
\begin{aligned}
& \tfrac{1}{n}\textstyle\sum_{i=1}^n V_{1,i}V_{3,i} \\
= \quad & \tfrac{1}{n}\textstyle\sum_{j=1}^n\sum_{i=1}^n (F_n(X_j)-F_{n,i}(X_j))[(F_n(X_i)-\tfrac{1}{2})(G_n(Y_i)-\tfrac{1}{2})-\rho^2/12](G_n(Y_j)-\tfrac{1}{2}) \\
= \quad & \tfrac{1}{n^2}\textstyle\sum_{j=1}^n\sum_{i=1}^n [I(X_i\le X_j)-F_n(X_j)][(F_n(X_i)-\tfrac{1}{2})(G_n(Y_i)-\tfrac{1}{2})-\rho^2/12](G_n(Y_j)-\tfrac{1}{2}) \\
\xrightarrow{p} \quad & \textstyle\iiiint_{x_2=x_1}^1 (y_2-\tfrac{1}{2})C(dx_2,dy_2)[(x_1-\tfrac{1}{2})(y_1-\tfrac{1}{2})-\rho^s/12]C(dx_1,dy_1) \\
= \quad & \textstyle\iiint (y_2-\tfrac{1}{2})(1-C_2(x_1,y_2))dy_2[(x_1-\tfrac{1}{2})(y_1-\tfrac{1}{2})-\rho^s/12]C(dx_1,dy_1) \\
= \quad & \textstyle\iiint -(y_2-\tfrac{1}{2})C_2(x_1,y_2)dy_2(x_1-\tfrac{1}{2})(y_1-\tfrac{1}{2})C(dx_1,dy_1)-(\rho^s/12)^2 \\
= \quad & \textstyle\iint \left(\int C(x_1,y_2)dy_2-\tfrac{1}{2}x_1\right)(x_1-\tfrac{1}{2})(y_1-\tfrac{1}{2})C(dx_1,dy_1)-(\rho^s/12)^2 \\
= \quad & \tfrac{1}{2}\textstyle\iint x_1C(x_1,y_2))dx_1dy_2-\tfrac{1}{4}\iint C(x,y)dxdy-\tfrac{1}{12}+\tfrac{1}{16}+\iint x_1C_1(x_1,y_1)(x_1-\tfrac{1}{2})dx_1dy_1 \\
& -\textstyle\iiint C_1(x_1,y_1)C(x_1,y_2)(x_1-\tfrac{1}{2})dx_1dy_1dy_2-(\rho^s/12)^2 \\
= \quad & \tfrac{1}{2}\textstyle\iint xC(x,y)dxdy-\tfrac{1}{48}+\tfrac{1}{4}-\iint xC(x,y)dxdy-\tfrac{1}{8}-\iiint x_1C_1(x_1,y_1)C(x_1,y_2)dx_1dy_1dx_2 \\
& +\tfrac{1}{2}\textstyle\iiint C_1(x_1,y_1)C(x_1,y_2)dx_1dy_1dy_2-\left(\iint C(x,y)dxdy-\tfrac{1}{4}\right)^2 \\
= \quad & \tfrac{1}{24}-\tfrac{1}{2}\textstyle\iint xC(x,y)dxdy-\iiint x_1C_1(x_1,y_1)C(x_1,y_2)dx_1dy_1dx_2 \\
& +\tfrac{1}{2}\textstyle\iiint C_1(x_1,y_1)C(x_1,y_2)dx_1dy_1dy_2+\tfrac{1}{2}\iint C(x,y)dxdy-\left(\iint C(x,y)dxdy\right)^2 \\
= \quad & -\tfrac{1}{12}-\tfrac{1}{2}\textstyle\iint xC(x,y)dxdy+\tfrac{1}{2}\iiint C_1(x_1,y_1)C(x_1,y_2)dx_1dy_1dy_2 \\
& +\tfrac{1}{2}\textstyle\iint C(x,y)dxdy-\mathbb{E}(A_1A_2),
\end{aligned} \tag{3.48}
$$

and

$$
\begin{aligned}
\tfrac{1}{n}\textstyle\sum_{i=1}^n V_{2,i}V_{3,i} \xrightarrow{p} \quad & -\tfrac{1}{12}-\tfrac{1}{2}\textstyle\iint yC(x,y)dxdy+\tfrac{1}{2}\iiint C_2(x_1,y_1)C(x_2,y_1)dx_1dy_1dy_2 \\
& +\tfrac{1}{2}\textstyle\iint C(x,y)dxdy-\mathbb{E}(A_1A_3).
\end{aligned} \tag{3.49}
$$

Hence, it follows from (3.35)–(3.37), (3.40)–(3.49) that

$$
\begin{aligned}
& \frac{1}{n}\sum_{i=1}^n (V_{1,i}+V_{2,i}+V_{3,i})^2 \\
= \quad & \frac{1}{n}\sum_{i=1}^n (V_{1,i}^2+V_{2,i}^2+V_{3,i}^2+2V_{1,i}V_{2,i}+2V_{1,i}V_{3,i}+2V_{2,i}V_{3,i}) \\
\xrightarrow{p} \quad & \mathbb{E}(A_1^2)+\mathbb{E}(A_2^2)+\mathbb{E}(A_3^2)-2\mathbb{E}(A_1A_2)-2\mathbb{E}(A_1A_3)+\mathbb{E}(A_2A_3),
\end{aligned}
$$

i.e.,

$$
\frac{1}{n}\sum_{i=1}^n (Z_i-\rho^s)^2 \xrightarrow{p} \sigma^2.
$$

□

Proof of Theorem 3.3.1. Since $V_{1,i}$, $V_{2,i}$ and $V_{3,i}$ defined in the proof of Lemma 1 are uniformly bounded for $i = 1, \cdots, n$, we have $\sup_{1 \le i \le n} |Z_i|$ is bounded. Hence, using the standard arguments in the empirical likelihood method (see Chapter 11 of Owen [73]), Lemmas 3.3.2 and 3.3.3, we obtain that

$$-2 \log L(\rho^s) = \frac{\{\sum_{i=1}^n (Z_i - \rho^s)\}^2}{\sum_{i=1}^n (Z_i - \rho^s)^2} + o_p(1) \xrightarrow{d} \chi^2(1).$$

□

3.4 Interval Estimation for Parametric Copulas

For fitting a parametric copula to multivariate data, a popular way is to employ the so-called pseudo maximum likelihood estimation proposed by Genest, Ghoudi and Rivest [42]. Although interval estimation can be obtained via estimating the asymptotic covariance of the pseudo maximum likelihood estimation, we propose a jackknife empirical likelihood method to construct confidence regions for the parameters without estimating any additional quantities such as the asymptotic covariance. A simulation study shows the advantages of the new method in case of strong dependence or having more than one parameter involved.

3.4.1 Introduction

Let $\mathbf{X}_1 = (X_{1,1}, \cdots, X_{1,d})^T, \cdots, \mathbf{X}_n = (X_{n,1}, \cdots, X_{n,d})^T$ be independent random vectors with common distribution function F and continuous marginal distributions $F_1, \cdots, F_d$. Then the copula of $\mathbf{X}_1$ is defined as

$$C(x_1, \cdots, x_d) = F(F_1^-(x_1), \cdots, F_d^-(x_d)) \tag{3.50}$$

for $0 \le x_1, \cdots, x_d \le 1$, where F_j^- denotes the inverse of F_j. Since the copula is independent of marginals, it becomes a more or less standard tool in modeling dependence in risk management. Many research papers and review papers have appeared in the literature with particular applications in insurance, finance and risk management; see references in Haug, Klüppelberg and Peng [44].

For fitting a family of parametric copulas $\{C(\cdot;\theta) : \theta \in \Theta \subset \mathbb{R}^q\}$ to a data set, a popular semi-parametric estimation is the so-called pseudo maximum likelihood estimation proposed by Genest, Ghoudi and Rivest [42]. That is, $\hat{\theta} = \arg\max \bar{L}(\theta)$, where $\bar{L}(\theta)$ is the pseudo likelihood function for θ defined as

$$\bar{L}(\theta) = \prod_{i=1}^{n} c(\hat{F}_1(X_{i,1}), \cdots, \hat{F}_d(X_{i,d}); \theta), \tag{3.51}$$

where $c(\cdot;\theta)$ denotes the density function of the parametric copula family $C(\cdot;\theta)$, and $\hat{F}_j(x) = \frac{1}{n+1}\sum_{i=1}^{n} I(X_{i,j} \leq x)$ for $j = 1, \cdots, d$. Alternatively, the pseudo maximum likelihood estimator can be defined as a root of the score equations

$$\sum_{i=1}^{n} \mathbf{l}(\hat{F}_1(X_{i,1}), \cdots, \hat{F}_d(X_{i,d}); \theta) = 0, \tag{3.52}$$

where $\mathbf{l}(x;\theta) = (l_1(x;\theta), \cdots, l_q(x;\theta))$ and $l_j(x;\theta) = \frac{\partial}{\partial\theta_j}\log c(x;\theta)$. Since $l_j((x_1, \cdots, x_d);\theta)$ may be infinity when one of $x_i's$ is one, we use $\hat{F}_j(x)$ as the empirical distribution function in this section instead of $\frac{1}{n}\sum_{i=1}^{n} I(X_{i,j} \leq x)$ to ensure $\max_{1\leq i\leq n} \hat{F}_j(X_{i,j}) < 1$. The asymptotic distribution of the above pseudo maximum likelihood estimator and a consistent estimator for the asymptotic variance are given in Genest, Ghoudi and Rivest [42]. Since the asymptotic covariance of the pseudo maximum likelihood estimator is complicated and involves the contribution from both the copula and marginals, it is of importance to seek a more efficient way to construct confidence regions for the parameters θ without estimating the asymptotic covariance.

In this section, we investigate the possibility of employing empirical likelihood methods. A key step in applying the empirical likelihood method is to formulate the nonparametric likelihood function. This is commonly done via estimating equations. Since the pseudo maximum likelihood estimator is a solution to the score equations (3.52), one may apply the method in Qin and Lawless [82] to construct confidence regions for β by defining the empirical likelihood function as

$$\begin{aligned} L_1(\theta) \ &= \sup\{\textstyle\prod_{i=1}^{n}(np_i) : p_1 \geq 0, \cdots, p_n \geq 0, \sum_{i=1}^{n} p_i = 1, \\ &\quad \textstyle\sum_{i=1}^{n} p_i \mathbf{l}(\hat{F}_1(X_{i,1}), \cdots, \hat{F}_d(X_{i,d}); \theta) = 0\}. \end{aligned}$$

Unfortunately, this likelihood function can not catch the variances of $\hat{F}_j's$ and thus Wilks' Theorem fails, i.e., $-2\log L_1(\theta)$ does not converge in distribution to a chi-square limit, due to the nonlinearity of $\mathbf{l}(\cdot;\theta)$. A common way to deal with nonlinear functionals is to linearize it before employing the empirical likelihood method; see Chen, Peng and Zhao [19] and Molanes-Lopez, Van Keilegom and Veraverbeke [66] for constructing confidence intervals for copula at a particular point. However, it remains unknown on how to linearize the score questions (3.52). In this section, we apply the jackknife empirical likelihood method to construct confidence intervals/regions for a parametric copula. When the copula is estimated nonparametrically, Peng, Qi and Van Keilegom [75] proposed a smoothed jackknife empirical likelihood method to construct confidence intervals for a copula at a fixed point.

We organize this section as follows. Section 3.4.2 presents the methodology and main results. A simulation study and a real data analysis are given in Section 3.4.3. All proofs are put in Section 3.4.4.

3.4.2 Methodology

In order to formulate an empirical likelihood function with $\hat{F}_j's$ taken into account, we consider the estimators $\frac{1}{n}\sum_{i=1}^n \mathbf{l}(\hat{F}_1(X_{i,1}),\cdots,\hat{F}_d(X_{i,d});\theta)$ and follow the idea in Jing, Yuan and Zhou [47] to construct a jackknife sample first and then apply the empirical likelihood method to the jackknife sample. Since the considered estimators are not U-statistics, we formulate the jackknife sample in a way different from that in Jing, Yuan and Zhou [47]. The details are as follows.

Like the definition of $\hat{F}_j(x)$ in the introduction, we define $\hat{F}_{j,-i}(x)=\frac{1}{n}\sum_{k=1,k\neq i}^n I(X_{k,j}\leq x)$ instead of $\frac{1}{n-1}\sum_{k=1,k\neq i}^n I(X_{k,j}\leq x)$ for $j=1,\cdots,d$ and $i=1,\cdots,n$. Further we formulate the jackknife sample as $\{\mathbf{Z}_i(\theta)=(Z_{i,1}(\theta),\cdots,Z_{i,q}(\theta))^T\}_{i=1}^n$, where

$$Z_{i,j}(\theta)=\sum_{k=1}^n l_j(\hat{F}_1(X_{k,1}),\cdots,\hat{F}_d(X_{k,d});\theta)-\sum_{k=1,k\neq i}^n l_j(\hat{F}_{1,-i}(X_{k,1}),\cdots,\hat{F}_{d,-i}(X_{k,d});\theta)$$

for $i=1,\cdots,n$ and $j=1,\cdots,q$. Based on this jackknife sample, we define the jackknife empirical likelihood function as

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

By the Lagrange multiplier technique, i.e., maximizing $\sum_{i=1}^{n} \log(np_i) + a(\sum_{i=1}^{n} p_i - 1) + b^T \sum_{i=1}^{n} p_i \mathbf{Z}_i(\theta)$ with respect to $p_1, \cdots, p_n, a, b$, we have $p_i = n^{-1}\{1+\lambda^T \mathbf{Z}_i(\theta)\}^{-1}$,

$$-2 \log L(\theta) = 2 \sum_{i=1}^{n} \log\{1 + \lambda^T \mathbf{Z}_i(\theta)\},$$

where $\lambda = (\lambda_1(\theta), \cdots, \lambda_q(\theta))^T$ satisfies

$$\sum_{i=1}^{n} \frac{\mathbf{Z}_i(\theta)}{1 + \lambda^T \mathbf{Z}_i(\theta)} = 0. \tag{3.53}$$

See Owen [71] for more details.

Before showing that Wilks' Theorem holds for the above jackknife empirical likelihood method, we list some regularity conditions. Throughout we use θ_0 to denote the true value of θ and define $r(u) = u(1-u)$.

A1) There exist some constants $0 < \alpha_1 < 1/2$ and $M_1 > 0$ such that, uniformly for $0 < u_1, \cdots, u_d < 1$,

$$|l_j(u_1, \cdots, u_d; \theta_0)| \leq M_1 \prod_{i=1}^{d} r(u_i)^{-\alpha_1},$$

$$|l_j^{(s)}(u_1, \cdots, u_d; \theta_0)| := |\frac{\partial}{\partial u_s} l_j(u_1, \cdots, u_d; \theta_0)| \leq M_1 r(u_s)^{-1} \prod_{i=1}^{d} r(u_i)^{-\alpha_1},$$

$$|l_j^{(sm)}(u_1, \cdots, u_d; \theta_0)| := |\frac{\partial^2}{\partial u_s \partial u_m} l_j(u_1, \cdots, u_d; \theta_0)| \leq M_1 r(u_s)^{-1} r(u_m)^{-1} \prod_{i=1}^{d} r(u_i)^{-\alpha_1},$$

and

$$\mathbb{E}[l_j^2(F_1(X_{1,1}), \cdots, F_d(X_{1,d}); \theta_0)] \leq M_1$$

for $j = 1, \cdots, q$ and $s, m = 1, \cdots, d$.

A2) For a given $0 < \alpha_2 < 1/2$, there exist some constants $0 < \alpha_3 < 1/2$ and $M_2 > 0$ such that, uniformly for $0 < u_1, \cdots, u_d < 1$

$$\int \cdots \int_{[0,1]^{d-1}} \prod_{i=1}^{d} r(u_i)^{-\alpha_2} c(u_1, \cdots, u_d; \theta_0)\, du_1 \cdots du_{s-1} du_{s+1} \cdots du_d \le M_2 r(u_s)^{-\alpha_3}$$

for $s = 1, \cdots, d$, and

$$\int \cdots \int_{[0,1]^{d-2}} \prod_{i=1}^{d} r(u_i)^{-\alpha_2} c(u_1, \cdots, u_d; \theta_0)\, du_1 \cdots du_{s-1} du_{s+1} \cdots du_{m-1} du_{m+1} \cdots du_d$$

$$\le M_2 r(u_s)^{-\alpha_3} r(u_m)^{-\alpha_3}$$

for $1 \le s < m \le d$.

Remark 3.4.1. Commonly used copulas such as Clayton, Frank, Gumbel, normal and t copulas satisfy A1) and A2).

Theorem 3.4.1. *Under conditions A1) and A2), we have*

$$-2 \log L(\theta_0) \xrightarrow{d} \chi^2(q) \quad as \quad n \to \infty.$$

Based on the above theorem, an empirical likelihood confidence interval/region for θ_0 with level ξ is $\{\theta : -2 \log L(\theta) \le \chi^2_{q,\xi}\}$, where $\chi^2_{q,\xi}$ is the ξ-th quantile of a chi-square distribution with q degrees of freedom.

3.4.3 Simulation study and data analysis

Simulation study. In this subsection, we examine the finite behavior of the proposed jackknife empirical likelihood method and compare it with the normal approximation method.

We draw $10,000$ random samples with size $n = 300$ from the Clayton copula $C(u_1, \cdots, u_d; \theta) = (1-d+u_1^{-\theta}+\cdots+u_d^{-\theta})^{-1/\theta}$, bivariate normal copula $C(u_1, u_2; \theta) = \Phi_\theta(\Phi^{-1}(u_1), \Phi^{-1}(u_2))$, where Φ denotes the standard normal distribution and Φ_θ denotes the standard bivariate normal distribution with correlation θ, and bivariate t-copula with $\theta = (\rho, \nu)$, where $\rho \in (-1, 1)$ and $\nu > 0$.

We employ the 'copula' package in R to calculate the pseudo maximum likelihood estimator and its asymptotic variance so as to construct a confidence interval/region for θ, denoted by NAM. We also denote the proposed jackknife empirical likelihood method by JELM. For calculating the score equations of the bivariate t-copula, we use the formulas in Dakovic and Czado [23] with some typos corrected. More specifically, i) the integrals in (7) and (8) have to be divided by 2; ii) x^2 in (8) is x_i^2; iii) the term $\frac{\nu+2}{2\nu}$ in the formula for $\frac{\partial l}{\partial \nu}(u_1, u_2)$ after (11) is $\frac{\nu-2}{2\nu}$. Note that equations (7), (8) and (11) mean those in Dakovic and Czado [23].

In Tables 3.7–3.9 we report coverage probabilities for these two methods with levels 0.9 and 0.95. Note that for the t-copula, the 'copula' package in R does not provide asymptotic covariance. Hence we only report the coverage probabilities for the proposed jackknife empirical likelihood method in this case. From these tables, we observe that (i) the proposed jackknife empirical likelihood method works better than the normal approximation methods for large θ in the Clayton and normal copula (i.e., strong dependence); (ii) results for the cases of $d = 4, \theta = 10, 15$ in Table 1 indicate that the asymptotic variance for the Clayton copula given in the 'copula' package may be problematic when the dimension is large; (iii) the proposed jackknife empirical likelihood method performs well for t-copulas, where the asymptotic variance in the **copula** package is not available.

Data analysis. We apply the proposed method to an insurance company data on losses and ALAEs. This particular data set has been analyzed by Frees and Valdez [39], Klugman and Parsa [58], Dupuis and Jones [30], and Peng [74]. Like Klugman and Parsa [58], we fit the Frank copula

$$C(u, v; \alpha) = -\frac{1}{\alpha} \log\{1 + \frac{(e^{-\alpha u} - 1)(e^{-\alpha v} - 1)}{e^{-\alpha} - 1}\}.$$

Using the `copula` package in `R`, we find the pseudo maximum likelihood estimator for α is 2.992 and the confidence intervals based on the normal approximation method

Table 3.7: Empirical coverage probabilities are reported for Clayton copulas with dimension $d = 2, 4$.

(d, θ)	JELM Level 0.9	NAM Level 0.9	JELM Level 0.95	NAM Level 0.95
(2,0.2)	0.8846	0.8875	0.9363	0.9417
(2,1)	0.8902	0.8950	0.9430	0.9448
(2,10)	0.9114	0.9162	0.9563	0.9566
(2,15)	0.9184	0.9160	0.9628	0.9582
(4,0.2)	0.8750	0.8734	0.9336	0.9331
(4,1)	0.8767	0.8791	0.9295	0.9294
(4,10)	0.9167	0.9418	0.9573	0.9703
(4,15)	0.9211	0.9519	0.9604	0.9781

Table 3.8: Empirical coverage probabilities are reported for the bivariate normal copula.

θ	JELM Level 0.9	NAM Level 0.9	JELM Level 0.95	NAM Level 0.95
0.2	0.8847	0.8851	0.9438	0.9434
0.5	0.8864	0.8750	0.9411	0.9314
0.8	0.8880	0.8818	0.9393	0.9331

Table 3.9: Empirical coverage probabilities are reported for the bivariate t copula.

$\theta = (\rho, \nu)$	JELM Level 0.9	JELM Level 0.95
(0.2, 3)	0.8853	0.9404
(0.5,3)	0.8874	0.9385
(0.8,3)	0.8945	0.9476
(0.2,8)	0.8808	0.9352
(0.5,8)	0.8861	0.9412
(0.8,8)	0.8878	0.9415

are $(2.694, 3.290)$ and $(2.637, 3.348)$ for levels 90% and 95%, respectively. The proposed jackknife empirical likelihood intervals are calculated to be $(2.702, 3.292)$ and $(2.653, 3.352)$ for levels 90% and 95%, respectively, which are slightly skewed to the right than the normal approximation based intervals.

3.4.4 Proofs

Lemma 3.4.2. *Under conditions of Theorem 3.4.1, we have*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbf{Z}_i(\theta_0)\xrightarrow{d}N(0,\Sigma)\quad as\quad n\to\infty,$$

where $\Sigma=(\sigma_{ij})_{1\le i,j\le q}$,

$$\sigma_{ij} = \mathbb{E}\left[\left(l_i(\mathbf{T}_1;\theta_0)+\sum_{s=1}^{d}W(i,s)\right)\left(l_j(\mathbf{T}_1;\theta_0)+\sum_{s=1}^{d}W(j,s)\right)\right]<\infty,$$

$\mathbf{T}_1=(F_1(X_{1,1}),\cdots,F_d(X_{1,d}))^T$ *and*

$$W(i,s)=\int_0^1\cdots\int_0^1 l_i^{(s)}(u_1,\cdots,u_d;\theta_0)(I(F_s(X_{1,s})\le u_s)-u_s)c(u_1,\cdots,u_d;\theta_0)\,du_1\cdots du_d.$$

Proof. We denote $\mathbf{T}_k=(F_1(X_{k,1}),\cdots,F_d(X_{k,d}))^T$, $\hat{\mathbf{T}}_k=(\hat{F}_1(X_{k,1}),\cdots,\hat{F}_d(X_{k,d}))^T$ and $\hat{\mathbf{T}}_{k,-i}=(\hat{F}_{1,-i}(X_{k,1}),\cdots,\hat{F}_{d,-i}(X_{k,d}))^T$ for $i,k=1,\cdots,n$. Write

$$\begin{aligned}
& Z_{i,j}(\theta_0)\\
= & \ l_j(\hat{\mathbf{T}}_k;\theta_0)+\sum_{k=1,k\ne i}^{n}\{l_j(\hat{\mathbf{T}}_k;\theta_0)-l_j(\hat{\mathbf{T}}_{k,-i};\theta_0)\}\\
= & \ l_j(\hat{\mathbf{T}}_k;\theta_0)+\sum_{k=1,k\ne i}^{n}\sum_{s=1}^{d}l_j^{(s)}(\hat{\mathbf{T}}_k;\theta_0)\{\hat{F}_s(X_{k,s})-\hat{F}_{s,-i}(X_{k,s})\}\\
& +\frac{1}{2}\sum_{k=1,k\ne i}^{n}\sum_{s=1}^{d}\sum_{t=1}^{d}l_j^{(st)}(\mathbf{Y}_{k,i};\theta_0)\{\hat{F}_s(X_{k,s})-\hat{F}_{s,-i}(X_{k,s})\}\{\hat{F}_t(X_{k,t})-\hat{F}_{t,-i}(X_{k,t})\}\\
= & \ l_j(\hat{\mathbf{T}}_k;\theta_0)+\frac{1}{n}\sum_{k=1,k\ne i}^{n}\sum_{s=1}^{d}l_j^{(s)}(\hat{\mathbf{T}}_k;\theta_0)\{I(X_{i,s}\le X_{k,s})-\hat{F}_s(X_{k,s})\}\\
& +\frac{1}{2n^2}\sum_{k=1,k\ne i}^{n}\sum_{s=1}^{d}\sum_{t=1}^{d}l_j^{(st)}(\mathbf{Y}_{k,i};\theta_0)\times\{I(X_{i,s}\le X_{k,s})-\hat{F}_s(X_{k,s})\}\\
& \quad\times\{I(X_{i,t}\le X_{k,t})-\hat{F}_t(X_{k,t})\}\\
=: & \ I_1(i,j)+I_2(i,j)+I_3(i,j), \qquad (3.54)
\end{aligned}$$

where

$$\mathbf{Y}_{k,i} = \beta_k \hat{\mathbf{T}}_k + (1-\beta_k)\hat{\mathbf{T}}_{k,-i}$$

and $\beta_k \in [0,1]$ depending on i and j. Since

$$\sup_{1\le i\le n} \frac{F_s(X_{i,s})}{\hat{F}_s(X_{i,s})} = O_p(1) \quad \text{and} \quad \sup_{1\le i\le n} \frac{1-F_s(X_{i,s})}{1-\hat{F}_s(X_{i,s})} = O_p(1) \tag{3.55}$$

(see (4) in Page 415 of Shorack and Wellner [94]), it follows from A1) that

$$\begin{aligned} & \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n I_2(i,j) \\ = \ & n^{-3/2} \textstyle\sum_{i=1}^n \sum_{k=1}^n \sum_{s=1}^d l_j^{(s)}(\hat{\mathbf{T}}_k;\theta_0)\{I(X_{i,s} \le X_{k,s}) - \hat{F}_s(X_{k,s})\} \\ & -n^{-3/2} \textstyle\sum_{i=1}^n \sum_{s=1}^d l_j^{(s)}(\hat{\mathbf{T}}_i;\theta_0)\{1-\hat{F}_s(X_{i,s})\} \\ = \ & -n^{-3/2} \textstyle\sum_{k=1}^n \sum_{s=1}^d l_j^{(s)}(\hat{\mathbf{T}}_k;\theta_0)\{1-2\hat{F}_s(X_{k,s})\} \\ = \ & O_p(n^{-3/2} \textstyle\sum_{i=1}^n \sum_{s=1}^d r(\hat{F}_s(X_{i,s}))^{-1} \prod_{t=1}^d r(\hat{F}_t(X_{i,t}))^{-\alpha_1}) \\ = \ & O_p(n^{-3/2} \textstyle\sum_{i=1}^n \sum_{s=1}^d r(F_s(X_{i,s}))^{-1} \prod_{t=1}^d r(F_t(X_{i,t}))^{-\alpha_1}). \end{aligned} \tag{3.56}$$

By A2) and choosing $\delta > 1$ and $\delta\alpha_3 < 1/2$, where α_3 is given in A2), we have for any $\epsilon > 0$

$$\begin{aligned} & \mathbb{P}(n^{-3/2} \textstyle\sum_{i=1}^n r(F_s(X_{i,s}))^{-1} \prod_{t=1}^d r(F_t(X_{i,t}))^{-\alpha_1} > \epsilon) \\ \le \ & \mathbb{P}(n^{-3/2} \textstyle\sum_{i=1}^n I(n^{-\delta} \le F_s(X_{i,s}) \le 1-n^{-\delta}) r(F_s(X_{i,s}))^{-1} \prod_{t=1}^d r(F_t(X_{i,t}))^{-\alpha_1} > \epsilon) \\ & +\mathbb{P}(\min_{1\le i\le n} F_s(X_{i,s}) < n^{-\delta}) + \mathbb{P}(\max_{1\le i\le n} F_s(X_{i,s}) > 1-n^{-\delta}) \\ \le \ & (n^{3/2}\epsilon)^{-1} \textstyle\sum_{i=1}^n \mathbb{E}[I(n^{-\delta} \le F_s(X_{i,s}) \le 1-n^{-\delta}) r(F_s(X_{i,s}))^{-1} \prod_{t=1}^d r(F_t(X_{i,t}))^{-\alpha_1}] \\ & +o(1) \\ \le \ & M_2 n^{-1/2}\epsilon^{-1} \mathbb{E}[I(n^{-\delta} \le F_s(X_{1,s}) \le 1-n^{-\delta}) r(F_s(X_{1,s}))^{-1-\alpha_3}] + o(1) \\ \le \ & M_2 n^{-1/2+\delta\alpha_3}\epsilon^{-1} + o(1) \\ = \ & o(1). \end{aligned} \tag{3.57}$$

Therefore, it follows from (3.56) and (3.57) that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n I_2(i,j) = o_p(1) \quad \text{for} \quad j = 1, \cdots, q. \tag{3.58}$$

By A1), (3.55) and noting that

$$\begin{aligned} & \textstyle\sum_{i=1}^n (I(X_{i,s} \le X_{k,s}) - \hat{F}_s(X_{k,s}))^2 \\ = \;& (n+1)\hat{F}_s(X_{k,s})(1-\hat{F}_s(X_{k,s})) - \hat{F}_s^2(X_{k,s}) \\ \le \;& (n+1) r(\hat{F}_s(X_{k,s})), \end{aligned}$$

we have

$$\begin{aligned} & |n^{-5/2} \textstyle\sum_{i=1}^n \sum_{k=1,k\neq i}^n l_j^{(st)}(\mathbf{Y}_{k,i};\theta_0)(I(X_{i,s} \le X_{k,s}) - \hat{F}_s(X_{k,s})) \\ & \qquad \times (I(X_{i,t} \le X_{k,t}) - \hat{F}_t(X_{k,t}))| \\ = \;& O_p(n^{-5/2} \textstyle\sum_{i=1}^n \sum_{k=1}^n r(F_s(X_{k,s}))^{-1} r(F_t(X_{k,t}))^{-1} \prod_{m=1}^d r(F_m(X_{k,m}))^{-\alpha_1} \\ & \times \{(I(X_{i,s} \le X_{k,s}) - \hat{F}_s(X_{k,s}))^2 + (I(X_{i,t} \le X_{k,t}) - \hat{F}_t(X_{k,t}))^2\}) \\ = \;& O_p(n^{-5/2} \textstyle\sum_{k=1}^n r(F_s(X_{k,s}))^{-1} r(F_t(X_{k,t}))^{-1} \prod_{m=1}^d r(F_m(X_{k,m}))^{-\alpha_1} \\ & \times (n+1)\{r(\hat{F}_s(X_{k,s})) + r(-\hat{F}_t(X_{k,t}))\}) \\ = \;& O_p(n^{-3/2} \textstyle\sum_{k=1}^n r(F_t(X_{k,t}))^{-1} \prod_{m=1}^d r(F_m(X_{k,m}))^{-\alpha_1}) \\ & + O_p(n^{-3/2} \textstyle\sum_{k=1}^n r(F_s(X_{k,s}))^{-1} \prod_{m=1}^d r(F_m(X_{k,m}))^{-\alpha_1}) \end{aligned} \tag{3.59}$$

for $s, t = 1, \cdots, q$. Like the proof of (3.57), we have

$$n^{-3/2} \sum_{k=1}^n r(F_t(X_{k,t}))^{-1} \prod_{m=1}^d r(F_m(X_{k,m}))^{-\alpha_1} = o_p(1)$$

for $t = 1, \cdots, d$, i.e.,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n I_3(i,j) = o_p(1) \quad \text{for} \quad j = 1, \cdots, q. \tag{3.60}$$

Write

$$\begin{aligned} I_1(i,j) \;& = l_j(\mathbf{T}_i;\theta_0) + \textstyle\sum_{s=1}^d l_j^{(s)}(\mathbf{T}_i;\theta_0)\{\hat{F}_s(X_{i,s}) - F_s(X_{i,s})\} \\ & \quad + \tfrac{1}{2} \textstyle\sum_{s=1}^d \sum_{t=1}^d l_j^{(st)}(\mathbf{Y}_i^*;\theta_0)\{\hat{F}_s(X_{i,s}) - F_s(X_{i,s})\}\{\hat{F}_t(X_{i,t}) - F_t(X_{i,t})\} \\ & =: II_1(i,j) + II_2(i,j) + II_3(i,j), \end{aligned}$$

where

$$\mathbf{Y}_i^* = \beta_i^* \hat{\mathbf{T}}_i + (1-\beta_i^*)\mathbf{T}_i$$

and $\beta_i^* \in [0,1]$.

Since

$$\max_{1\le i\le n}|\frac{\sqrt{n}\{\hat{F}_s(X_{i,s})-F_s(X_{i,s})\}}{F_s^{1/2}(X_{i,s})(1-F_s(X_{i,s}))^{1/2}}|=O_p(\log n) \tag{3.61}$$

for $s=1,\cdots,d$ (see Mason [64]), using the same arguments in proving (3.57), we can show that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}II_3(i,j)=o_p(1) \quad \text{for} \quad j=1,\cdots,q. \tag{3.62}$$

It is easy to check that

$$\begin{aligned}&\mathbb{E}(\{\hat{F}_{s,-i}(X_{i,s})-\tfrac{n-1}{n}F_s(X_{i,s})\}\{\hat{F}_{s,-k}(X_{k,s})-\tfrac{n-1}{n}F_s(X_{k,s})\}|X_i,X_k)\\ =\;&\tfrac{n-2}{n^2}\{F_s(X_{i,s}\wedge X_{k,s})-F_s(X_{i,s})F_s(X_{k,s})\}\end{aligned} \tag{3.63}$$

for $i\neq k$. Put

$$W_1(i,j,s)=l_j^{(s)}(\mathbf{T}_i;\theta_0)\{\hat{F}_s(X_{i,s})-F_s(X_{i,s})\},$$

$$W_2(i,j,s)=l_j^{(s)}(\mathbf{T}_i;\theta_0)\{\hat{F}_{s,-i}(X_{i,s})-\frac{n-1}{n}F_s(X_{i,s})\}$$

and

$$\begin{aligned}W_3(i,j,s)\;=\;&\int_0^1\cdots\int_0^1 l_j^{(s)}(u_1,\cdots,u_d;\theta_0)\{I(F_s(X_{i,s})\le u_s)-u_s\}\times\\ &c(u_1,\cdots,u_d;\theta_0)\,du_1\cdots du_d.\end{aligned}$$

Since

$$W_1(i,j,s)=\frac{n}{n+1}W_2(i,j,s)+l_j^{(s)}(\mathbf{T}_i;\theta_0)\{\frac{1}{n+1}-\frac{2}{n+1}F_s(X_{i,s})\},$$

it follows from the same arguments in proving (3.57) that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_1(i,j,s)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}W_2(i,j,s)+o_p(1) \tag{3.64}$$

for $j=1,\cdots,q$ and $s=1,\cdots,d$. By (3.61), we have

$$\begin{aligned}&\max_{1\le i\le n}|\tfrac{\sqrt{n}\{\hat{F}_{s,-i}(X_{i,s})-\frac{n-1}{n}F_s(X_{i,s})\}}{F_s^{1/2}(X_{i,s})(1-F_s(X_{i,s}))^{1/2}}|\\ \le\;&\max_{1\le i\le n}\tfrac{n+1}{n}|\tfrac{\sqrt{n}\{\hat{F}_s(X_{i,s})-F_s(X_{i,s})\}}{F_s^{1/2}(X_{i,s})(1-F_s(X_{i,s}))^{1/2}}|\\ &+\max_{1\le i\le n}\{\sqrt{n}F_s^{1/2}(X_{i,s})(1-F_s(X_{i,s}))^{1/2}\}^{-1}\\ =\;&O_p(\log n)\end{aligned} \tag{3.65}$$

for $s = 1, \cdots, d$. Using (3.65) and the same arguments in proving (3.57), we have

$$\frac{1}{n}\sum_{i=1}^{n} W_2^2(i,j,s) = o_p(1) \quad \text{for} \quad j = 1, \cdots, q, s = 1, \cdots, d. \tag{3.66}$$

By (3.63) and (3.66), we have

$$\begin{aligned}
& \mathbb{E}\{\tfrac{1}{n}\textstyle\sum_{i,k=1,i\neq k}^{n} W_2(i,j,s)W_2(k,j,s)\} \\
= \; & \mathbb{E}(\mathbb{E}\{\tfrac{1}{n}\textstyle\sum_{i,k=1,i\neq k}^{n} W_2(i,j,s)W_2(k,j,s)|X_i, X_k\}) \\
= \; & \mathbb{E}\{\tfrac{n-2}{n^3}\textstyle\sum_{i,k=1,i\neq k}^{n} l_j^{(s)}(\mathbf{T}_i;\theta_0)l_j^{(s)}(\mathbf{T}_k;\theta_0)(F_s(X_{i,s}) \wedge F_s(X_{k,s}) - F_s(X_{i,s})F_s(X_{k,s}))\} \\
= \; & \textstyle\int_0^1 \cdots \int_0^1 l_j^{(s)}(u_1, \cdots, u_d;\theta_0)l_j^{(s)}(v_1, \cdots, v_d;\theta_0)(u_s \wedge v_s - u_s v_s)\times \\
& c(u_1, \cdots, u_d;\theta_0)c(v_1, \cdots, v_d;\theta)\, du_1 \cdots du_d dv_1 \cdots dv_d + o(1),
\end{aligned} \tag{3.67}$$

$$\begin{aligned}
& \mathbb{E}\{\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\sum_{k=1}^{n} W_2(i,j,s)W_3(k,j,s)\} \\
= \; & \mathbb{E}(\mathbb{E}\{\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\sum_{k=1}^{n} W_2(i,j,s)W_3(k,j,s)|X_i, X_k\}) \\
= \; & \mathbb{E}(\mathbb{E}\{\tfrac{1}{n}\textstyle\sum_{i,k=1,i\neq k}^{n} W_2(i,j,s)W_3(k,j,s)|X_i, X_k\}) \\
= \; & \mathbb{E}\{\tfrac{1}{n^2}\textstyle\sum_{i,k=1,i\neq k}^{n} l_j^{(s)}(\mathbf{T}_i;\theta_0)(I(X_{k,s} \leq X_{i,s}) - F_s(X_{i,s}))W_3(k,j,s)\} \\
= \; & \textstyle\int_0^1 \cdots \int_0^1 l_j^{(s)}(u_1, \cdots, u_d;\theta_0)l_j^{(s)}(v_1, \cdots, v_d;\theta_0)(u_s \wedge v_s - u_s v_s)\times \\
& c(u_1, \cdots, u_d;\theta_0)c(v_1, \cdots, v_d;\theta_0)\, du_1 \cdots du_d dv_1 \cdots dv_d + o(1)
\end{aligned} \tag{3.68}$$

and

$$\begin{aligned}
& \mathbb{E}\{\tfrac{1}{n}\textstyle\sum_{i=1}^{n}\sum_{k=1}^{n} W_3(i,j,s)W_3(k,j,s)\} \\
= \; & \mathbb{E}\{\tfrac{1}{n}\textstyle\sum_{i=1}^{n} W_3^2(i,j,s)\} \\
= \; & \textstyle\int_0^1 \cdots \int_0^1 l_j^{(s)}(u_1, \cdots, u_d;\theta_0)l_j^{(s)}(v_1, \cdots, v_d;\theta_0)(u_s \wedge v_s - u_s v_s)\times \\
& c(u_1, \cdots, u_d;\theta_0)c(v_1, \cdots, v_d;\theta_0)\, du_1 \cdots du_d dv_1 \cdots dv_d
\end{aligned} \tag{3.69}$$

for $j = 1, \cdots, q$ and $s = 1, \cdots, d$. Hence, it follows from (3.66)–(3.69) that for any

$\epsilon > 0$

$$\begin{aligned}
& \mathbb{P}(|\tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n (W_2(i,j,s) - W_3(i,j,s))| > \epsilon) \\
= \; & \mathbb{P}(\tfrac{1}{n} \textstyle\sum_{i=1}^n W_2^2(i,j,s) + \tfrac{1}{n} \sum_{i,k=1, i\neq k}^n W_2(i,j,s) W_2(k,j,s) \\
& - \tfrac{2}{n} \textstyle\sum_{i=1}^n \sum_{k=1}^n W_2(i,j,s) W_3(k,j,s) + \tfrac{1}{n} \sum_{i=1}^n \sum_{k=1}^n W_3(i,j,s) W_3(k,j,s) > \epsilon^2) \\
\leq \; & \mathbb{P}(\tfrac{1}{n} \textstyle\sum_{i=1}^n W_2^2(i,j,s) > \epsilon^2/2) \\
& + \tfrac{2}{\epsilon^2} \mathbb{E}\{\tfrac{1}{n} \textstyle\sum_{i,k=1, i\neq k}^n W_2(i,j,s) W_2(k,j,s) - \tfrac{2}{n} \sum_{i=1}^n \sum_{k=1}^n W_2(i,j,s) W_3(k,j,s) \\
& + \tfrac{1}{n} \textstyle\sum_{i=1}^n \sum_{k=1}^n W_3(i,j,s) W_3(k,j,s)\} \\
= \; & o(1).
\end{aligned} \tag{3.70}$$

By (3.62), (3.64) and (3.70), we have

$$\begin{aligned}
& \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n I_1(i,j) \\
= \; & \tfrac{1}{\sqrt{n}} \textstyle\sum_{i=1}^n l_j(\mathbf{T}_i; \theta_0) + \tfrac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{s=1}^d \int_0^1 \cdots \int_0^1 l_j^{(s)}(u_1, \cdots, u_d; \theta_0) \\
& \times (I(F_s(X_{i,s}) \leq u_s) - u_s) c(u_1, \cdots, u_d; \theta_0)\, du_1 \cdots du_d + o_p(1)
\end{aligned} \tag{3.71}$$

for $j = 1, \cdots, q$. Note that by A1) and A2),

$$\begin{aligned}
& \mathbb{E}\left[\int_0^1 \cdots \int_0^1 l_j^{(s)}(u_1, \cdots, u_d; \theta_0)(I(F_s(X_{i,s}) \leq u_s) - u_s) c(u_1, \cdots, u_d; \theta_0)\, du_1 \cdots du_d\right]^2 \\
= \; & \int_0^1 \cdots \int_0^1 l_j^{(s)}(u_1, \cdots, u_d; \theta_0) c(u_1, \cdots, u_d; \theta_0) l_j^{(s)}(v_1, \cdots, v_d; \theta_0) c(v_1, \cdots, v_d; \theta_0) \\
& \times (\min\{u_s, v_s\} - u_s v_s)\, du_1 \cdots du_d dv_1 \cdots dv_d \\
< \; & \infty.
\end{aligned}$$

Hence, the lemma follows from (3.58), (3.60), (3.71) and the central limit theorem.

Lemma 3.4.3. *Under conditions of Theorem 3.4.1, we have*

$$\frac{1}{n} \sum_{i=1}^n \mathbf{Z}_i(\theta_0) \mathbf{Z}_i^T(\theta_0) \xrightarrow{p} \Sigma \quad \text{as} \quad n \to \infty,$$

where Σ is defined in Lemma 3.4.2.

Proof. Using the same notation in the proof of Lemma 3.4.2, we can show that for fixed $j, m = 1, \cdots, q$,

$$\frac{1}{n} \sum_{i=1}^n I_1(i,j) I_1(i,m) = \mathbb{E}[l_j(\mathbf{T}_1; \theta_0) l_m(\mathbf{T}_1; \theta_0)] + o_p(1),$$

$$\frac{1}{n}\sum_{i=1}^{n} I_3(i,j)\{I_1(i,m)+I_2(i,m)\} = o_p(1), \quad \frac{1}{n}\sum_{i=1}^{n} I_3(i,j)I_3(i,m) = o_p(1),$$

$$\frac{1}{n}\sum_{i=1}^{n} I_1(i,j)I_2(i,m) = \mathbb{E}\left[l_j(\mathbf{T}_1;\theta_0)\sum_{s=1}^{d} W(m,s)\right] + o_p(1),$$

and

$$\frac{1}{n}\sum_{i=1}^{n} I_2(i,j)I_2(i,m) = \mathbb{E}\left[\sum_{s=1}^{d}\sum_{t=1}^{d} W(j,s)W(m,t)\right] + o_p(1),$$

which implies that

$$\frac{1}{n}\sum_{i=1}^{n} Z_{i,j}(\theta_0)Z_{i,m}(\theta_0) \xrightarrow{p} \sigma_{jm} \quad \text{for} \quad j,m = 1,\cdots,q,$$

i.e., the lemma holds.

Lemma 3.4.4. *Under conditions of Theorem 3.4.1, we have for $j = 1,\cdots,q$,*

$$\max_{1\le i\le n} |Z_{i,j}(\theta_0)| = o_p(n^{1/2}).$$

Proof. We shall use the same notation in the proof of Lemma 3.4.2. For any $M > 0$, we have

$$\begin{aligned} & \mathbb{P}\left(\max_{1\le i\le n} |I_2(i,j)| \ge n^{1/2}M\right) \\ \le\ & \mathbb{P}\left(\max_{1\le i\le n} \frac{1}{n}\sum_{k=1,k\ne i}^{n}\sum_{s=1}^{d} |l_j^{(s)}(\hat{\mathbf{T}}_k;\theta_0)| \ge n^{1/2}M\right) \\ \le\ & \mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n}\sum_{s=1}^{d} |l_j^{(s)}(\hat{\mathbf{T}}_k;\theta_0)| \ge n^{1/2}M\right). \end{aligned}$$

Hence by the same arguments in (3.56) and (3.57) we have

$$n^{-3/2}\sum_{k=1}^{n}\sum_{s=1}^{d} |l_j^{(s)}(\hat{\mathbf{T}}_k;\theta_0)| = o_p(1),$$

i.e., $\mathbb{P}\left(\max_{1\le i\le n} |I_2(i,j)| \ge n^{1/2}M\right) = o(1)$, which implies that

$$\max_{1\le i\le n} |I_2(i,j)| = o_p(n^{1/2}). \tag{3.72}$$

Note that in (3.59) and (3.60), we actually showed

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} |I_3(i,j)| = o_p(1),$$

which implies

$$\max_{1\le i\le n} |I_3(i,j)| = o_p(n^{1/2}). \tag{3.73}$$

Similarly, we have

$$\max_{1\le i\le n} |II_2(i,j)| = o_p(n^{1/2}) \quad \text{and} \quad \max_{1\le i\le n} |II_3(i,j)| = o_p(n^{1/2}). \tag{3.74}$$

Since $\mathbb{E}[l_j^2(\mathbf{T}_1;\theta_0)] < \infty$, we have $n\mathbb{P}(l_j^2(\mathbf{T}_1;\theta_0) \ge n) = o(1)$, i.e.,

$$\max_{1\le i\le n} |II_1(i,j)| = o_p(n^{1/2}). \tag{3.75}$$

Hence the lemma follows from (3.72) to (3.75).

Proof of Theorem 3.4.1. It follows from Lemmas 3.4.2-Lemma 3.4.4 and the standard arguments in the empirical likelihood method for a mean vector.

CHAPTER IV

COMPLETELY MIXABLE DISTRIBUTIONS AND THEIR APPLICATIONS IN RISK MANAGEMENT

In this chapter, we introduce the theory of completely mixable distributions. We give the definition and study the properties of CM distributions. We prove a few classes of distributions are CM. The idea of CM distributions can be used to provide valuable implications in variance minimization, multivariate dependence and risk management. The content of this chapter is mainly based on the following papers and preprints.

1. Wang, R., Peng, L. and Yang, J. (2012). Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities. *Preprint*.

2. Puccetti, G., Wang, B. and Wang, R. (2012). Advances in complete mixability. *Journal of Applied Probability*, to appear.

3. Wang, B and Wang, R. (2011). The complete mixability and convex minimization problems for monotone marginal distributions. *Journal of Multivariate Analysis* **102**, 1344-1360.

4.1 Introduction

Rüschendorf and Uckelmann [87] investigated *random variables with constant sums* and associated it with variance minimization problems in Fréchet class. In this chapter, a distribution function F is called *n-completely mixable* (n-CM) if there exist n random variables $X_1, \ldots, X_n$ identically distributed as F having constant sum, that is satisfying

$$X_1 + \cdots + X_n = n\mu.$$

This property was studied by Gaffke and Rüschendorf [41] in the case of uniform distributions. The case of distributions with symmetric and unimodal density was studied for $n = 3$ by Knott and Smith [60], [61] and for the general case $n \geq 2$ by Rüschendorf and Uckelmann [87] using a different method.

The concept of complete mixability is related to some Fréchet class optimization problems in the theory of optimal couplings. Let $\mathbf{X} = (X_1, \cdots, X_n)$, $S = X_1 + \cdots + X_n$ and define the homogenous *Fréchet class* as in Chapter I

$$\mathfrak{F}_n(F_1, \cdots, F_n) = \{\mathbf{X} : \ X_i \sim F_i, \ i = 1, \cdots, n\}.$$

Question A: the expectation of a convex function. Find

$$\inf_{\mathbf{X} \in \mathfrak{F}_n(F_1, \cdots, F_n)} \mathbb{E}[f(S)] \tag{4.1}$$

for f being a convex function.

Question B: the distribution of the total risk. Find bounds on the distribution of S:

$$m_+(s) = \inf_{\mathbf{X} \in \mathfrak{F}_n(F_1, \cdots, F_n)} \mathbb{P}(S < s); \tag{4.2}$$

$$M_+(s) = \sup_{\mathbf{X} \in \mathfrak{F}_n(F_1, \cdots, F_n)} \mathbb{P}(S < s). \tag{4.3}$$

As introduced in Chapter I, Questions A and B have relevant applications in quantitative risk management, where they are needed to assess the aggregate risk of a portfolio of losses for regulatory issues. Later we will review the literature on these problems in Section 4.5 and Section 4.6. For more details on the motivation of these problems within quantitative risk management, we refer to Embrechts and Puccetti [37].

The rest of this chapter is organized as follows. We give the formal definition and basic properties of CM distributions in Section 4.2. Completeness and Decomposition theorems are given in Section 4.3. We prove three classes of distributions are CM in Section 4.4. Section 4.5 addresses Question A and Section 4.6 addresses Question B, both using the idea of CM distributions. Some technical proofs are put in Section 4.7. Throughout this chapter, we identify probability measures with the corresponding distribution functions.

4.2 Definition and Basic Properties

In this thesis, we call the marginal distribution of random variables with a constant sum a *completely mixable distribution*, as in the following definition.

Definition 4.2.1. A distribution function F on $\mathbb{R}$ is called *n-completely mixable* (n-CM) if there exist n random variables $X_1, \ldots, X_n$ identically distributed as F such that

$$P(X_1 + \cdots + X_n = n\mu) = 1, \tag{4.4}$$

for some $\mu \in \mathbb{R}$. Any such μ is called a center of F and any vector $(X_1, \ldots, X_n)$ satisfying (4.4) with $X_i \sim F, 1 \leq i \leq n$, is called an *n-complete mix*.

Sometimes we say a distribution is CM omitting the integer n which should be clear from the context. We denote by $\mathcal{M}_n(\mu)$ the set of all n-CM distributions with center μ, and by $\mathcal{M}_n = \bigcup_{\mu\in\mathbb{R}} \mathcal{M}_n(\mu)$ the set of all n-CM distributions on $\mathbb{R}$.

Proposition 4.2.1. ***(Basic properties.)*** *For simplicity, in the following we let F_X be the distribution of X for any random variable X.*

(1) (Invariance under affine transformations) Suppose $F_X \in \mathcal{M}_n(\mu)$, then $F_{aX+b} \in \mathcal{M}_n(a\mu + b)$ for any constants a, b.

(2) (Center of the complete mixability) Suppose $F_X \in \mathcal{M}_n(\mu)$ and follows the weak law of large numbers (WLLN), then μ is unique. If $\mathbb{E}(X)$ exists, then $\mu = \mathbb{E}(X)$.

(3) (Additivity 1: distribution-wise) Suppose $F, G \in \mathcal{M}_n(\mu)$. Then for any $\lambda \in [0, 1]$, $\lambda F + (1 - \lambda)G \in \mathcal{M}_n(\mu)$.

(4) *(Additivity 2: index-wise) Suppose* $F \in \mathcal{M}_n, G \in \mathcal{M}_k$, *then* $\frac{n}{n+k}F + \frac{k}{n+k}G \in \mathcal{M}_{n+k}$. *As a consequence, if* $F \in \mathcal{M}_n \cap \mathcal{M}_k$, *then* $F \in \mathcal{M}_{n+k}$.

(5) *(Additivity 3: random-variable-wise) Suppose* X *and* Y *are independent,* $F_X, F_Y \in \mathcal{M}_n$, *then* $F_{X+Y} \in \mathcal{M}_n$.

(6) *(Mean condition) Suppose the distribution* $F_X \in \mathcal{M}_n(\mu)$. *Let* $a = \sup\{x : \mathbb{P}(X \leq x) = 0\}$ *and* $b = \sup\{x : \mathbb{P}(X \leq x) < 1\}$. *If one of* a *and* b *is finite, then the other one is finite,* $\mu = \mathbb{E}(X)$ *and*

$$a + \frac{b-a}{n} \leq \mu \leq b - \frac{b-a}{n}. \tag{4.5}$$

Proof.

(1) This follows immediately from the definition.

(2) Assume $\mathbb{E}(X)$ exists and $(X_1, \cdots, X_n)$ is an n-complete mix with marginal distribution F_X. Taking expectation on both sides of $\mu = \frac{1}{n}(X_1 + \cdots + X_n)$ gives us $\mu = \mathbb{E}(X)$. Now suppose F_X follows WLLN. We can take independent copies of $(X_1, \cdots, X_n)$, denoted by $\{(X_{1,i}, \cdots, X_{n,i})\}_{i=1}^{\infty}$, and take their average

$$\begin{aligned} n\mu &= \frac{1}{k}\sum_{i=1}^{k}(X_{1,k} + \cdots + X_{n,k}) \\ &= \frac{1}{k}\sum_{i=1}^{k} X_{1,k} + \cdots + \frac{1}{k}\sum_{i=1}^{k} X_{n,k} \\ &= n\mathbb{E}(XI_{\{|X_1| \leq k\}}) + o_p(1) \end{aligned}$$

as k goes to infinity. Therefore $\mathbb{E}(XI_{\{|X_1| \leq k\}}) \to \mu$ and μ is unique.

(3) Suppose $X_1 + \cdots + X_n = n\mu$, $X_i \sim F$ and $Y_1 + \cdots + Y_n = n\mu$, $Y_i \sim G$, $i = 1, \cdots, n$. Let Z be a Bernoulli(λ) random variable independent of $\{X_i\}_{i=1}^n$ and $\{Y_i\}_{i=1}^n$. Set $Z_i = I_{\{Z=1\}}X_i + I_{\{Z=0\}}Y_i$, then $Z_1 + \cdots + Z_n = n\mu$ and $Z_i \sim \lambda F + (1-\lambda)G$, $i = 1, \cdots, n$.

(4) Suppose $X_1 + \cdots + X_n = n\mu$, $X_i \sim F$, $i = 1, \cdots, n$ and $Y_1 + \cdots + Y_k = k\nu$, $Y_j \sim G$, $j = 1, \cdots, k$. Let σ be a random permutation uniformly distributed on the set of all $(n+k)$-permutations and independent of $X_1, \cdots, X_n, Y_1, \cdots, Y_k$. Denote

$$(Z_1, \cdots, Z_{n+k}) = \sigma(X_1, \cdots, X_n, Y_1, \cdots, Y_k),$$

then $Z_1 + \cdots + Z_{n+k} = n\mu + k\nu$ and $Z_i \sim \frac{n}{n+k}F + \frac{k}{n+k}G$, $i = 1, \cdots, n+k$.

(5) Let $X_i \sim F_X$, $Y_i \sim F_Y$, $i = 1, \cdots, n$ such that $X_1 + \cdots + X_n$ and $Y_1 + \cdots + Y_n$ are constants. Denote $\mathbf{X} = (X_1, \cdots, X_n)$, $\mathbf{Y} = (Y_1, \cdots, Y_n)$ and let $F_{\mathbf{X}}$ and $F_{\mathbf{Y}}$ be the distributions of $\mathbf{X}$ and $\mathbf{Y}$. Let $\hat{\mathbf{X}} = (\hat{X}_1, \cdots, \hat{X}_n) \sim P_{\mathbf{X}}$ and $\hat{\mathbf{Y}} = (\hat{Y}_1, \cdots, \hat{Y}_n) \sim P_{\mathbf{Y}}$ be independent random vectors. Then we have $\hat{X}_1 + \cdots + \hat{X}_n$ and $\hat{Y}_1 + \cdots + \hat{Y}_n$ are both constants. Denoting $\hat{F}$ by the distribution of $\hat{\mathbf{X}} + \hat{\mathbf{Y}}$, the 1-marginal distribution of $\hat{F}$ is identical with F_{X+Y}. Now $X_i + Y_i \sim F_{X+Y}$, $i = 1, \cdots, n$ and $\sum_{i=1}^n (X_i + Y_i)$ is a constant. Hence $F_{X+Y} \in \mathcal{M}_n$.

(6) Let $X_i \sim F_X$, $i = 1, \cdots, n$, $X_1 + \cdots + X_n = n\mu$ and suppose $a > -\infty$. Note that if $\mu < a + \frac{b-a}{n}$, then $X_1 = n\mu - (X_2 + \cdots + X_n) \leq n\mu - (n-1)a < b$, which contradicts the fact that $b = \sup\{x : \mathbb{P}(X \leq x) < 1\}$. Thus $\mu \geq a + \frac{b-a}{n}$ and $b < \infty$. The inequality $\mu \leq b - \frac{b-a}{n}$ and the case given $b < \infty$ can be obtained similarly. $\square$

The mean condition (4.5) is very important in the theory of CM distributions as a necessary condition. Later we will see that the condition (4.5) is sufficient for some classes of distributions. Also note that the uniqueness of the center of a CM distribution is still unknown.

Proposition 4.2.2. ***(Examples.)***

(i) F is 1-CM if and only if F is the distribution of a constant.

(ii) *F is 2-CM if and only if F is symmetric, i.e. $X \sim F$ and $a - X \sim F$ for some constant $a \in \mathbb{R}$.*

(iii) *The Binomial distribution $B(n, p/q), p, q \in \mathbb{N}$, is q-CM.*

(iv) *The Gaussian and the Cauchy distributions are n-CM for $n \geq 2$.*

(v) *The uniform distribution on the interval $[a, b]$ is n-CM for any $n \geq 2$ and $a < b$.*

(vi) *The Beta distribution with parameters $\alpha, \beta > 0$ where $(\alpha - 1)(\beta - 1) \leq 0$ is n-CM for $\frac{1}{n} \leq \frac{\alpha}{\alpha+\beta} \leq \frac{n-1}{n}$.*

(vii) *The Beta distribution Beta(α, β) with $1 \leq \alpha, \beta \leq 2$ is n-CM for $n \geq 3$.*

(viii) *Any triangular distribution is n-CM for $n \geq 3$.*

Some of the examples come from theorems later in Section 4.4. We put the proof in Section 4.7.

Before closing this section, we generalize the idea of CM distributions to the non-homogenous case.

Definition 4.2.2. The univariate distribution functions $F_1, \cdots, F_n$ are *jointly mixable* (JM) if there exist n random variables $X_1, \cdots, X_n$ with distribution functions $F_1, \cdots, F_n$ respectively, such that

$$P(X_1 + \cdots + X_n = n\mu) = 1, \tag{4.6}$$

holds for some $C \in \mathbb{R}$.

Obviously, $F_1, \cdots, F_n$ are JM distributions when $F_1 = \cdots = F_n = F$ and F is n-CM. The following proposition gives a necessary condition for JM distributions and the condition is sufficient for n normal distributions. The proof is given in Section 4.7.

Proposition 4.2.3.

1. *Suppose $F_1, \cdots, F_n$ are JM with finite variance $\sigma_1^2, \cdots, \sigma_n^2$. Then*

$$\max_{1 \le i \le n} \sigma_i \le \frac{1}{2} \sum_{i=1}^{n} \sigma_i. \tag{4.7}$$

2. *Suppose F_i is $N(\mu_i, \sigma_i^2)$ for $i = 1, \cdots, n$. Then $F_1, \cdots, F_n$ are JM if and only if (4.7) holds.*

4.3 Completeness and Decomposition Theorems

In this section, we show that any n-CM distribution can be obtained as the limit of a convex combination of discrete n-CM distributions. First, we show that the sets $\mathcal{M}_n(\mu)$ and $\mathcal{M}_n$ are complete under weak convergence, that is any n-CM distributions can be seen as the the limit of n-CM discrete distributions.

Theorem 4.3.1. *The following statements hold for weak convergence.*

(a) The limit of a sequence of n-CM distribution functions (with center μ) is n-CM (with center μ).

(b) Any n-CM distribution function with center μ is the limit of a sequence of discrete n-CM distribution function with center μ.

(c) A distribution function is n-CM (with center μ) if and only if it is the limit of a sequence of discrete n-CM distribution functions (with center μ).

Proof.

(a) Denote by $F^k, k \in \mathbb{N}$ a sequence of n-CM distributions having limit F. Since $F^k \in \mathcal{M}_n$, for any $k \in \mathbb{N}$ it is possible to find $X_1^k, \ldots, X_n^k$ such that $X_i^k \sim F^k, 1 \le i \le n$ and

$$P(X_1^k + \cdots + X_n^k = c_k) = 1, \tag{4.8}$$

for some $c_k \in \mathbb{R}$. As $F^k \xrightarrow{d} F$, there also exist n random variables $X_1, \ldots, X_n$ identically distributed as F for which $X_i^k \xrightarrow{d} X_i, 1 \le i \le n$ and, therefore, such that

$$(X_1^k + \cdots + X_n^k) \xrightarrow{d} (X_1 + \cdots + X_n). \tag{4.9}$$

Combining (4.8) and (4.9), we find that $X_1 + \cdots + X_n = c = \lim c_k$ holds a.s.. Since $X_i \sim F, 1 \le i \le n$, this implies that F is n-CM. If we have $c_k = n\mu$ for all $k \in \mathbb{N}$, then $c = n\mu$.

(b) Let $\mathbf{X} = (X_1, \dots, X_n)$ be an n-complete mix on $\mathbb{R}^n$ with $X_i \sim F$, $1 \le i \le n$ and

$$X_1 + \cdots + X_n = n\mu, \text{ a.s..}$$

As $\mathbf{X}$ is supported on the set $S_n(\mu) = \{\mathbf{x} \in \mathbb{R}^n : \sum_{i=1}^n x_i = n\mu\} \subset \mathbb{R}^n$, we can find a sequence $F^k, k \in \mathbb{N}$ of discrete distributions on $S_n(\mu)$ converging weakly to the distribution of $\mathbf{X}$. The theorem follows by noting that F_1^k, the first marginal of F^k, is n-CM since F^k is supported on $S_n(\mu)$ and the sequence $F_1^k, k \in \mathbb{N}$ converges weakly to F.

(c) This is a corollary of points (a) and (b).

□

Now, we prove a decomposition theorem for n-CM distributions. In the following, we call an *n-discrete uniform* distribution a uniform distribution on n points, that is giving mass $1/n$ at each of the n points in its support.

Lemma 4.3.2. *An n-discrete uniform distribution is n-CM.*

Proof. Let F be an n-discrete uniform distribution on the points $y_1, \dots, y_n$. Let $\mathbf{X} = (X_1, \dots, X_n)$ be a random vector uniformly distributed on the $n!$ vectors

$$(y_{\pi(1)}, \dots, y_{\pi(n)}), \ \pi \in \mathcal{P}_n,$$

where $\mathcal{P}_n$ is the set of all permutations of $\{1, \dots, n\}$. In the support of $\mathbf{X}$, there are exactly $(n-1)!$ vectors having the value y_j as i-th component. Therefore, we have

$$P(X_i = y_j) = \frac{(n-1)!}{n!} = 1/n, \ 1 \le i, j \le n.$$

As a consequence, $\mathbf{X}$ has marginal distributions identically distributed as F. Since $\sum_{i=1}^{n} y_{\pi(i)}$ is constant on π, $\mathbf{X}$ is an n-complete mix and F is n-CM. □

We denote by $\mathcal{M}_n^S(\mu)$ the set of all n-discrete uniform distributions with mean μ and by $L\left(\mathcal{M}_n^S(\mu)\right)$ be the set of all countable convex combinations of elements in $\mathcal{M}_n^S(\mu)$, that is

$$L\left(\mathcal{M}_n^S(\mu)\right) = \left\{\sum_{k=1}^{\infty} a_k F^k; F^k \in \mathcal{M}_n^S(\mu),\ a_k \geq 0, \sum_{k=1}^{\infty} a_k = 1\right\}.$$

We show that any discrete n-CM distribution can be obtained as the countable convex combination of n-discrete uniform distributions.

Theorem 4.3.3. *The following statements hold:*

(a) The countable convex combination of n-CM distribution functions with center μ is n-CM with center μ.

(b) If F is discrete, then $F \in \mathcal{M}_n(\mu)$ if and only if $F \in L\left(\mathcal{M}_n^S(\mu)\right)$.

(c) If $F \in L\left(\mathcal{M}_n^S(\mu)\right)$ with $F = \sum_{k\in\mathbb{N}} a^k F^k$, the joint distribution G of an n-complete mix with marginals F is given by

$$G(x_1, \cdots, x_n) = \sum_{k\in\mathbb{N}} \frac{a_k}{n!} \prod_{i=1}^{n} [nF^k(x_{[i]}) - i + 1]^+,$$

where $x_{[i]}$ is the i-th order statistics of $\{x_1, \cdots, x_n\}$.

Proof.

(a) The statement for finite convex combinations follows by induction from Proposition 4.2.1(3). Now let $a_k, k \in \mathbb{N}$ be a sequence of nonnegative values with $\sum_{k=1}^{+\infty} a_k = 1$ and $F^k \in \mathcal{M}_n(\mu), k \in \mathbb{N}$ be a sequence of n-CM distributions having center μ. W.l.o.g., we can assume $a_1 > 0$ and define the new sequence

$$G^k = \frac{\sum_{i=1}^{k} a_i F^i}{\sum_{i=1}^{k} a_i}, k \in \mathbb{N}.$$

Any G^k is the finite convex sum of n-CM distributions, thus it is n-CM. Since $G^k \xrightarrow{d} G = \sum_{k=1}^{+\infty} a_k F^k$, we have that G is n-CM by point (a) in Theorem 4.3.1.

(b) The inclusion $L\left(\mathcal{M}_n^S(\mu)\right) \subset \mathcal{M}_n(\mu)$, follows from (a). Then, it is sufficient to show $\mathcal{M}_n(\mu) \subset L\left(\mathcal{M}_n^S(\mu)\right)$. Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a complete mix with center μ and discrete marginals identically distributed as F. Denoting by $\{\mathbf{x}^j, j \in A \subset \mathbb{N}\}$ the countable support of $\mathbf{X}$, we have

$$F(s) = \frac{1}{n}\sum_{i=1}^{n} P(X_i \leq s) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j\in A} P\left(X_i \leq s|\mathbf{X} = \mathbf{x}^j\right) P(\mathbf{X} = \mathbf{x}^j)$$
$$= \sum_{j\in A} P(\mathbf{X} = \mathbf{x}^j)\left(\frac{1}{n}\sum_{i=1}^{n} P\left(X_i \leq s|\mathbf{X} = \mathbf{x}^j\right)\right) = \sum_{j\in A} a_j \left(\frac{1}{n}\sum_{i=1}^{n} 1_{\{x_i^j \leq s\}}\right),$$

where x_i^j denotes the i-th component of the vector $\mathbf{x}^j$ and $a_j = P(\mathbf{X} = \mathbf{x}_j), j \in A$. Note that the a_j's are nonnegative, $\sum_{j\in A} a_j = 1$ and, for any $j \in A$, the function $\sum_{i=1}^n 1_{\{x_i^j \leq s\}}$ is the distribution function of a random variable uniformly distributed on $\{x_1^j, \ldots, x_n^j\}$. Being $\mathbf{X}$ an n-complete mix, we have that $\sum_{i=1}^n x_i^j = n\mu$ when $a_j > 0$. As a result, F can be written as a countable convex sum of distributions in $\mathcal{M}_n^S(\mu)$, that is $F \in L(\mathcal{M}_n^S(\mu))$.

(c) First, note that G has marginals identically distributed as F since

$$\lim_{x_i \to +\infty, i\neq j} R(x_1, \ldots, x_n) = \sum_{k\in\mathbb{N}} a^k F^k(x_j) = F(x_j),\ 1 \leq j \leq n.$$

In order to show that G is the distribution an n-complete mix, we prove that

$$G^k(x_1, \ldots, x_n) = \frac{1}{n!}\prod_{i=1}^{n}[nF^k(x_{[i]}) - i + 1]^+$$

is the distribution of an n-complete mix with center μ, for any $k \in \mathbb{N}$.

Since $F^k \in \mathcal{M}_n^S(\mu)$, there exist $y_1^k \leq \cdots \leq y_n^k$ such that $\sum_{i=1}^n y_i^k = n\mu$ and $F^k(y_i^k) = 1/n\sum_{j=1}^n 1_{\{y_j^k \leq y_i^k\}}$. Noting that

$$\frac{1}{n!}\prod_{i=1}^{n}[nF^k(x_{[i]}) - i + 1]^+ = \frac{1}{n!}\sum_{\pi\in\mathcal{P}_n} 1_{\left\{y^k_{\pi(i)} \leq x_i,\ 1\leq i\leq n\right\}},$$

we have that, for any $k \in \mathbb{N}$, G^k is uniformly distributed on the $n!$ vectors

$$(y^k_{\pi(1)}, \ldots, y^k_{\pi(n)}),\ \pi \in \mathcal{P}_n, k \in \mathbb{N}.$$

Thus, G^k is the distribution of an n-complete mix with center $1/n \sum_{i=1}^n y_i^k = \mu$, from which it follows that also $G = \sum_{k \in \mathbb{N}} a_k G^k$ is the distribution of an n-complete mix with center μ.

□

Remark 4.3.1. There are some points to remark about Theorem 4.3.3:

(i) Similarly to what done in the proof of point (b), one can show that an arbitrary n-CM distribution with center μ can be written as an integral of n-discrete uniform distributions with center μ.

(ii) Using the notation introduced in the proof of point (c), the distribution G can be seen as the distribution of the random variable $\sum_{k \in \mathbb{N}} 1_{\{Z=k\}} G^k$, where Z a discrete random variable giving mass a_k to $k \in \mathbb{N}$ and independent from the G^k's. Note, however, that the distribution of an n-complete mix for a discrete F may not be unique.

(iii) A number of the n points of the support of an n-discrete distribution can be chosen to be equal. The set of n-discrete uniform distributions therefore includes all distributions giving masses $(k/n), k \in \mathbb{N}$ to at most n different points.

(iv) The convex combination of n-discrete distributions with different centers may fail to be n-CM. For example, the Bernoulli distribution $F(s) = \left(1_{\{0 \leq s\}} + 1_{\{1 \leq s\}}\right)/2$ is the convex sum of two 1-CM distributions but it is not 1-CM. Therefore, the assumption of a common center cannot be dropped in all points of Theorem 4.3.3.

As a corollary of Theorem 4.3.1 (c) and Theorem 4.3.3 (b), we find the main result of this section.

Corollary 4.3.4. *A distribution is n-CM with center μ if and only if is the limit of a sequence of a countable convex combination of n-discrete uniform distributions with center μ.*

4.4 Classes of CM Distributions

One nice result for the complete mixability is given in Rüschendorf and Uckelmann [87]. We cite this result in a rewritten form in the following theorem.

Theorem 4.4.1. ***(Rüschendorf and Uckelmann)*** *Suppose the probability density function $p(x)$ of a distribution P is symmetric and unimodal, then P is n-CM for $n \geq 2$.*

In this section, we characterizes three more classes of CM distributions.

4.4.1 Distributions with a monotone density

In this section, we will see that the mean condition (4.5) is sufficient for a monotone density.

Theorem 4.4.2. *Suppose the probability density function $p(x)$ of a distribution F with mean μ is monotone on $[a, b]$ and $p(x) = 0$ elsewhere. Then F is n-CM if*

$$a + \frac{1}{n}(b - a) \leq \mu \leq b - \frac{1}{n}(b - a).$$

Proof of Theorem 4.4.2. For $n = 1$ or 2, the proof is trivial since no distribution satisfies the assumption when $n = 1$, and only one distribution, namely the uniform distribution, satisfies the assumption when $n = 2$. Hence we only need to prove the case of $n \geq 3$. Since the complete mixability is invariant under affine transformations, without losing generality we assume the center to be 0.

We start the proof with the discrete version of Theorem 4.4.2. We say a CM distribution A is CM on a set S, if A is supported in the set S. Let d and N be positive integers, where $d = n-1 \geq 2$, and let $S_N^d := \{-N, \cdots, -1, 0, 1, \cdots, dN\}$ be a set of $(d+1)N+1$ points. In the following proof, we identify a discrete probability distribution with its probability mass function A for simplicity.

Lemma 4.4.3. *Suppose the mass function A is supported in S_N^d, and the pair (A, N) satisfies*

(i) (decreasing mass)

$$A(-N+1) \geq \cdots \geq A(0) \geq \cdots \geq A(dN) \geq 0, \tag{4.10}$$

(ii) (boundary condition)

$$C_N(A) = A(-N) - [d \times A(dN) + (d-1) \times A(dN-1) + \cdots + 1 \times A(dN-d+1)] \geq 0, \tag{4.11}$$

(iii) (zero center of mass)

$$\sum_{i=-N}^{dN} i \times A(i) = 0. \tag{4.12}$$

Then A is $(d+1)$-CM on S_N^d.

Proof. We prove this lemma by induction over N. Our idea is to write $A = \bar{A} + \sum_{i=0}^{K} b_i B_i$ such that for each i, $b_i \geq 0$, B_i is a $(d+1)$-discrete uniform distribution centered at 0 (on S_N^d if not specified) mass function, $\bar{A}$ is supported in S_{N-1}^d, and $(\bar{A}, N-1)$ satisfies (i) and (ii). Note that (iii) is automatically satisfied. First we need the following fact.

Lemma 4.4.4. *If* (4.10), (4.12) *in Lemma 4.4.3 hold and $A(-N) \geq \frac{d+1}{2d} A(-N+1)$, then* (4.11) *holds.*

The proof of Lemma 4.4.4 will be presented in Section 4.7. This lemma implies that if $A(-N) \geq A(-N+1)$, (4.10) and (4.12) hold, then (4.11) holds. Thus, a decreasing mass function with zero center is sufficient for Lemma 4.4.3.

Now suppose Lemma 4.4.3 holds for the case of $N-1$ (here $N \geq 2$).

Case 1. $C_N(A) = 0$.

If $A(-N)$=0 then (4.11) implies that $A(dN) = A(dN-1) = \cdots = A(dN-d+1) = 0$. Thus A is supported in S^d_{N-1} and $(A, N-1)$ satisfies (i), (ii) and (iii). Therefore A is $(d+1)$-CM on S^d_{N-1} (and hence on S^d_N).

If $A(-N) > 0$, we construct B_i, $i = 0, 1, \cdots, d-1$ such that $B_i(-N) = d-i$, $B_i(-N+1) = i$, $B_i(dN-i) = 1$ and 0 otherwise. Obviously each B_i is simply mixable. Let $b_i = A(dN-i)$ and $\bar{A} = A - \sum_{i=0}^{d-1} b_i B_i$. It is straightforward to check $\bar{A}$ is still a mass function and is supported in S^d_{N-1}. Clearly $\bar{A}(i) = A(i)$ for $i = -N+2, \cdots, dN-d$, and hence (i) is satisfied by $(\bar{A}, N-1)$.

The rest work is to check (ii) $C_{N-1}(\bar{A}) \geq 0$. It is just some algebraic calculation and we leave it in Section 4.7. Thus $\bar{A}$ is $(d+1)$-CM on S^d_{N-1}. This shows $A = \bar{A} + \sum_{i=0}^{d-1} b_i B_i$ is $(d+1)$-CM (on S^d_N).

Case 2. $C_N(A) > 0$.

Denote $M = M_A = \max\{i : A(i) > 0\}$. By (i) and $A(-N) > 0$, it follows that $N \leq M \leq dN$. Let q and r be integers such that

$$(d+1)N = (N+M)q + r, \quad 0 \leq r < N+M.$$

Obviously $q < d$. For $i = 0, 1, \cdots, M+N-r$, Let $B_i(-N) = d-q$, $B_i(M) = q-1$, $B_i(r-N+i) = B_i(M-i) = 1$ and 0 elsewhere. It is easy to check each B_i is $(d+1)$-discrete uniform and centered at 0.

Let $T = T_A = \sum_{i=0}^{M+N-r} B_i$. Then T is $(d+1)$-CM, $T(-N) = (d-q)(M+N-r+1)$, $T(M) = (q-1)(M+N-r+1)+2$, $T(r-N) = T(r-N+1) = \cdots = T(M-1) = 2$

and 0 otherwise. We have

$$C_N(T) = \begin{cases} (d-q)(M+N-r+1), & M \le dN-d, \\ (d-1)((d+1)N-2r+1)-(d-r+1)(d-r), & M > dN-d. \end{cases}$$

Thus $C_N(T) > 0$. Let $b_A = \max\{x : xT(M) \le A(M), xC_N(T) \le C_N(A))\}$. For each mass function A, we define an operator $\mathcal{R}A := A - b_A T_A$. Note that $C_N(\mathcal{R}A) = C_N(A) - b_A C_N(T)$. It is straightforward to check $\mathcal{R}A$ is still a mass function, $(\mathcal{R}A, N)$ satisfies (i), (ii), (iii) and either $\mathcal{R}A(M) = 0$ or $C_N(\mathcal{R}A) = 0$.

If $C_N(\mathcal{R}A) = 0$, then $\mathcal{R}A$ fits into Case 1, being $(d+1)$-CM and therefore $A = \mathcal{R}A + b_A T_A$ is $(d+1)$-CM.

If $C_N(\mathcal{R}A) > 0$, then $\mathcal{R}A(M) = 0$ and $M_{\mathcal{R}A} \le M-1$. Now we consider $\mathcal{R}^k A$, $k = 2, 3, \cdots$. Since $M_{\mathcal{R}^k A} \ge 0$ for all k as long as $\mathcal{R}^k A \ne 0$, we have $C_N(\mathcal{R}^k A) = 0$ for some k. Thus $\mathcal{R}^k A$ is $(d+1)$-CM and so is $A = \mathcal{R}^k A + \sum_{i=0}^{k-1} b_{\mathcal{R}^i A} T_{\mathcal{R}^i A}$.

Now it is only left to show that the lemma holds for $N = 1$. Let T_A and M_A be defined as in Case 2. When $N = 1$, (iii) becomes $C_1(A) = 0$, therefore $C_1(T_A) = 0$ since $(T_A, 1)$ satisfies (iii). For $A(-1) = 0$, $A = 0$ on $S_1^d \setminus \{0\}$ and the lemma is trivial. For $A(-1) > 0$, let $b_A = A(M_A)/T_A(M_A)$ and $\mathcal{R}A := A - b_A T_A$. Similar to case 2, $\mathcal{R}A$ is still a mass function, $(\mathcal{R}A, 1)$ satisfies (i), (ii), (iii) and $\mathcal{R}A(M_A) = 0$. We consider $\mathcal{R}^k A$, $k = 2, 3, \cdots$ and eventually $M_{\mathcal{R}^k A} = 0$ for some k. Hence $R^k A$ is $(d+1)$-CM and so is A. This completes the proof. □

The following lemma is an immediate consequence of Lemma 4.4.3 and Lemma 4.4.4.

Lemma 4.4.5. *Suppose the probability mass function of a distribution P with mean 0 is decreasing on S_N^d and is 0 elsewhere, then P is $d+1$-CM.*

Now let us write any distribution with a monotone density as the limit of discrete distributions in Lemma 4.4.5. Let $S_N = \{-N/N, (-N+1)/N, \cdots, (dN-1)/N, dN/N\}$. For each continuous distribution P on $[-1, d]$ with mean zero and decreasing density, let $Y \sim P$. Denote $\bar{P}_N$ the distribution function of $\lfloor NY \rfloor / N$ and $\hat{P}_N$ the discrete uniform distribution on S_N. Since

$$-\frac{1}{N} \leq \int y \bar{P}_N(dy) \leq 0$$

and

$$\int y \hat{P}_N(dy) = \frac{d-1}{2} \geq \frac{1}{2},$$

there exists λ_N: $0 \leq \lambda_N < 2/N$ such that

$$\int y((1-\lambda_N)\bar{P}_N + \lambda_N \hat{P}_N)(dy) = 0.$$

Then the distributions $\{(1-\lambda)\bar{P}_N + \lambda \hat{P}_N\}$ are decreasing on S_N, with mean zero, and converge weakly to P as $N \to \infty$. This argument shows that there exist $P_k \xrightarrow{d} P$ and each P_k is $d+1$-CM and centered at 0. Then by Theorem 4.3.1, as the limit of completely mixable distributions, each continuous distribution P on $[-1, d]$ with mean 0 and decreasing density is $d+1$-CM.

Finally, by Proposition 4.2.1(1), each continuous distribution P on $[0, 1]$ with mean $1/n$ and decreasing density is n-CM. Just note that any decreasing density on $[0, 1]$ is also an decreasing density on $[0, a]$, hence each continuous distribution P on $[0, 1]$ with mean a/n, $a \geq 1$ and decreasing density is n-CM. Using Proposition 4.2.1(1) once again and the proof of Theorem 4.4.2 is complete. □

Remark 4.4.1. By Proposition 4.2.1(6), the condition in Theorem 4.4.2 is necessary and sufficient for a distribution P with monotone density on $[a, b]$ (where a and b are the infimum and the supremum of $\{x : p(x) > 0\}$) to be n-CM.

Remark 4.4.2. As a consequence of Theorem 4.4.2, the uniform distributions and distributions with a unimodal density (which are convex combinations of uniform distributions with the same center) are n-CM for $n \geq 2$. This is another proof of Theorem 4.4.1, different from the one given in [87].

4.4.2 Distributions with a concave density

In this section, we show that any continuous distribution with a concave density is completely mixable. Similarly to the method used in the proof of Theorem 4.4.2, we will first prove complete mixability of a particular class of discrete distributions with concave mass function.

Theorem 4.4.6. *Suppose F is a discrete distributions on the set*

$$S_{N,M} = \{-N, -N+1, \ldots, -1, 0, 1, \ldots, M-1, M\},\ N, M \in \mathbb{N}_0,$$

having mean $\mu = 0$ and mass function $f : S_{N,M} \to [0,1]$ satisfying $f(-N), f(M) > 0$ and

$$f(i-1) + f(i+1) \leq 2f(i),\ -N+1 \leq i \leq M-1. \tag{4.13}$$

Then, F is n-CM for any $n \geq 3$.

In order to prove Theorem 4.4.6, we need the following lemma.

Lemma 4.4.7. *Under the assumptions of Theorem 4.4.6, we have*

$$M \leq 2N \text{ and } N \leq 2M.$$

Proof. We only need to prove that $M \leq 2N$, as $N \leq 2M$ follows by symmetry. The condition $\mu = 0$ implies that $M = 0$ if and only if $N = 0$, thus we can assume M, N to be both positive. It is easy to see that (4.13) is equivalent to

$$A(v) \geq \frac{(w-v)A(u) + (v-u)A(w)}{w-u}, \tag{4.14}$$

for all $u, v, w \in S_{N,M}$ such that $u \leq v \leq w$ and $u < w$. For instance, the two inequalities

$$f(v) \geq \frac{f(v-1) + f(v+1)}{2} \text{ and } f(v-1) \geq \frac{f(v-2) + f(v)}{2}$$

imply

$$f(v) \geq \frac{f(v-2) + 2f(v+1)}{3}.$$

As particular cases of (4.14), we get

$$f(i) \geq \frac{(M-i)f(0)+if(M)}{M} > \frac{M-i}{M}f(0),\ 0 \leq i \leq M, \tag{4.15a}$$

$$f(0) \geq \frac{Mf(-j)+jf(M)}{M+j} > \frac{M}{M+j}f(-j),\ 0 \leq j \leq N. \tag{4.15b}$$

Since $\mu = \sum_{i \in S_{N,M}} if(i) = 0$, (4.15) implies that

$$\frac{f(0)M(M-1)(M+1)}{6M} = \frac{f(0)}{M}\sum_{i=1}^{M} i(M-i)$$
$$< \sum_{i=1}^{M} if(i) = \sum_{j=0}^{N} jf(-j) < \frac{f(0)}{M}\sum_{j=1}^{N} j(M+j) = \frac{f(0)N(N+1)(3M+2N+1)}{6M},$$

from which we have

$$M(M+1)(M-1) < N(N+1)(3M+2N+1).$$

In the above equation, the right-hand side is increasing in N and equality holds when $N = (M+1)/2$. Therefore, we have $N > (M-1)/2$, namely $M \leq 2N$. □

Proof of Theorem 4.4.6. We will prove the theorem by induction over $M+N$, the cardinality of the set $S_{N,M}$. Note that, if $M = N = 0$, F is the unit mass at 0 and thus is completely mixable for any n. Moreover, the case $M+N = 1$ is not allowed by the zero mean condition. Therefore, the first step of the induction will be $M+N = 2$. In this case the zero mean condition combined with (4.13) forces F to be supported on $\{-1, 0, 1\}$ with masses $f(-1) = f(1) = a$ and $f(0) = 1-2a$ with $a < 0 \leq 1/3$. We can write F as

$$F = (3a)G + (1-3a)H, \tag{4.16}$$

where G is the uniform distribution on $\{-1, 0, 1\}$ and H is the unit mass at 0. Being a unit mass, H is n-CM for any $n \in \mathbb{N}$, while G satisfies the assumptions of Lemma 4.4.5 with $d = n-1$ and, then , is n-CM for any $n \geq 2$. Equation (4.16) states that F is the convex sum of two n-CM distributions with center $\mu = 0$. By Theorem 4.3.3(a), F is n-CM, for any $n \geq 2$.

Now, we assume that the theorem holds for any distribution H satisfying the assumption of the theorem with $N+M \leq (K-1)$ points in $S_{N,M}$ and prove that it holds for any distribution F with K points in $S_{N,M}$, $K \geq 3$. As illustrated for $N+M=2$, the idea of the proof is to decompose F as the convex sum of such an H and another n-CM distribution G.

Let F a distribution satisfying the assumption of the theorem with $N+M = K, K \geq 3$. W.l.o.g., in what follows we assume $M \geq N$ (the theorem holds symmetrically for $N \leq M$). We denote by G the discrete distribution having mass function $g : S_{N,M} \to [0,1]$ given by

$$g(-N) = \frac{(M-N+1)}{(M+N+1)},\ g(-N+1) = \cdots = g(M) = \frac{2N}{(M+N+1)(M+N)}.$$

Elementary calculations show that the distribution G has first moment $\mu = 0$ and, being $M \geq N$, that g is decreasing. From Lemma 4.4.7, we have that $M \leq 2N \leq (n-1)N$ for any $n \geq 3$, and, then, the distribution G satisfies the assumption of Lemma 4.4.5 with $d = n-1$. As a consequence, G is n-CM. Now, we define the function $\hat{f} : S_{N,M} \to \mathbb{R}$ as

$$\hat{f} = f - k_1 g, \tag{4.17}$$

where

$$k_1 = \min\left\{\frac{f(-N)}{g(-N)}, \frac{f(M)}{g(M)}\right\} > 0.$$

Note that we have

$$\hat{f}(-N) = f(-N) - k_1 g(-N) \geq f(-N) - \frac{f(-N)}{g(-N)} g(-N) = 0, \tag{4.18a}$$

$$\hat{f}(M) = f(M) - k_1 g(M) \geq f(M) - \frac{f(M)}{g(M)} g(M) = 0. \tag{4.18b}$$

Since g is convex on $S_{N,M}$, the function $\hat{f}$ is the sum of two concave densities and, therefore, is concave. Concavity of $\hat{f}$, combined with (4.18), implies that $\hat{f}$ is also nonnegative on $S_{N,M}$. At this point, it is possible to define the discrete distribution

H as the one having concave mass function

$$h = \hat{f}/k_2, \tag{4.19}$$

where

$$k_2 = \sum_{i \in S_{N,M}} \hat{f}(i).$$

Note that the distribution H has mean $\mu = 0$ as

$$\sum_{i=-N}^{M} ih(i) = \frac{1}{k_2}\left(\sum_{i=-N}^{M} if(i) - k_1 \sum_{i=-N}^{M} ig(i)\right) = 0.$$

Moreover, at least one of the values $\hat{f}(-N)$ and $\hat{f}(M)$ is equal to zero. In conclusion, H is a distribution function on a subset of $S_{N,M}$ containing at most $K-1$ points, having mean $\mu = 0$ and concave mass function h. By the induction assumption, H is n-CM. Combining (4.17) and (4.19), we obtain that

$$F = k_1 G + k_2 H, \text{ with } k_1 + k_2 = 1.$$

Thus, F is the convex combination of two n-CM distributions and, then, F is n-CM. □

Theorem 4.4.8. *Any continuous distribution on a bounded interval (a, b) having a concave density is n-CM for any $n \geq 3$.*

Proof. The proof is analogous to the part of the proof of Theorem 4.4.2 following Lemma 4.4.5. For any F with a concave density, we find a sequence of discrete concave distributions that goes to F. Note that a distribution with concave density on $(0, 1)$ is n-CM for all $n \geq 3$, hence the mean condition

$$1/n \leq \mu \leq 1 - 1/n$$

is automatically satisfied for $n \geq 3$. □

According to Theorem 4.4.8, The Beta(α, β) distribution with parameters $1 \leq \alpha, \beta \leq 2$ is n-completely mixable for $n \geq 3$. Any triangular distribution has a concave density and hence it is n-completely mixable for $n \geq 3$.

4.4.3 Radially symmetric distributions

In this section, we show that any n-radially symmetric distribution is completely mixable. The definition of an n-radially symmetric distribution which we give here is an extension of the one introduced in Knott and Smith [61].

Definition 4.4.1. Suppose that U is a random variable uniformly distributed on $(0,1)$ and let $\mathbf{A} = (A_1, \ldots, A_n)$, $\mathbf{B} = (B_1, \ldots, B_n)$ be two random vectors on $\mathbb{R}^n$ independently distributed from U. A random variable X and its distribution are called $n-$*radially symmetric* if

$$X = a + \sum_{k=1}^{n} \left(A_k \cos(2\pi kU) + B_k \sin(2\pi kU)\right), \tag{4.20}$$

for some constant $a \in \mathbb{R}$.

In the above definition, the random vectors $\mathbf{A}$ and $\mathbf{B}$ can be chosen to have an arbitrary distribution on $\mathbb{R}^n$.

Theorem 4.4.9. *Any n-radially symmetric distribution is m-CM for any $m \geq n+1$.*

Proof. Let F be the $n-$radially symmetric distribution of a random variable X of the form (4.20), for some U uniformly distributed on $(0,1)$ and $\mathbf{A}$ and $\mathbf{B}$ distributed independently from U. Fixed an integer $m \geq n+1$, let the m random variables $X_1, \ldots, X_m$ be defined as

$$X_i = a + \sum_{k=1}^{n} \left(A_k \cos\left(2\pi k \left(V + \frac{i}{m}\right)\right) + B_k \sin\left(2\pi k\left(V + \frac{i}{m}\right)\right)\right), \ 1 \leq i \leq m,$$

where V is random variable uniformly distributed on $(0,1)$ and independent from $\mathbf{A}$ and $\mathbf{B}$. Note that

$$\cos\left(2\pi k\left(V + \frac{i}{m}\right)\right) \sim \cos(2\pi kU) \text{ and } \sin\left(2\pi k\left(V + \frac{i}{m}\right)\right) \sim \sin(2\pi kU),$$

for $1 \leq i \leq m$ and $1 \leq k \leq n$. Therefore, the X_i's are all identically distributed as F. To complete the proof, we show that their sum is, a.s, the constant ma.

For $1 \leq i \leq m$, let $\xi_i = e^{\mathbf{i}2\pi ki/m}$, where $\mathbf{i}$ is the imaginary unit. We denote by $d_k = \gcd(k, m)$ the greatest common divisor of k and m. Since $m \geq n+1$, we have that $k \leq n \leq m-1$ and, thus, $d_k < m$ for $1 \leq k \leq n$. When $d_k = 1$, the m values $\xi_1, \ldots, \xi_m$ are all the roots of the equation $\xi^m = 1$ and, therefore, $\sum_{i=1}^m \xi_i = 0$. If, instead, $1 < d_k < m$, then the m/d_k values $\xi_1, \ldots, \xi_{m/d_k}$ are all the roots of the equation $\xi^{m/d_k} = 1$ and, again, we have $\sum_{i=1}^m \xi_i = d_k \sum_{i=1}^{m/d_k} \xi_i = 0$. From this, it easily follows that

$$\begin{aligned} \sum_{i=1}^m \left(\cos\left(2\pi k \left(V + \frac{i}{m}\right)\right) + \mathbf{i} \sin\left(2\pi k \left(V + \frac{i}{m}\right)\right)\right) &= \sum_{i=1}^m e^{\mathbf{i}2\pi k(V+i/m)} \\ &= e^{\mathbf{i}2\pi kV} \sum_{i=1}^m \xi_i = 0. \end{aligned}$$

The above equality implies that

$$\sum_{i=1}^k \cos\left(2\pi k\left(V + \frac{i}{m}\right)\right) = \sum_{i=1}^k \sin\left(2\pi k\left(V + \frac{i}{m}\right)\right) = 0$$

and, therefore, that

$$\begin{aligned} \sum_{i=1}^m X_i &= ma + \sum_{i=1}^m \sum_{k=1}^n \left(A_k \cos\left(2\pi k\left(V + \frac{i}{m}\right)\right) + B_k \sin\left(2\pi k\left(V + \frac{i}{m}\right)\right)\right) \\ &= ma + \sum_{k=1}^n \left(A_k \sum_{i=1}^m \cos\left(2\pi k\left(V + \frac{i}{m}\right)\right) + B_k \sum_{i=1}^m \sin\left(2\pi k\left(V + \frac{i}{m}\right)\right)\right) \\ &= ma. \end{aligned}$$

□

An interesting example of a radially symmetric distribution is given by the continuous random variable $X = \cos(2\pi U)$, where U is uniformly distributed on $(0, 1)$. By Theorem 4.4.9, the distribution of X is n-CM for $n \geq 2$. As illustrated in Figure 4.1, the density of X is a convex function on the interval $[-1, 1]$. Therefore, Theorem 4.4.9 indicates that there exist continuous n-CM distributions with a large density at both endpoints of their support. As the set of n-CM distributions with a given center is convex, Theorem 4.4.9 is no doubt useful to construct new classes of completely mixable distributions.

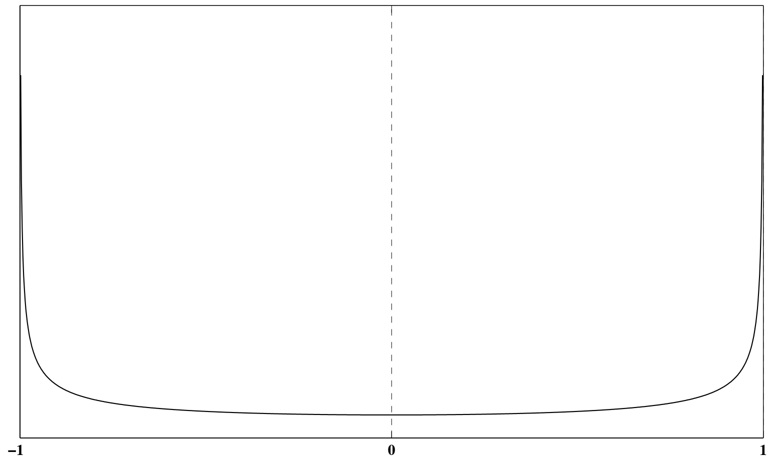

Figure 4.1: The density of the random variable $X = \cos(2\pi U)$.

4.5 Convex minimization problems

Fréchet class problems are of great interest in actuarial science, and mathematical finance, as introduced in Chapter I.

In this section, we study a convex minimization problem in a homogenous Fréchet class using the idea of CM distributions. Throughout this section, let $\mathbf{X} = (X_1, \cdots, X_n)$, $S = X_1 + \cdots + X_n$ and define the homogenous Fréchet class as Section 4.1 and Chapter I

$$\mathfrak{F}_n(F) = \mathfrak{F}_n(F, \cdots, F) = \{\mathbf{X} : \ X_i \sim F, \ i = 1, \cdots, n\}.$$

$\mathfrak{F}_n(F)$ is the set of random vectors with a given marginal distribution F. We are interested in the total risk S when $\mathbf{X} \in \mathfrak{F}_n(F)$.

Question A: the expectation of a convex function. Find

$$\inf_{\mathbf{X} \in \mathfrak{F}_n(F)} \mathbb{E}[f(S)] \tag{4.21}$$

for f being a convex function.

The expectation of convex (concave) functions plays an important role in the study of insurance, finance, and economics. For instance, $\mathbb{E}[f(S)]$ includes important quantities such as the variance, stop-loss premium, excess of loss, prices of the European options and multivariate Spearman's rho. More over, risk-avoiding (risk-seeking) utility functions are concave (convex), while $\mathbb{E}[f(S)]$ is the expected utility. $\mathbb{E}[f(S)]$ also appears in the convex ordering and optimization problems. Therefore, Question A is related to various topics in statistics, risk theory, copulas and stochastic orders. We refer to Rüschendorf and Uckelmann [87] and Hammersley and Handscomb [43] for variante minimization problems, Embrechts, Lindskog and McNeil [33] and Embrechts, McNeil and Straumann [34] for problems of bounds in risk theory, Nelsen [69] for copulas, Joe [48] for Fréchet classes and Shaked and Shanthikumar [92] for stochastic orders.

As introduced in Chapter I, it is well-known that the maximum of $\mathbb{E}[f(S)]$ over $\mathbf{X} \in \mathfrak{F}_n(F)$ is obtained by letting $X_1 = \cdots = X_n$. However, the infimum stays a mystery for $n \geq 3$.

Jensen's inequality connects Question A with the CM distributions.

Proposition 4.5.1. *If f is a (strictly) convex function and $\mu < \infty$ is the mean of F, then*

$$\inf_{\mathbf{X}\in\mathfrak{F}(F;n)} \mathbb{E}[f\left(X_1 + \cdots + X_n\right)] \geq f(n\mu), \tag{4.22}$$

and the equality in (4.22) *holds if (and only if) F is n-CM.*

This is a direct application of the Jensen's inequality. Thus, the identification of CM distributions immediately leads to the solution to Question A.

Practically, some risks are unbounded from one side, hence the distribution violates the mean condition (4.5) for CM. Risks with a decreasing density, such as Perato or exponentially distributed risks, are commonly used in practice. Hence, in the following we solve Question A for distributions with a monotone density. We first introduce a class of copulas Q_n^P in Section 4.5.1. Then we give our main theorem for problem (4.21) in Section 4.5.2 and illustrate some applications in Section 4.5.3.

4.5.1 The copula Q_n^P

Let P be a distribution with monotone density on its support, and $f:\mathbb{R}\to\mathbb{R}$ is a convex function. In the following we denote G the inverse cdf of $Y_i \sim P$, then $Y_i = G(X_i)$ for some $X_i \sim \mathrm{U}[0,1]$, $i = 1,\cdots,n$ and (4.21) reads as

$$\min_{\mathbf{X}\in\mathfrak{F}_n(\mathrm{U}[0,1])} \mathbb{E}f\left(G(X_1)+\cdots+G(X_n)\right) = \min_{C\in\mathfrak{C}_n} \int G(x_1)+\cdots+G(x_n)dC(x_1,\cdots,x_n), \tag{4.23}$$

where $\mathfrak{C}_n$ is the set of all n-copulas. In the following, we use the setting (4.23) for Question A, and $X_1,\cdots,X_n$ represent uniform [0,1] random variables.

Remark 4.5.1.

1. P having an increasing (decreasing) density is equivalent to G being continuous and concave (convex). Thus both f and G have convexity in this problem and another equivalent setting for (4.21) is

$$\min_{\mathbf{X}\in\mathfrak{F}_n(\mathrm{U}[0,1])} \mathbb{E}f\left(G(X_1)+\cdots+G(X_n)\right)$$

   for $f:\mathbb{R}\to\mathbb{R}$ being convex and $G:[0,1]\to\mathbb{R}$ being concave (convex), continuous and increasing.

2. If $X\sim P$ and P has decreasing density, we can simply replace X by $-X$ (note that $f(-x)$ is also convex). Thus without loss of generality, in the following we will assume P has increasing density.

To obtain an optimal coupling for problem (4.23), we construct n-copulas $Q_n^P(c)$ ($n\geq 2$) for some $0\leq c\leq 1/n$. For P with an increasing density and a constant $c\in[0,\frac{1}{n}$, we define a copula by $Q_n^P(c)$, if $(X_1,\cdots,X_n)\sim Q_n^P(c)$ satisfying

(a) For each $i=1,\cdots,n$, the joint-density of $X_1,\cdots,X_n$ given $X_i\in[0,c]$ is uniformly supported on line segments $x_j = 1-(n-1)x_i,\ \forall j\neq i,\ x_i\in[0,c]$; and

(b) $G(X_1)+\cdots+G(X_n)$ is a constant when $X_i\in(c,1-(n-1)c)$ for any $i=1,\cdots,n$.

Note that such a copula may not exist for some $c > 0$.

Proposition 4.5.2. *Denote*

$$H(x) = G(x) + (n-1)G(1-(n-1)x). \tag{4.24}$$

There exists a copula $Q_n^P(c)$ *satisfying* (a) *and* (b) *if*

$$\int_c^{\frac{1}{n}} H(t)dt \leq (\frac{1}{n} - c)H(c). \tag{4.25}$$

Proof. We first take random variables $Y_1, \cdots, Y_n \sim \mathrm{U}([0,c] \cup [1-(n-1)c, 1])$ such that the joint-density of $Y_1, \cdots, Y_n$ is uniformly supported on each line segment $y_j = 1-(n-1)y_i,\ \forall j \neq i,\ y_i \in [0,c]$. By Theorem 4.4.2, there exist $Z_1, \cdots, Z_n \sim \mathrm{U}[c, 1-(n-1)c]$ such that $G(Z_1) + \cdots + G(Z_n)$ is a constant since $G(Z_i)$ has an increasing density and that (4.25) implies

$$\mathbb{E}(G(Z_1)) \leq G(c) + \frac{n}{n-1}[G(1-(n-1)c) - G(c)].$$

Let $U \sim \mathrm{U}[0,1]$ be independent of $(Y_1, \cdots, Y_n, Z_1, \cdots, Z_n)$ and $X_i = I_{\{U<nc\}}Y_i + I_{\{U \geq nc\}}Z_i$, then $X_i \sim \mathrm{U}[0,1]$ for $i = 1, \cdots, n$. Properties (a) and (b) are satisfied by the joint distribution of $X_1, \cdots, X_n$, which shows that $Q_n^P(c)$ exists. □

Remark 4.5.2.

1. Property (a) describes the joint distribution on the set $\bigcup_{i=1}^n \{0 \leq x_i \leq c, 1-(n-1)c \leq x_j \leq 1, j \neq i\}$, and property (b) describes it on the set $(c, 1-(n-1)c)^n$. These two sets are disjoint and their union is $[0,1]^n$.

2. The key idea of constructing $Q_n^P(c)$ is that when X_i is small, we let other random variables $X_j,\ j \neq i$ be large. When each of X_i, $i = 1, \cdots, n$ is of medium size, we let $G(X_1) + \cdots + G(X_n)$ be a constant. This could be a good candidate of optimal coupling since the variance of $G(X_1) + \cdots + G(X_n)$ is largely reduced. Later we will show that $Q_n^P(c)$ is optimal for the smallest possible c.

3. $Q_n^P(c)$ does not always exist for arbitrary c and it may not be unique while exists. However, when $\mathbf{X} \sim Q_n^P(c)$, $\mathbb{E}[f(G(X_1) + \cdots + G(X_n)]$ is determined by properties (a) and (b). Therefore, in the following $Q_n^P(c)$ is just one representative in the family of copulas satisfying (a) and (b).

4. It is easy to check that when $Q_2^P(c)$ exists, it is exactly the Fréchet-Hoeffding lower bound $W_2(u,v) = (u+v-1)_+$.

We denote c_n the smallest c such that $Q_n^P(c)$ exists and let $Q_n^P := Q_n^P(c_n)$. Note that $c_n = 0$ if and only if P is n-CM. In the following we will find c_n.

Proposition 4.5.3. *The smallest possible c is given by*

$$c_n = \min\{c \in [0, \frac{1}{n}] : \int_c^{\frac{1}{n}} H(t)dt \leq (\frac{1}{n} - c)H(c)\}. \tag{4.26}$$

Proof. Suppose $Q_n^P(c)$ exists. By (b), when any of $X_i \in (c, 1-(n-1)c)$, $G(X_1) + \cdots + G(X_n)$ is a constant, namely

$$\begin{aligned} G(X_1) + \cdots + G(X_n) &= \mathbb{E}(G(X_1) + \cdots + G(X_n) | c \leq X_i \leq 1-(n-1)c) \\ &= \frac{n}{1-nc} \int_c^{1-(n-1)c} G(t)dt. \end{aligned}$$

Noting that the conditional distribution of $G(X_i)$ on the set $\{X_i \in (c, 1-(n-1)c)\}$ is completely mixable, by Proposition 4.2.1(6) its conditional mean is less than or equal to $G(c)/n + (n-1)G(1-(n-1)c)/n$. Thus we have a necessary condition on c,

$$\int_c^{1-(n-1)c} G(t)dt \leq (\frac{1}{n} - c)[G(c) + (n-1)G(1-(n-1)c)]. \tag{4.27}$$

Together with (4.24), we obtain (4.25) from (4.27).

Note that $H(x)$ is concave on $[0, \frac{1}{n}]$ since $G(x)$ is concave. Hence the set of c satisfying (4.27) is a closed interval $[\hat{c}_n, \frac{1}{n}]$. (4.25) becomes $\hat{c}_n \leq c \leq \frac{1}{n}$ and therefore $c_n \geq \hat{c}_n$. By Proposition 4.5.2 we know $Q_n^P(\hat{c}_n)$ exists and thus $c_n = \hat{c}_n$. □

Now we have c_n and $Q_n^P = Q_n^P(c_n)$.

4.5.2 Main theorem

In the next we show the minimality of Q_n^P. The following lemma (see Theorem 3.A.5 in Shaked and Shanthikumar [92]) will be used.

Lemma 4.5.4. *Suppose X and Y with distribution functions F_1, F_2 respectively satisfy $\mathbb{E}X = \mathbb{E}Y$ and for any c in $[0,1]$, $\int_0^c F_1^-(t)dt \geq \int_0^c F_2^-(t)dt$, where $F_1^-(t) = \sup\{x : F_1(x) < t\}$ and $F_2^-(t) = \sup\{y : F_2(y) < t\}$. Then for any convex function f, $\mathbb{E}(f(X)) \leq \mathbb{E}(f(Y))$.*

Theorem 4.5.5. *Suppose P is a distribution with increasing density and G is the inverse cdf of P, then for any convex function f,*

$$\min_{(Z_1,\cdots,Z_n)\in\mathfrak{F}_n(P)} \mathbb{E}f(Z_1 + \cdots + Z_n) = \mathbb{E}f\left(G(X_1) + \cdots + G(X_n)\right), \tag{4.28}$$

where $(X_1, \cdots, X_n) \sim Q_n^P$.

Proof. Let $(X_1, \cdots, X_n) \sim Q_n^P$ and $Z_i = G(Y_i)$ where $Y_i \sim \mathrm{U}[0,1]$, $i = 1, \cdots, n$. Denote $X = G(X_1) + \cdots + G(X_n)$ and $Y = G(Y_1) + \cdots + G(Y_n)$. Let F_1 and F_2 be the cdf of X and Y respectively, $F_1^-(t) = \sup\{x : F_1(x) < t\}$ and $F_2^-(t) = \sup\{y : F_2(y) < t\}$. We will show that for any $c \in [0,1]$,

$$\int_0^c F_1^-(t)dt \geq \int_0^c F_2^-(t)dt.$$

To obtain this, denote $A_X(u) = \bigcup_i\{X_i < u\}$, $A_Y(u) = \bigcup_i\{Y_i < u\}$ and let $W(u) = \mathbb{P}(A_Y(u))$. Obviously $u \leq W(u) \leq nu$ and W is invertible. For $c \in [0, nc_n]$, let $u^\star = W^{-1}(c)$, it then follows that $c \geq u^\star \geq c/n$ and $\{Y_i \in [0, c/n]\} \subset \{Y_i \in [0, u^\star]\} \subset A_Y(u^\star)$.

By the definition of Q_n^P, for each i, $\{X_i \in [0, c/n] \cup [1-(n-1)c/n, 1]\} = A_X(c/n)$. Note that $X_i \stackrel{d}{=} Y_i \sim \mathbb{U}$ and $\mathbb{P}(A_X(c/n)) = \mathbb{P}(A_Y(u^\star)) = c$, therefore

$$\mathbb{P}(A_Y(u^\star) \setminus \{Y_i \in [0, c/n]\}) = c - c/n = \mathbb{P}(Y_i \in [1-(n-1)c/n, 1]).$$

Since G is increasing and the above two sets are equally measured, we have

$$\mathbb{E}[I_{\{Y_i\in[1-(n-1)c/n,1]\}}G(Y_i)] \geq \mathbb{E}[I_{A_Y(u^\star)\setminus\{Y_i\in[0,c/n]\}}G(Y_i)].$$

It follows that

$$\begin{aligned}
\mathbb{E}(I_{A_X(c/n)}G(X_i)) &= \mathbb{E}[(I_{\{X_i\in[0,c/n]\}} + I_{\{X_i\in[1-(n-1)c/n,1]\}})G(X_i)] \\
&= \mathbb{E}[(I_{\{Y_i\in[0,c/n]\}} + I_{\{Y_i\in[1-(n-1)c/n,1]\}})G(Y_i)] \\
&\geq \mathbb{E}[(I_{\{Y_i\in[0,c/n]\}} + I_{A_Y(u^\star)\setminus\{Y_i\in[0,c/n]\}})G(Y_i)] \\
&= \mathbb{E}(I_{A_Y(u^\star)}G(Y_i)).
\end{aligned}$$

Thus we have

$$\mathbb{E}(I_{A_X(c/n)}X) \geq \mathbb{E}(I_{A_Y(u^\star)}Y). \tag{4.29}$$

Note that $H(x)$ is concave and differentiable. By the definition of c_n, the mean of $H(x)$ on $[c_n, \frac{1}{n}]$ is $H(c_n)$. With $H(x)$ being concave, we have $H'(c_n) \geq 0$ and thus $H(x)$ is increasing on $[0, c_n]$. Note that on the set $A_X(c_n)$,

$$X = \sum_{i=1}^{n} I_{\{X_i<c_n\}}[G(X_i) + (n-1)G(1-(n-1)X_i)] = \sum_{i=1}^{n} I_{\{X_i<c_n\}}H(X_i),$$

and the events $\{X_i < c_n\}$ $i = 1, \cdots, n$ are disjoint. It follows that for $t \leq H(c_n)$, $F_1(t) = \mathbb{P}(X \leq t) = n\mathbb{P}(H(X_1) \leq t)$. Thus for $c \leq nc_n$, $F_1^-(c) = H(c/n)$ and

$$\mathbb{E}(I_{A_X(c/n)}X) = n\int_0^{c/n} H(t)dt = \int_0^c H(t/n)dt = \int_0^c F_1^-(t)dt. \tag{4.30}$$

Also note that

$$\mathbb{E}(I_{A_Y(u^\star)}Y) \geq \int_0^c F_2^-(t)dt \tag{4.31}$$

since $\mathbb{P}(A_Y(u^\star)) = c$. It follows from (4.39), (4.30) and (4.31) that for any $c \in [0, nc_n]$,

$$\int_0^c F_1^-(t)dt \geq \int_0^c F_2^-(t)dt.$$

For $c \in (nc_n, 1]$, note that $H_1(x) := \int_0^x F_1^-(t)dt$ and $H_2(x) := \int_0^x F_2^-(t)dt$ are convex functions and $\mathbb{E}(X) = \mathbb{E}(Y)$ thus $H_1(1) = H_2(1)$. Furthermore we have $F_1^-(t)$

is a constant when $t \geq c_n$ since Q_n^P satisfies (b). By the facts that $H_1(c_n) \geq H_2(c_n)$, $H_1(1) = H_2(1)$, H_1 is linear over $[nc_n, 1]$ and H_1, H_2 are convex, we conclude

$$\int_0^c F_1^-(t)dt \geq \int_0^c F_2^-(t)dt$$

for any $c \in [0,1]$. By Lemma 4.5.4 we obtain

$$\mathbb{E}f(G(Y_1) + \cdots + G(Y_n)) \leq \mathbb{E}f\left(G(X_1) + \cdots + G(X_n)\right)$$

and it completes the proof. □

Remark 4.5.3.

1. In stochastic orderings, the above result is interpreted in the following way: suppose $Y_1, \cdots, Y_n, Z_1, \cdots, Z_n \sim P$ and $Z_1, \cdots, Z_n$ have copula Q_n^P, then

$$Z_1 + \cdots + Z_n \leq_{\mathrm{cx}} Y_1 + \cdots + Y_n \leq_{\mathrm{cx}} nY_1.$$

   Thus $Z_1 + \cdots + Z_n$ is the lower bound in the convex order on the sum $Y_1 + \cdots + Y_n$ with given marginal distributions $Y_i \sim P$. This completes the result of bounds in the convex order on the sum in the Fréchet class $\mathfrak{F}_n(P)$. For an overview of the stochastic orderings, see Shaked and Shanthikumar [92].

2. The optimal copula Q_n^P solving (4.21) depends only on the marginal distribution P, but not on the convex function f.

3. Although we are able to show the existence and minimality, we are unable to write the function Q_n^P explicitly.

Theorem 4.5.6. *We have*

$$\min_{(Y_1,\cdots,Y_n)\in\mathfrak{F}_n(P)} \mathbb{E}f\left(Y_1 + \cdots + Y_n\right) = n\int_0^{c_n} f(H(x))dx + (1 - nc_n)f(H(c_n)), \quad (4.32)$$

where $H(x)$ and c_n are defined as in (4.24) *and* (4.26).

Proof. Let $(X_1, \cdots, X_n) \sim Q_n^P$. By Theorem 4.5.5,

$$
\begin{aligned}
& \min_{(Y_1,\cdots,Y_n)\in\mathfrak{F}_n(P)} \mathbb{E}f\left(Y_1+\cdots+Y_n\right) \\
= \; & \mathbb{E}f\left(G(X_1)+\cdots+G(X_n)\right) \\
= \; & n\mathbb{E}[f(G(X_1)+\cdots+G(X_n))I_{\{X_1\in[0,c_n]\}}] \\
& +\mathbb{E}[f(G(X_1)+\cdots+G(X_n))I_{\{X_1\in[c_n,1-(n-1)c_n]\}}] \\
= \; & n\mathbb{E}[f(H(X_1))I_{\{X_1\in[0,c_n]\}}]+\mathbb{E}[f(H(c_n))I_{\{X_1\in[c_n,1-(n-1)c_n]\}}] \\
= \; & n\int_0^{c_n} f(H(x))dx+(1-nc_n)f(H(c_n)).
\end{aligned}
$$

□

Corollary 4.5.7. *If the density of P is monotone and supported in a finite interval $[a, b]$, then*

$$\min_{(X_1,\cdots,X_n)\in\mathfrak{F}_n(P)} \mathbb{E}f\left(X_1+\cdots+X_n\right)=f(n\mu)$$

for n sufficiently large, where μ is the mean of P.

Proof. We have $a < \mu < b$ since P is a continuous distribution. Hence there exists N such that $b - \frac{1}{n}(b-a) > \mu$ for $n \geq N$. By Theorem 3.3.2 we know P is n-CM and centered at μ. Thus we have

$$\mathbb{E}[f(n\mu)] \geq \min_{(X_1,\cdots,X_n)\in\mathfrak{F}_n(P)} \mathbb{E}f\left(X_1+\cdots+X_n\right) \geq f(n\mu)$$

by Jensen's inequality. This shows that

$$\min_{(X_1,\cdots,X_n)\in\mathfrak{F}_n(P)} \mathbb{E}f\left(X_1+\cdots+X_n\right)=f(n\mu)$$

for n sufficiently large. □

4.5.3 Examples

The minimum of the expected product of uniform random variables. Let us look at the problem

$$\Lambda_n := \min_{X_1,\cdots,X_n\sim\mathbb{U}} \mathbb{E}(X_1X_2\cdots X_n). \tag{4.33}$$

Problem (4.33) has a long history. For $n = 3$ and $X, Y, Z \sim \mathrm{U}[0,1]$, Rüschendorf [86] found 1/24 as a lower bound for $\mathbb{E}(XYZ)$, but apparently the bound is not sharp. Baiocchi [6] constructed a discretization of X, Y and Z and applied a linear programming to approximate the minimum, which leads to a value ≈ 0.06159. Bertino [8] obtained an upper bound ≈ 0.05481 for Λ_3, by manually taking the limit of one class of discretizations of X, Y, Z. He conjectured that this upper bound was the true value of Λ_3. Recently, Nelsen and Ubeda-Flores [70] introduced the coefficients of directional dependence, whose lower bound has not been found and equals a function of the lower bound for $\mathbb{E}(XYZ)$.

This problem is a special case of problem (4.1). By letting P be the distribution of $\log(X)$, $X \sim \mathrm{U}[0,1]$ (namely, $P = -\mathrm{Expo}(1)$) and $f(x) = \exp(x)$, we can use Theorem 4.5.5 and Theorem 4.5.6 to solve (4.33).

Corollary 4.5.8. *Let $(X_1, \cdots, X_n) \in \mathfrak{F}_n(P)$ have copula Q_n^P. We have*

$$\begin{aligned}\Lambda_n =&\mathbb{E}(X_1 \cdots X_n)\\ =&\frac{1}{(n-1)^2}\left(\frac{1}{n+1} - (1-(n-1)c_n)^n + \frac{n}{n+1}(1-(n-1)c_n)^{n+1}\right) \\ &+ (1-nc_n)c_n(1-(n-1)c_n)^{n-1},\end{aligned} \tag{4.34}$$

where c_n is the unique solution to

$$\log(1-(n-1)c) - \log(c) = n - n^2 c,\ 0 \le c < 1/n. \tag{4.35}$$

It is an immediate application of Theorem 4.5.5 and Theorem 4.5.6, hence we omit the proof here.

The numerical values of Λ_n for different n are presented in Table 4.1. One may suggest that $\Lambda_n \sim e^{-n}$ as n goes to infinity.

Corollary 4.5.9. *We have*

$$\Lambda_n = e^{-n} + \frac{n}{2}e^{-2n} + O(n^4 e^{-3n}).$$

n	Λ_n	c_n	e^{-n}	$\Lambda_n e^n$
1	$1/2$	N/A	3.6788×10^{-1}	1.3591
2	$1/6$	$1/2$	1.3533×10^{-1}	1.2315
3	5.4803×10^{-2}	9.4542×10^{-2}	4.9787×10^{-2}	1.1008
4	1.9098×10^{-2}	2.5406×10^{-2}	1.8316×10^{-2}	1.0427
5	6.8604×10^{-3}	7.9597×10^{-3}	6.7379×10^{-3}	1.0182
10	4.5410×10^{-5}	4.5589×10^{-5}	4.5400×10^{-5}	1.0002
20	2.0612×10^{-9}	2.0612×10^{-9}	2.0612×10^{-9}	1.0000
50	1.9287×10^{-22}	1.9287×10^{-22}	1.9287×10^{-22}	1.0000
100	3.7201×10^{-44}	3.7201×10^{-44}	3.7201×10^{-44}	1.0000

Table 4.1: Numerical values of Λ_n

See Section 4.7 for the proof.

Remark 4.5.4.

1. In fact this approximating procedure can be done infinitely further. For $n = 10$, $\Lambda_{10} - e^{-10} = 1.0323 \times 10^{-8}$, $5e^{-20} = 1.0306 \times 10^{-8}$. We cam see that the approximation is already very precise.

2. Nelsen and Ubeda-Flores [70] introduced the directional dependence coefficients $\rho_n^{(\alpha_1 \cdots, \alpha_n)}$, $\alpha_i \in \{-1, 1\}$, $i = 1, \cdots, n$. The lower bound on $\rho_n^{(\alpha_1 \cdots, \alpha_n)}$ can be written as

$$\rho_n^{(\alpha_1 \cdots, \alpha_n)} \geq \min_{X_1, \cdots, X_n \sim \mathbb{U}} \{2^n \mathbb{E}(X_1 \cdots X_n) - 1\} = 2^n \Lambda_n - 1,$$

and Corollary 4.5.8 provides this value.

Stop-loss premiums of the total risk. Let $X_1, X_2, \cdots, X_n \geq 0$ be n individual risks with the same marginal distributions P. Their stop-loss premium is defined as $\mathbb{E}[(X_1 + \cdots + X_n - t)_+]$ where $t \geq 0$ is a constant and $(\cdot)_+ = \max\{\cdot, 0\}$. See Kaas, Goovaerts, Dhaene and Denuit [54] for references of this topic. An important problem in variance reduction is to determine the minimum of the stop-loss premium over all possible dependence structure, i.e.

$$\min_{X_1, \cdots, X_n \sim P} \mathbb{E}[(X_1 + \cdots + X_n - t)_+] = \min_{C \in \mathfrak{C}_n} \mathbb{E}[(G(U_1) + \cdots + G(U_n) - t)_+] \qquad (4.36)$$

where G is the pseudo-inverse of the cdf of $X_i \sim P$, C is the copula of $(U_1, \cdots, U_n)$ and $\mathfrak{C}_n$ is the set of n-copulas. Our result solves (4.36) for monotone distributions P. By Theorem 4.5.5, we have

$$\begin{aligned}\min_{X_1,\cdots,X_n \sim P} \mathbb{E}[(X_1 + \cdots + X_n - t)_+] &= \mathbb{E}^{Q_n^P}[(G(U_1) + \cdots + G(U_n) - t)_+] \\ &= n \int_0^{c_n} [H(u) - t]_+ du + (1 - nc_n)[H(c_n) - t]_+.\end{aligned}$$

We provide a numerical result to compare the stop-loss premium $\mathbb{E}[(X_1+X_2+X_3-t)_+]$ for 4 different cases when $n = 3$. Suppose P is the exponential distribution with parameter 1 and $X_1, X_2, X_3 \sim P$.

- Case 1. X_1, X_2 and X_3 are comonotonic (see Denneberg [25]), i.e. $X_1 = X_2 = X_3$ almost surely. This case gives the maximum stop-loss premium.
- Case 2. X_1, X_2 and X_3 are independent.
- Case 3. X_1, X_2 and X_3 are negatively correlated with copula $C^{(1,2,3)}$ in Yang, Qi and Wang. [112] (i.e. the corresponding uniform random variables U_1, U_2 and U_3 in (4.36) satisfy $U_1 = 1 - U_3$ and U_2 is independent of U_1 and U_3).
- Case 4. X_1, X_2 and X_3 have copula Q_3^P. This case gives the minimum stop-loss premium.

The result is given in Figure 4.2. From the figure, we can see that the minimum stop-loss premium and the stop-loss premium for independent risks have a significant difference, especially for large values of t.

4.6 Bounds on the distribution of the total risk

In this section, we study the bounds on the distributions of S in a homogenous Fréchet class using the idea of CM distributions. Here we use the same notations from Section 4.5.

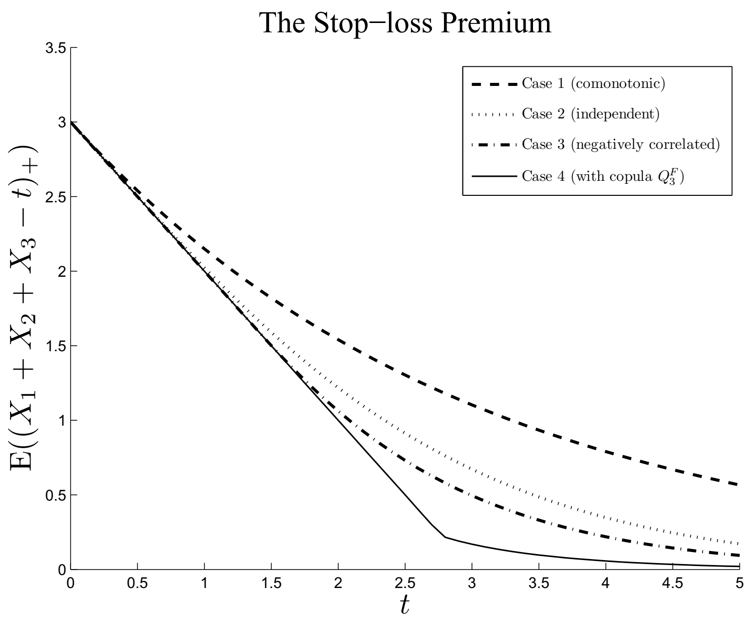

Figure 4.2: The stop-loss premium for different dependence structures

Let $\mathbf{X} = (X_1, \cdots, X_n)$ be a risk vector with known marginal distributions $F_1, \cdots, F_n$, denoted as $X_i \sim F_i, i = 1, \cdots, n$ and let $S = X_1 + \cdots + X_n$ be the total risk. For the purpose of risk management, it is of importance to find the best-possible bounds for the distribution of the total risk S when the dependence structure is unspecified:

Question B: the distribution of the total risk. Find bounds on the distribution of S:

$$m_+(s) = \inf_{\mathbf{X} \in \mathfrak{F}_n(F_1, \cdots, F_n)} \mathbb{P}(S < s); \tag{4.37}$$

$$M_+(s) = \sup_{\mathbf{X} \in \mathfrak{F}_n(F_1, \cdots, F_n)} \mathbb{P}(S < s). \tag{4.38}$$

See Embrechts and Puccetti [37] for discussions on such problems in risk management. Since techniques for handling $M_+(s)$ are very similar to those for $m_+(s)$, we shall focus on $m_+(s)$ in this section.

First let us review some known results on $m_+(s)$. Rüschendorf [85] found $m_+(s)$ when all marginal distributions have the same uniform or binomial distribution; Denuit, Genest and Marceau [26] and Embrechts, Höing and Juri [32] used copulas to yield the so-called *standard bounds*, which are no longer sharp for $n \geq 3$, and discussed some applications; Embrechts and Puccetti [35] provided a better lower bound (still not sharp) when all marginal distributions are the same and continuous, and some results when partial information on the dependence structure is available; Embrechts and Höing [31] provided a geometric interpretation to highlight the shape of the dependence structures with the worst VaR scenarios; Embrechts and Puccetti [36] extended this problem to multivariate marginal distributions and provided results similar to the univariate case. In summary, for $n \geq 3$, exact bounds were only found for the homogenous case ($F_1 = \cdots = F_n = F$) in Rüschendorf [85] where F is uniform or binomial. Besides the above results on $m_+(s)$, Rüschendorf [85] associated an equivalent dual optimization problem with the bounds for a general function of $X_1, \cdots, X_n$ instead of the total risk S.

The bounds $m_+(s)$ and $M_+(s)$ directly lead to the sharp bounds on quantile-based risk measures of S. A widely used measure is the so-called Value-at-Risk (VaR) at level α, defined as

$$\mathrm{VaR}_\alpha(S) = \inf\{s \in \mathbb{R} : \mathbb{P}(S \leq s) \geq \alpha\}.$$

The bound on the above VaR is called the worst Value-at-Risk scenario. Deriving sharp bounds for the worst VaR is of great interest in the recent research of quantitative risk management; see Embrechts and Puccetti [37] and Kaas, Laeven and Nelsen [55] for more details.

The section is organized as follows. We first provide a new lower bound on $m_+(s)$ in Section 4.6.1. When all the marginal distributions are identical and have a monotone or tail-monotone density, we employ the technique of Q_n^F introduced in Section 4.5 to find $m_+(s)$ in Section 4.6.2 and the worst Value-at-Risk for S in Section 4.6.3. Some examples are given in Section 4.6.4

4.6.1 General bounds

For any distribution F, we use $F^{-1}(t) = \inf\{s \in \mathbb{R} : F(s) \geq t\}$ to denote the (generalized) inverse function and we denote by $\tilde{F}_a$ the conditional distribution of F on $[F^{-1}(a), \infty)$ for $a \in [0, 1)$, i.e., $\tilde{F}_a(x) = \max\left\{\frac{F(x)-a}{1-a}, 0\right\}$ for $x \in \mathbb{R}$. It is straightforward to check that for $u \in [0, 1]$, $\tilde{F}_a^{-1}(u) = F^{-1}((1-a)u + a)$. In addition, let $\tilde{F}_1 = \lim_{a \to 1-} \tilde{F}_a$.

In the next we will give a general lower bound on $m_+(s)$. Before showing this bound, we need some definitions and lemmas.

Definition 4.6.1. The random vector $\mathbf{X} = (X_1, \cdots, X_n)$ with marginal distributions $F_1, \cdots, F_n$ is called an *optimal coupling* for $m_+(s)$ if

$$\mathbb{P}(X_1 + \cdots + X_n < s) = m_+(s).$$

It is known that the optimal coupling for $m_+(s)$ always exists (see the introduction in Rüschendorf [86] for instance). The following lemma is Proposition 3(c) of Rüschendorf [85], which will be used later.

Lemma 4.6.1. *Suppose $F_1, \cdots, F_n$ are continuous. Then there exits an optimal coupling $\mathbf{X} = (X_1, \cdots, X_n)$ for $m_+(s)$ such that $\{S \geq s\} = \{X_i \geq F_i^{-1}(m_+(s))\}$ for each $i = 1, \cdots, n$.*

Before presenting the main results on the relationship between the bounds on $m_+(s)$ and the jointly mixable distributions, we define the conditional moment function $\Phi(t)$ which plays an important role in the problem of finding $m_+(s)$. Suppose $X_i \sim F_i$ for $i = 1, \cdots, n$. Define

$$\Phi(t) = \sum_{i=1}^{n} \mathbb{E}(X_i | X_i \geq F_i^{-1}(t))$$

for $t \in (0, 1)$, and let

$$\Phi(1) = \lim_{t \to 1-} \Phi(t), \quad \Phi(0) = \lim_{t \to 0+} \Phi(t).$$

Obviously $\Phi(t)$ is increasing and continuous when $F_i, i = 1, \cdots, n$ are continuous. Let

$$\Phi^{-1}(x) = \inf\{t \in [0,1] \ : \ \Phi(t) \geq x\}$$

for $x \leq \Phi(1)$ and $\Phi^{-1}(x) = 1$ for $x > \Phi(1)$.

Theorem 4.6.2. *Suppose the distributions $F_1, \cdots, F_n$ are continuous.*

(1) We have

$$m_+(s) \geq \Phi^{-1}(s); \tag{4.39}$$

(2) For each fixed $s \geq \Phi(0)$, the equality

$$m_+(s) = \Phi^{-1}(s) \tag{4.40}$$

holds if and only if the conditional distributions $\tilde{F}_{1,a}, \cdots, \tilde{F}_{n,a}$ are jointly mixable, where $a = \Phi^{-1}(s)$.

Proof.

(1) It is trivial to prove the result when $\Phi(0) = \infty$. So we assume $\Phi(0) < \infty$. Note that from Lemma 4.6.1 we know that there exists an optimal coupling $\mathbf{X} = (X_1, \ldots, X_n)$ for $m_+(s)$ such that $\{S \geq s\} = \{X_i \geq F_i^{-1}(m_+(s))\}$ for each $i = 1, \cdots, n$. Hence

$$s \leq \mathbb{E}[S|S \geq s] = \sum_{i=1}^{n} \mathbb{E}[X_i|X_i \geq F_i^{-1}(m_+(s))] = \Phi(m_+(s)),$$

which implies (4.39).

(2) Suppose $\mathbf{X} = (X_1, \cdots, X_n)$ is an optimal coupling for $m_+(s)$ such that $\{S \geq s\} = \{X_i \geq F_i^{-1}(m_+(s))\}$ for each i. When $m_+(s) = \Phi^{-1}(s)$, it follows from the proof of part (1) that $\mathbb{E}(S|S \geq s) = s$, which implies that the conditional distributions of $X_1, \cdots, X_n$ on the set $\{S \geq s\}$ are JM, i.e., the conditional distributions $\tilde{F}_{1,a}, \cdots, \tilde{F}_{n,a}$ are JM.

Conversely, assume that $\tilde{F}_{1,a}, \cdots, \tilde{F}_{n,a}$ are JM. Then there exist $Y_1 \sim \tilde{F}_{1,a}, \cdots, Y_n \sim \tilde{F}_{n,a}$ such that

$$Y_1 + \cdots + Y_n = \mathbb{E}(Y_1 + \cdots + Y_n) = \Phi(a) \geq s.$$

Let

$$X_i = F_i^{-1}(U) I_{\{U \leq a\}} + Y_i I_{\{U > a\}}, \tag{4.41}$$

where $U \sim \mathrm{U}[0,1]$ and is independent of $(Y_1, \cdots, Y_n)$. Then it is easy to verify that X_i has the distribution function F_i for $i = 1, \cdots, n$ and

$$m_+(s) \leq \mathbb{P}(S < s) \leq a = \Phi^{-1}(s).$$

The other inequality $m_+(s) \geq \Phi^{-1}(s)$ is shown in part (1).

□

In the next we apply Theorem 4.6.2 to the homogenous case, i.e. $F_1 = \cdots = F_n \equiv F$. For $X \sim F$, define

$$\psi(t) = \mathbb{E}(X | X \geq F^{-1}(t))$$

for $t \in (0, 1)$,

$$\psi(1) = \lim_{t \to 1-} \psi(t), \quad \psi(0) = \lim_{t \to 0+} \psi(t),$$

$$\psi^{-1}(x) = \inf\{t \in [0,1] \ : \ \psi(t) \geq x\}$$

for $x \leq \psi(1)$ and $\psi^{-1}(x) = 1$ for $x > \psi(1)$. The following corollary follows from Theorem 4.6.2 immediately.

Corollary 4.6.3. *Suppose $F_1 = \cdots = F_n \equiv F$ and F is continuous.*

(1) We have

$$m_+(s) \geq \psi^{-1}(s/n). \tag{4.42}$$

(2) For each fixed $s \geq n\psi(0)$, the equality

$$m_+(s) = \psi^{-1}(s/n) \tag{4.43}$$

holds if and only if the conditional distribution function $\tilde{F}_a$ is n-completely mixable, where $a = \psi^{-1}(s/n)$.

Embrechts and Puccetti [35] also gave a lower bound for $m_+(s)$ in the homogeneous case. Different from the bound in [35], Theorem 4.6.2 deals with a more general case, where the random variables $X_1, \cdots, X_n$ do not need to be identically distributed and positive. Moreover, the bound in Theorem 4.6.2 is easier to calculate. Note that infinite support generally implies that the mixable condition in Theorem 4.6.2 and Corollary 4.6.3 does not hold.

4.6.2 Homogenous case with monotone marginal densities

In this section, we investigate the homogenous case when $F_1 = \cdots = F_n = F$ and F has either a monotone density or a tail-monotone density on its support. Since the case of $n = 1$ is trivial, we assume $n \geq 2$.

When the support of the distribution F is unbounded, the mixable condition in Theorem 4.6.2 and Corollary 4.6.3 is not satisfied by Proposition 4.2.1(6), i.e., the bound $\psi^{-1}(s/n)$ is not sharp. In this section, we find a formula for calculating the bound $m_+(s)$ for any distribution with a monotone density or a tail-monotone density, and obtain the corresponding correlation structure. This partially answers the question of optimal coupling for $m_+(s)$, which has remained open for decades. As a direct application, the bounds on $\mathrm{VaR}_\alpha(S)$ are obtained as well.

To calculate $m_+(s)$ for F having a monotone marginal density, we will use the copula Q_n^F ($n \geq 2$) in Section 4.5. More specifically, for some $0 \leq c \leq 1/n$ and random vector $(U_1, \cdots, U_n)$ with uniform marginal distributions on [0,1], we say $(U_1, \cdots, U_n) \sim Q_n^F(c)$ if

(a) For each $i = 1, \cdots, n$, given $U_i \in [0, c]$, we have $U_j = 1 - (n-1)U_i, \ \forall j \neq i$.

(b) $F^{-1}(U_1) + \cdots + F^{-1}(U_n)$ is a constant when any one of $U_i' s$ lies in $(c, 1-(n-1)c)$.

Denote $Q_n^F = Q_n^F(c_n)$ where c_n is the smallest possible c such that $Q_n^F(c)$ exists. Note that $c_n = 0$ if and only if F is n-CM. Define

$$H(x) = F^{-1}(x) + (n-1)F^{-1}(1-(n-1)x) \quad \text{for } F \text{ with a non-decreasing density.} \tag{4.44}$$

From Section 4.5, the smallest possible c for F with an increasing density is

$$c_n = \min\{c \in [0, \frac{1}{n}] : \int_c^{\frac{1}{n}} H(t)dt \leq (\frac{1}{n} - c)H(c)\} \tag{4.45}$$

and for any convex function f,

$$\min_{X_1, \cdots, X_n \sim F} \mathbb{E} f(X_1 + \cdots + X_n) = \mathbb{E}^{Q_n^F} f\left(F^{-1}(U_1) + \cdots + F^{-1}(U_n)\right). \tag{4.46}$$

For F with a decreasing density ($n \geq 2$), we define $Q_n^F(c)$ similarly as follows. For some $0 \leq c \leq 1/n$, we say $(U_1, \cdots, U_n) \sim Q_n^F(c)$ if

(a') For each $i = 1, \cdots, n$, given $U_i \in [1-c, 1]$, we have $U_j = (n-1)(1-U_i), \ \forall j \neq i$.

(b') $F^{-1}(U_1) + \cdots + F^{-1}(U_n)$ is a constant when any one of U_i lies in $((n-1)c, 1-c)$.

Define

$$H(x) = (n-1)F^{-1}((n-1)x) + F^{-1}(1-x) \quad \text{for } F \text{ with a decreasing density.} \tag{4.47}$$

As for the distribution of Z with a decreasing density, the distribution of $-Z$ has an increasing density, thus the above properties hold for F with a decreasing density. That is, the smallest possible c for F with a decreasing density is

$$c_n = \min\{c \in [0, \frac{1}{n}] : \int_c^{\frac{1}{n}} H(t)dt \geq (\frac{1}{n} - c)H(c)\}. \tag{4.48}$$

And for a distribution F with a decreasing density and any convex function f the equation (4.46) holds.

Although

$$m_+(s) = \min_{X_1,\cdots,X_n\sim F} \mathbb{E}(I_{\{S<s\}}),$$

the above results can not be applied directly to solve $m_+(s)$ since the indicator function $I_{(-\infty,s)}(\cdot)$ is not a concave function. Here we propose to find $m_+(s)$ for F with a monotone marginal density based on the following properties of Q_n^F.

Proposition 4.6.4. *Suppose F admits a monotone density on its support.*

1. *If $(U_1,\cdots,U_n)\sim Q_n^F(c)$ and F has an increasing density, then $I_{\{U_i\in(c,1-(n-1)c)\}} = I_{\{U_1\in(c,1-(n-1)c)\}}$ a.s. for $i=1,\cdots,n$.*

2. *If $X_1,\cdots,X_n\sim F$ with copula Q_n^F, then*

$$S = X_1+\cdots+X_n = \begin{cases} H(U/n)I_{\{U\le nc_n\}} + H(c_n)I_{\{U>nc_n\}}, & c_n>0; \\ n\mathbb{E}(X_1), & c_n=0 \end{cases} \tag{4.49}$$

for some $U\sim \mathrm{U}[0,1]$.

The proof of Proposition 4.6.4 is given in the appendix.

Now we are ready to give a computable formula for $m_+(s)$. In the following we define a function $\phi(x)$ which works similarly as $\Phi(x)$ in the CM case.

For F with a decreasing density and $a\in[0,1]$, define

$$H_a(x) = (n-1)F^{-1}(a+(n-1)x) + F^{-1}(1-x) \tag{4.50}$$

for $x\in[0,\frac{1-a}{n}]$ and

$$c_n(a) = \min\{c\in[0,\frac{1}{n}(1-a)] : \int_c^{\frac{1}{n}(1-a)} H_a(t)dt \ge (\frac{1}{n}(1-a)-c)H_a(c)\}. \tag{4.51}$$

Write

$$\phi(a) = \begin{cases} H_a(c_n(a)) & \text{if } c_n(a)>0, \\ n\psi(a) & \text{if } c_n(a)=0. \end{cases} \tag{4.52}$$

On the other hand, for F with an increasing density and $a \in [0,1]$, define

$$H_a(x) = F^{-1}(a+x) + (n-1)F^{-1}(1-(n-1)x), \tag{4.53}$$

$$c_n(a) = \min\{c \in [0, \frac{1}{n}(1-a)] : \int_c^{\frac{1}{n}(1-a)} H_a(t)dt \leq (\frac{1}{n}(1-a)-c)H_a(c)\} \tag{4.54}$$

and

$$\phi(a) = \begin{cases} H_a(0) & \text{if } c_n(a) > 0, \\ n\psi(a) & \text{if } c_n(a) = 0. \end{cases} \tag{4.55}$$

Some probabilistic interpretation of the functions $H_a(x)$ and $\phi(a)$ is given in the following remark. Technical details are put in Lemma 4.6.5 later.

Remark 4.6.1. Suppose $Y_1, \cdots, Y_n \sim \tilde{F}_a$ with copula $Q_n^{\tilde{F}_a}$. By (4.49) we have

$$Y_1 + \cdots + Y_n = \begin{cases} \tilde{H}(U/n)I_{\{U \leq n\tilde{c}_n\}} + \tilde{H}(\tilde{c}_n)I_{\{U > n\tilde{c}_n\}}, & \tilde{c}_n > 0, \\ n\mathbb{E}(Y_1), & \tilde{c}_n = 0 \end{cases}$$

for some $U \sim \mathrm{U}[0,1]$, where $\tilde{H}(x)$ and $\tilde{c}_n$ are $H(x)$ and c_n defined in (4.44), (4.45), (4.47) and (4.48) by replacing F with $\tilde{F}_a$. It is easy to check that $\tilde{H}(x) = H_a((1-a)x)$, $\tilde{c}_n = c_n(a)/(1-a)$ and $\tilde{H}(\tilde{c}_n) = H_a(c_n(a))$. For $c_n(a) > 0$, later we will show that $H_a(x)$, $x \in [0, c_n(a)]$ attains its minimum value at $H_a(c_n(a))$ for $\tilde{F}_a$ with a decreasing density and at $H_a(0)$ for $\tilde{F}_a$ with an increasing density. Therefore, the minimum possible value of $Y_1 + \cdots + Y_n$ is

$$\min_{x \in [0, c_n(a)]} H_a(x)I_{\{c_n(a)>0\}} + n\mathbb{E}(Y_1)I_{\{c_n(a)=0\}} = \phi(a).$$

Thus, $\mathbb{P}(Y_1 + \cdots + Y_n \geq \phi(a)) = 1$, which leads to $\mathbb{P}(S < \phi(a)) \leq a$ by setting $X_i = F^{-1}(V)I_{\{V \leq a\}} + Y_i I_{\{V > a\}}$ where $V \sim \mathrm{U}[0,1]$ is independent of $Y_1, \cdots, Y_n$. This suggests $m_+(s) \leq \phi^{-1}(a)$, i.e., $\phi^{-1}(a)$ is potentially an optimal bound. In order to prove the optimality of $\phi^{-1}(a)$, more details of the functions $H_a(x)$ and $\phi(a)$ are given in the following lemma, whose proof is put in the appendix.

Lemma 4.6.5. *Suppose F admits a monotone density.*

(i) If F has a decreasing density, then given $a \in [0,1)$, $H_a(x)$ is decreasing and differentiable for $x \in [0, c_n(a)]$.

(ii) If F has an increasing density, then given $a \in [0,1)$, $H_a(x)$ is increasing and differentiable for $x \in [0, c_n(a)]$.

(iii) If F has a decreasing density, then $\phi(a) = n\mathbb{E}[F^{-1}(V_a)]$ where $V_a \sim \mathrm{U}[a + (n-1)c_n(a), 1 - c_n(a)]$.

(iv) For any random variables $U_1, \cdots, U_n \sim \mathrm{U}[a,1]$ and $0 \le a < b \le 1$, we have $\mathbb{E}(F^{-1}(U_i)|A) < \mathbb{E}[F^{-1}(V_b)]$ for $i = 1, \cdots, n$, where V_b is defined in (iii) and $A = \bigcap_{i=1}^{n}\{U_i \in [a, 1 - c_n(b)]\}$.

(v) Suppose $Y_1, \cdots, Y_n \sim \tilde{F}_a$ with copula $Q_n^{\tilde{F}_a}$, then $\mathbb{P}(Y_1 + \cdots + Y_n \ge \phi(a)) = 1$.

(vi) $\phi(a)$ is continuous and strictly increasing for $a \in [0,1)$.

Since $\phi(a)$ is continuous and strictly increasing, its inverse function $\phi^{-1}(a)$ exists. Put $\phi^{-1}(t) = 0$ if $t < \phi(0)$ and $\phi^{-1}(t) = 1$ if $t > \phi(1)$.

Theorem 4.6.6. *Suppose the distribution $F(x)$ has a decreasing density on its support and $\phi(a)$ is defined in (4.52), or the distribution $F(x)$ has an increasing density on its support and $\phi(a)$ is defined in (4.55). Then we have $m_+(s) = \phi^{-1}(s)$.*

Proof.

(a) We first prove $m_+(s) \le \phi^{-1}(s)$. Write $a = \phi^{-1}(s)$. For $i = 1, \cdots, n$, let $Y_1, \cdots, Y_n \sim \tilde{F}_a$ with copula $Q_n^{\tilde{F}_a}$ and $X_i = F^{-1}(V)I_{\{V \le a\}} + Y_i I_{\{V > a\}}$ where $V \sim \mathrm{U}[0,1]$ is independent of $Y_1, \cdots, Y_n$. It is easy to check that $X_i \sim F$ and by Lemma 4.6.5(v),

$$m_+(s) \le \mathbb{P}(S < \phi(a)) = 1 - \mathbb{P}(S \ge \phi(a)) \le 1 - \mathbb{P}(Y_1 + \cdots + Y_n \ge \phi(a))\mathbb{P}(V > a) = a.$$

Thus $m_+(s) \le \phi^{-1}(s)$.

(b) Next we prove $m_+(s) \geq \phi^{-1}(s)$ for the case when $F(x)$ has a decreasing density.

Suppose $a = m_+(s) < \phi^{-1}(s) = b$ and $\mathbf{X} = (X_1, \cdots, X_n)$ is an optimal coupling for $m_+(s)$ such that $\{S \geq s\} = \{X_i \geq F^{-1}(a)\}$ for each i. Hence there exist $U_{a,1}, \cdots, U_{a,n} \sim \mathrm{U}[a, 1]$ such that $F^{-1}(U_{a,1}) + \cdots + F^{-1}(U_{a,n}) \geq s$ with probability 1. By Lemma 4.6.5(iii) and (iv), we have

$$s \leq \mathbb{E}[\sum_{i=1}^{n} F^{-1}(U_{a,i})|A] < n\mathbb{E}(F^{-1}(V_b)) = \phi(b) = s.$$

This leads to a contradiction. Thus $m_+(s) = \phi^{-1}(s)$.

(c) Finally we prove $m_+(s) \geq \phi^{-1}(s)$ for the case when $F(x)$ has an increasing density. In this case $F^{-1}(1) < \infty$.

Write $a = m_+(s)$ and let $\mathbf{X} = (X_1, \cdots, X_n)$ be an optimal coupling for $m_+(s)$ such that $\{S \geq s\} = \{X_i \geq F^{-1}(a)\}$ for each i. It is clear that

$$\begin{aligned}&\mathbb{P}(S < F^{-1}(a) + (n-1)F^{-1}(1) + \epsilon | S \geq s) \\ &\geq \mathbb{P}(X_i < F^{-1}(a) + \epsilon | X_i \geq F^{-1}(a)) > 0\end{aligned}$$

for any $\epsilon > 0$. Note that $\mathbb{P}(S < s | S \geq s) = 0$ and thus

$$s \leq F^{-1}(a) + (n-1)F^{-1}(1) = H_a(0).$$

This shows $s \leq H_a(0)$. The inequality $s \leq n\psi(a)$ is given by Theorem 4.6.2. Hence $s \leq \phi(a)$ and $a \geq \phi^{-1}(s)$. □

The proof of the above theorem suggests to construct the optimal correlation structure as follows. In both cases, for $a = \phi^{-1}(s)$ let $U_{a,1}, \cdots, U_{a,n} \sim \mathrm{U}[a, 1]$ with copula $Q_n^{\tilde{F}_a}$ and $U \sim$U[0,1] is independent of $(U_{a,1}, \cdots, U_{a,n})$. Define

$$U_i = U_{a,i} I_{\{U \geq a\}} + U I_{\{U < a\}} \tag{4.56}$$

for $i = 1, \cdots, n$. Then

$$\mathbb{P}(F^{-1}(U_1) + \cdots + F^{-1}(U_n) < s) = \phi^{-1}(s).$$

Remark 4.6.2.

1. The copula Q_n^F plays an important role for the bounds on both the convex minimization problem (4.46) and the $m_+(s)$ problem for monotone marginal densities. Note that Q_n^F may not be unique, hence the structure (4.56) may not be unique. Also, on the set $\{S < s\}$, the dependence structure of $X_1, \cdots, X_n$ can be arbitrary.

2. The value $\phi^{-1}(s)$ is accurate even when $\mathbb{E}(\max\{X_1, 0\}) = \infty$. When the distribution $\tilde{F}_a$ is n-CM, Theorem 4.6.6 gives the sharp bound $\Phi^{-1}(s)$ in Theorem 4.6.2. The problem of $M_+(s)$ for monotone densities is also solved by the above theorem.

3. Figure 4.3 shows the sketch of an optimal coupling for F with a decreasing density, some $a > 0$ and $c_n(a) > 0$. Here $U_1, \cdots, U_n \sim \mathrm{U}[0,1]$ and $\mathbb{P}(F^{-1}(U_1) + \cdots + F^{-1}(U_n) < s) = \phi^{-1}(s)$.

   (i) When $U_i \in [0, a]$, U_i is arbitrarily coupled to all other U_j in Part A.

   (ii) When $U_i \in [a, a + (n-1)c_n(a)]$, U_i is coupled to other U_j, $j \neq i$ in Part B and Part D. For $j \neq i$, either $U_i - a = (n-1)(1 - U_j)$ or $U_j = U_i$.

   (iii) When $U_i \in [a + (n-1)c_n(a), 1 - c_n(a)]$, U_i is coupled to all other U_j, $j \neq i$ in Part C, and $F^{-1}(U_1) + \cdots + F^{-1}(U_n) = \phi(a)$. It is the completely mixable part.

   (iv) When $U_i \in [1 - c_n(a), 1]$, U_i is coupled to other U_j, $j \neq i$ in Part B. For $j \neq i$, $U_j - a = (n-1)(1 - U_i)$.

4. Figure 4.4 shows the real values of $m_+(s)$ in Theorem 4.6.6 and the lower bound $\psi^{-1}(s/n)$ in Theorem 4.6.2 for the Pareto(2,1) distribution. Note that the real values are equal to the bound in Embrechts and Puccetti [35], which suggests that the bound in [35] may be sharp for Pareto distributions.

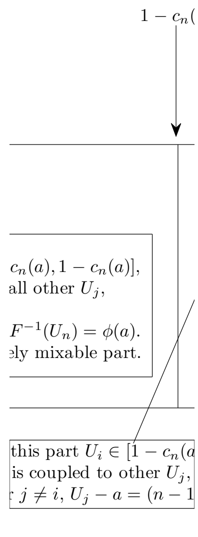

Figure 4.3: Sketch of the optimal coupling

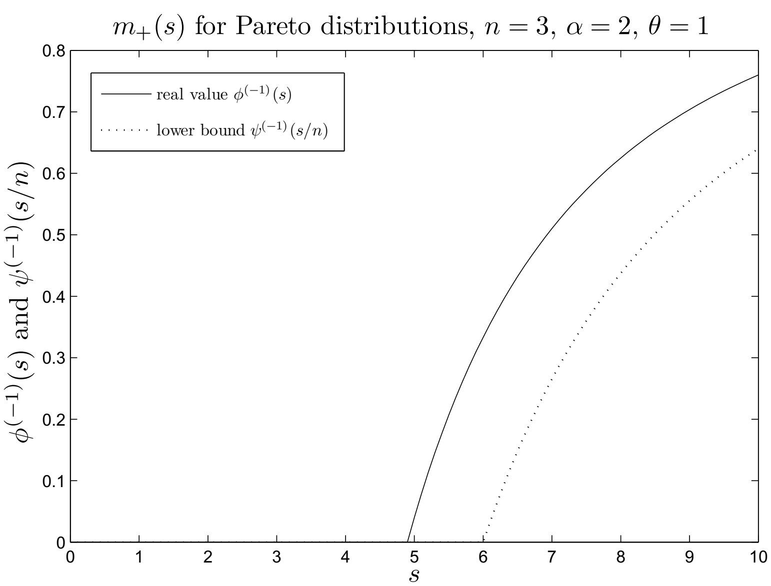

Figure 4.4: $m_+(s)$ and $\psi^{-1}(s/n)$ for a Pareto distribution

For the distribution F with density $p(x)$, we say $p(x)$ is tail-monotone, if for some $b \in \mathbb{R}$, $p(x)$ is decreasing for $x > b$ or $p(x)$ is increasing for $x < b$. We are particularly interested in the case when $p(x)$ is tail-decreasing ($p(x)$ is decreasing for $x > b$) since the risks are usually positive random variables. Note that for most risk distributions the tail-decreasing property is satisfied. For example, the Gamma distribution with shape parameter α for $\alpha > 1$ and the F-distribution with d_1, d_2 degrees of freedom for $d_1 > 2$ do not have a monotone density, but they have a tail-decreasing density.

In the VaR problems, one is concerned with the tail behavior of the distribution. From the proof of Theorem 4.6.6, information on the left tail of F does not play any role in the calculation of $m_+(s)$. Based on this observation, we have the following theorem, which solves $m_+(s)$ for F with tail-decreasing density and some large s.

Theorem 4.6.7. *Suppose the density function of F is decreasing on $[b, \infty)$, and $\phi(a)$ is defined in* (4.52)*. Then for $s \geq \phi(F(b))$, $m_+(s) = \phi^{-1}(s)$.*

Proof. Since the density function of F is decreasing on $[b, \infty)$, the conditional distribution $\tilde{F}_{F(b)}$ has a decreasing density. Note that $H_a(x)$, $c_n(a)$ and $\phi(a)$ only depend on the conditional distribution $\tilde{F}_a$, hence they are well defined for $F(b) \leq a \leq 1$.

Since $s \geq \phi(F(b))$, $\phi^{-1}(s) \geq F(b)$ and the conditional distribution $\tilde{F}_{\phi^{-1}(s)}$ has a decreasing density. Theorem 4.6.7 follows from the same arguments as in the proof of Theorem 4.6.6, where no condition on the distribution of X_i on $\{X_i < F^{-1}(\phi^{-1}(s))\}$ is used. □

4.6.3 The worst Value-at-Risk scenarios

The Value-at-Risk (VaR) is an important risk measure in risk management; see Embrechts and Puccetti [37] and references therein. Recall that VaR is the α-quantile of the distribution, i.e.,

$$\mathrm{VaR}_\alpha(S) = F_S^{-1}(\alpha) = \inf\{s \in \mathbb{R} : F_S(s) \geq \alpha\}, \tag{4.57}$$

where F_S is the distribution of S. Typical values of the level α are 0.95, 0.99 or even 0.999. As mentioned in Embrechts and Puccetti [37], banks are concerned with an upper bound on $\mathrm{VaR}(\sum_{i=1}^{d} X_i)$ when the correlation structure between $\mathbf{X} = (X_1, \cdots, X_d)$ is unspecified.

Finding the bounds on the VaR is equivalent to finding the inverse function of $m_+(s)$ (note that $m_+(s)$ is non-decreasing). Using Theorem 4.6.6 and Theorem 4.6.7, we are able to obtain the explicit value of the upper bound on the VaR, namely, the worst Value-at-Risk. The proof follows directly from the fact that $\sup_{X_i \sim F, 1 \le i \le n} \mathrm{VaR}_\alpha(S) = m_+^{-1}(\alpha)$ when $m_+(s)$ is continuous and strictly increasing.

Theorem 4.6.8. *Suppose that the density function of the marginal distribution F is decreasing on $[b, \infty)$ and $\phi(a)$ is defined in (4.52). Then for $\alpha \ge F(b)$, the worst VaR of $S = X_1 + \cdots + X_n$ is*

$$\sup_{X_i \sim F, 1 \le i \le n} \mathrm{VaR}_\alpha(S) = m_+^{-1}(\alpha) = \phi(\alpha). \tag{4.58}$$

In particular, (4.58) holds for all α if the marginal distribution F has decreasing density on its support and an optimal correlation structure is given by (4.56).

For arbitrary marginal distributions $F_1, \cdots, F_n$, Theorem 4.6.2 gives an upper bound for the worst-VaR problem as follows.

Corollary 4.6.9. *For arbitrary marginal distributions,*

$$\sup_{X_i \sim F_i, i=1,\cdots,n} \mathrm{VaR}_\alpha(S) \le m_+^{-1}(\alpha) \le \Phi(\alpha), \tag{4.59}$$

where $\Phi(\alpha)$ is defined in Section 2.

Figure 4.5 shows the explicit worst-VaR, best-VaR and VaR in the independent case for the distribution Pareto(3,2), $n = 3$ and $0.9 \le \alpha \le 0.995$.

4.6.4 Examples

Here we give some examples to show how to compute $m_+(s)$.

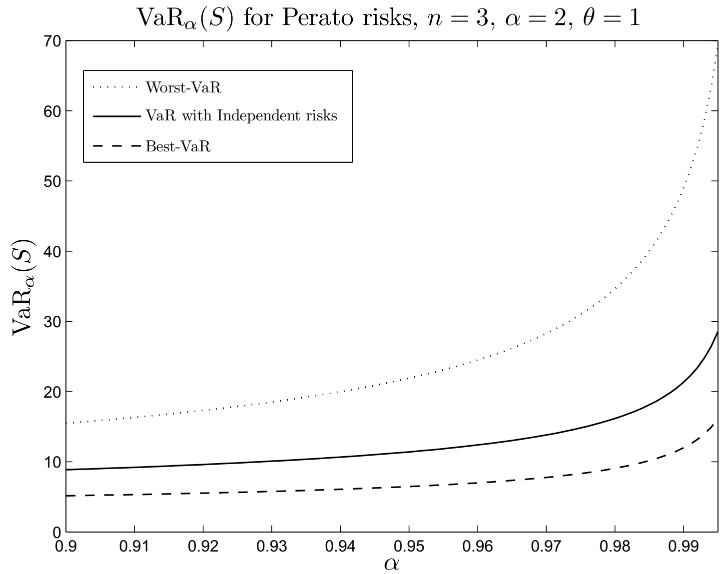

Figure 4.5: Worst and best VaR for a Pareto distribution

Example 4.6.1. Assume that $X \sim \mathrm{U}[0,1]$, the uniform distribution on [0,1]. Then

$$p(x) = 1,\ F(x) = x,\ x \in [0,1],\ F^{-1}(t) = t,\ t \in [0,1].$$

Further we have $c_n(a) = 0$ for all $0 \le a \le 1$ and $\phi(t) = n\psi(t) = n\mathbb{E}(X|X > t) = \frac{n(1+t)}{2}$ for $t \in [0,1]$. Thus

$$m_+(s) = \phi^{-1}(s) = 1 \wedge \left(\frac{2s}{n} - 1\right)_+ .$$

This result indeed is the same as that in Rüschendorf [85]. One optimal correlation structure is also given in Rüschendorf and Uckelmann [87].

Example 4.6.2. Assume that $X \sim \mathrm{Pareto}(\alpha, \theta), \alpha > 1, \theta > 0$ with density function

$$p(x) = \alpha\theta^\alpha x^{-\alpha-1},\ x \ge \theta.$$

Then

$$F(x) = 1 - \left(\frac{x}{\theta}\right)^{-\alpha},\ x \ge \theta,\ F^{-1}(t) = \theta(1-t)^{-1/\alpha},\ t \in [0,1].$$

Further we have that $c_n(a)$ is the smallest $c \in [0, \frac{1}{n}(1-a)]$ such that

$$\frac{\alpha}{\alpha-1}((1-a-(n-1)c)^{1-1/\alpha} - c^{1-1/\alpha}) \ge (\frac{1}{n}(1-a)-c)((n-1)(1-a-(n-1)c)^{-1/\alpha} + c^{-1/\alpha}).$$

The numerical values of $m_+(s)$ for two Pareto distributions and $n = 3$ are plotted in Figure 4.6. A possible correlation structure is given in (4.56).

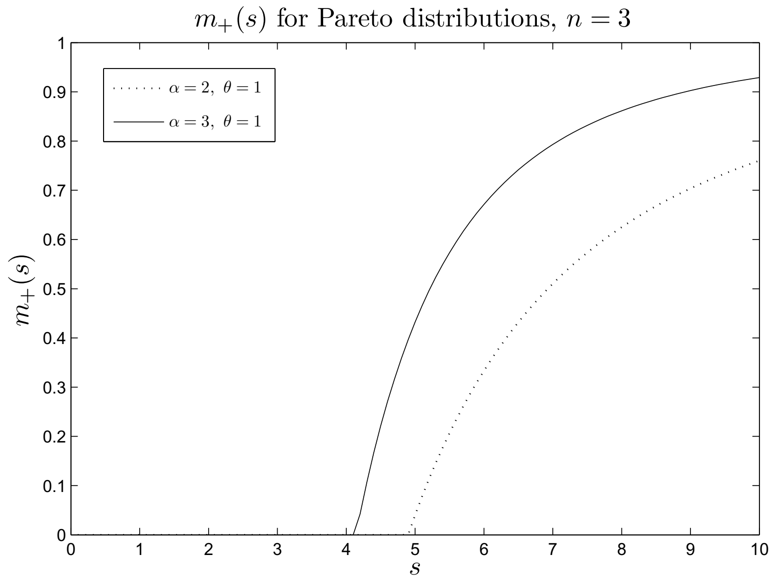

Figure 4.6: $m_+(s)$ for Pareto distributions

Example 4.6.3. Assume that $X \sim \text{Gamma}(\alpha, \lambda), \alpha \leq 1, \lambda > 0$ with density function

$$p(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}.$$

Then

$$F(x) = \gamma(\alpha, \lambda x),\ \ x > 0,$$

where $\gamma(\alpha, t) = \int_0^t \frac{1}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx$ is the lower incomplete Gamma function. Further $c_n(a)$ is the smallest $c \in [0, \frac{1}{n}(1-a)]$ such that

$$\frac{\alpha}{\lambda}(\gamma(\alpha+1, \lambda F^{-1}(1-c)) - \gamma(\alpha+1, \lambda F^{-1}(a+(n-1)c))) \geq (\frac{1}{n}(1-a) - c)H_a(c),$$

which can be calculated numerically. The numerical values of $m_+(s)$ for two Gamma distributions and $n = 3$ are plotted in Figure 4.7. A possible correlation structure is given in (4.56).

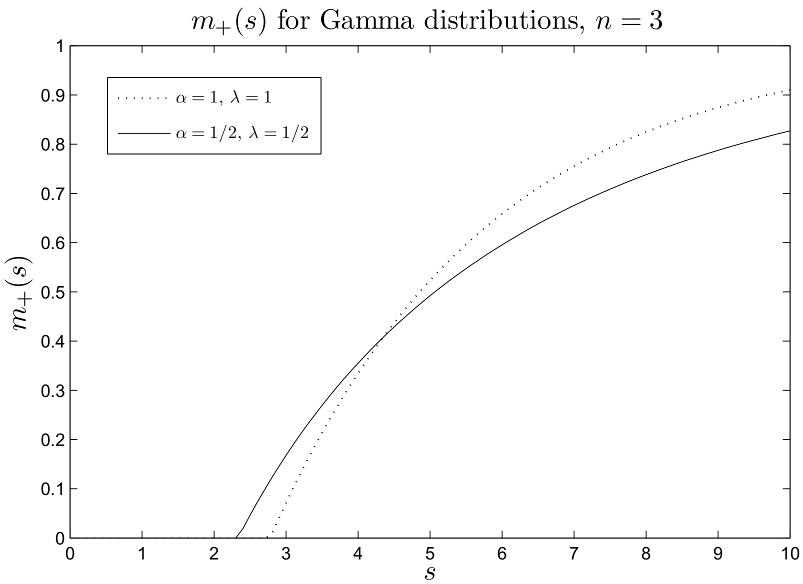

Figure 4.7: $m_+(s)$ for Gamma distributions

4.7 Technical Proofs

Proof of Proposition 4.2.2. (i) and (ii) are obvious. For (iii), let $S = (\underbrace{0, \cdots, 0}_{q-p}, \underbrace{1, \cdots, 1}_{p})$, σ be a random permutation uniformly distributed on the set of all q-permutations, and the random vector $\mathbf{X} = (X_1, \cdots, X_q) = \sigma(S)$. We can check that $X_i \sim B(1, r)$ for $i = 1, \cdots, q$ and $X_1 + \cdots + X_q = p$ is a constant. Hence $B(1, r)$ is q-CM. The rest part of (iii) follows from Proposition 4.2.1(5). (iv) is an application of Theorem 4.4.1. The uniform distribution in (v) and the Beta distribution in (vi) have monotone densities, hence (v) and (iv) follow from Theorem 4.4.2. The Beta distribution in (vii) and the triangular distribution in (viii) have concave densities, then (vii) and (viii) follow from Theorem 4.4.8. □

Proof of Lemma 4.4.4. (4.12) reads as

$$N \times A(-N) + \cdots + 1 \times A(-1) = 1 \times A(1) + \cdots + dN \times A(dN). \tag{4.60}$$

The left-hand side of (4.60) is

$$\begin{aligned} N \times A(-N) + \cdots + 1 \times A(-1) &\le N \times A(-N) + \frac{(N-1)N}{2} \times A(-N+1) \\ &\le \left(N + \frac{N(N-1)}{2}\frac{2d}{d+1}\right) A(-N) \\ &= \frac{N(dN+1)}{d+1} \times A(-N). \end{aligned}$$

The right-hand side of (4.60) is

$$\begin{aligned} & 1 \times A(1) + \cdots + dN \times A(dN) \\ \ge\ & \frac{(dN-d+1)(dN-d+2)}{2} \times A(dN-d+1) \\ & +(dN-d+2) \times A(dN-d+2) + \cdots + dN \times A(dN) \qquad (4.61) \\ \ge\ & \frac{N(dN+1)}{d+1} \times (1 \times A(dN-d+1) + 2 \times A(dN-d+2) + \cdots + d \times A(dN)) \qquad (4.62) \end{aligned}$$

The last inequality is due to the fact that $A(dN-d+1) \ge \cdots \ge A(dN)$, the summation of all coefficients in (4.61) equals that in (4.62) and for each i and the summation of all coefficients from term $A(dN-d+1)$ to $A(dN-d+i)$ in (4.60) is greater than that in (4.62). Therefore we get

$$1 \times A(dN-d+1) + 2 \times A(dN-d+2) + \cdots + d \times A(dN) \le A(-N),$$

and thus $C_N(A) \ge 0$. □

Proof of $C_{N-1}(\bar{A}) \ge 0$. Note that $\bar{A}(-N+1) = A(-N+1) - \sum_{i=1}^{d-1} iA(dN-i)$. Comparing the left-hand side and right-hand side of (4.60), we get

$$\begin{aligned} & N \times A(-N) + \frac{N(N-1)}{2} \times A(-N+1) \\ \ge\ & \text{LHS of (4.60)} \\ =\ & \text{RHS of (4.60)} \\ \ge\ & \frac{(dN-d+1)(dN-d+2)}{2} \times A(dN-d+1) + \sum_{i=2}^{d} (dN-d+i) \times A(dN-d+i). \end{aligned}$$

Plugging $C_N(A) = 0$ in and after simplification (here we divide both sides by $N-1$, hence $N \geq 2$ is needed), the above inequality reads as

$$\begin{aligned} N \times A(-N+1) &\geq 2 \times A(dN-1) + \cdots + 2(d-2) \times A(dN-d+2) \\ &\quad + \frac{(d^2N - d^2 + 3d - 2)(N-1)}{2} \times A(dN-d+1). \end{aligned}$$

Since $A(dN-1) \leq A(dN-2) \leq \cdots \leq A(dN-d+1)$, we can conclude

$$A(-N+1) \geq \frac{2d}{d-1}[1 \times A(dN-1) + \cdots + (d-1) \times A(dN-d+1)].$$

This leads to

$$\bar{A}(-N+1) \geq A(-N+1) - \frac{d-1}{2d}A(-N+1) \geq \frac{d+1}{2d}A(-N+2) = \frac{d+1}{2d}\bar{A}(-N+2).$$

By **Lemma 4.4.4** we know $(\bar{A}, N-1)$ satisfies (ii). □

Proof of Corollary 4.5.9. In the following we let P_n be the unique solution to

$$\log P = \frac{nP - n}{n + P - 1}, \; P > 1. \tag{4.63}$$

One can show (4.63) has unique solution other than $P = 1$ by the following argument. Let $f(x) = \log x - n + \frac{n^2}{n+x-1}$. Then $f'(x) = \frac{1}{x} - \frac{n^2}{(n+x-1)^2}$, hence $f'(x)$ only has one root other than $x = 1$. This shows $f(x) = 0$ has at most one root other than $x = 1$. Note that $f(2) < 0$ and $f(e^n) > 0$, thus it has unique root other than $x = 1$.

Let $c_n = \frac{1}{P_n+n-1}$ ($P_n = \frac{1-(n-1)c_n}{c_n}$) and plug it in (4.63), we get c_n is the unique solution to (4.35).

For any $0 < \eta < 1$,

$$f(\eta e^n) = \log \eta + \frac{n^2}{n + \eta e^n - 1} < 0$$

for large n, hence there is a solution to $f(x) = 0$ between ηe^n and e^n. Since P_n is the solution, we know $P_n \sim e^n$, therefore $c_n = \frac{1}{P_n+n-1} \sim e^{-n}$.

Furthermore, it follows from $\log(P_n/e^n) = -n^2/(n+P_n-1)$ and $P_n \sim e^n$ that

$$\begin{aligned}
P_n/e^n &= 1 - \frac{n^2}{P_n+n-1} + \frac{n^4}{2(P_n+n-1)^2} + O(\frac{n^6}{(P_n+n-1)^3}) \\
&= 1 - \frac{n^2}{e^n} + \frac{n^2(P_n+n-1-e^n)}{e^n(P_n+n-1)} + \frac{n^4}{2(P_n+n-1)^2} + O(\frac{n^6}{e^{3n}}) \\
&= 1 - \frac{n^2}{e^n} + \frac{n^2(-n^2+n-1)}{e^{2n}} + O(\frac{n^6}{e^{3n}}) + \frac{n^4}{2e^{2n}} + O(\frac{n^6}{e^{3n}}) + O(\frac{n^6}{e^{3n}}) \\
&= 1 - n^2 e^{-n} + \frac{-n^4+2n^3-2n^2}{2} e^{-2n} + O(n^6 e^{-3n}).
\end{aligned}$$

Consequently

$$\begin{aligned}
c_n &= e^{-n} + (\frac{1}{P_n+n-1} - e^{-n}) \\
&= e^{-n} + \frac{e^n - (P_n+n-1)}{e^n(P_n+n-1)} \\
&= e^{-n} + (n^2-n+1)e^{-2n} + O(n^4 e^{-3n}),
\end{aligned}$$

and

$$\begin{aligned}
\Lambda_n =& n\int_0^{c_n} x(1-(n-1)x)^{n-1}dx + (1-nc_n)c_n(1-(n-1)c_n)^{n-1} \\
=& n\int_0^{c_n} x(1-(n-1)x)^{n-1}dx + c_n[1-((n-1)^2+n)c_n + O(n^3c_n^2)] \\
=& n\int_0^{c_n} x(1-(n-1)x)^{n-1}dx + c_n - (n^2-n+1)c_n^2 + O(n^3c_n^3). \\
=& \frac{n}{2}c_n^2 + O(n^3c_n^3) + c_n - (n^2-n+1)c_n^2 + O(n^3c_n^3). \\
=& e^{-n} + \frac{n}{2}e^{-2n} + O(n^4e^{-3n}).
\end{aligned}$$

□

Proof of Proposition 4.2.3.

1. The case $n=1$ is trivial. For $n \geq 2$, by the definition of JM distributions, there exist $X_1 \sim F_1, \cdots, X_n \sim F_n$ such that $\mathrm{Var}(X_1+\cdots+X_n) = 0$. Since

$$\begin{aligned}
\sqrt{\mathrm{Var}(X_1+X_2+\cdots+X_n)} &\geq \sqrt{\mathrm{Var}(X_1)} - \sqrt{\mathrm{Var}(X_2+\cdots+X_n)} \\
&\geq \sigma_1 - \sum_{i=2}^n \sigma_i,
\end{aligned}$$

we have $2\sigma_1 - \sum_{i=1}^n \sigma_i \le 0$. Similarly, we can show that $2\sigma_k - \sum_{i=1}^n \sigma_i \le 0$ for any $k = 1, \cdots, n$, i.e., (4.7) holds.

2. We only need to prove the "$\Leftarrow$" part for $n \ge 2$. Without loss of generality, we assume $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_n$. Let $\mathbf{X} = (X_1, \cdots, X_n)$ be a multivariate Gaussian random vector with known marginal distributions $F_1, \cdots, F_n$ and an unspecific correlation matrix Γ. We want to show there exists a correlation matrix Γ such that $\mathrm{Var}(X_1 + \cdots + X_n) = 0$.

   Let T be the correlation matrix of $(X_2, \cdots, X_n)$ and $Y = X_2 + \cdots + X_n$. Define $f(T) = \sqrt{\mathrm{Var}(X_1)} - \sqrt{\mathrm{Var}(Y)}$. Obviously $f(T)$ is a continuous function of T with canonical distance measure. It is easy to check that $f(T) = \sigma_1 - \sum_{i=2}^n \sigma_i \le 0$ when $X_2 = \sigma_2 Z + \mu_2, \cdots, X_n = \sigma_n Z + \mu_n$ for some $Z \sim N(0,1)$. Since $\sigma_1 \ge \sigma_2 \ge \cdots \ge \sigma_n$, we also have $f(T) = \sigma_1 - |\sum_{i=2}^n (-1)^i \sigma_i| \ge 0$ when $X_i = (-1)^i \sigma_i Z + \mu_i$ for $i = 2, \cdots, n$. Hence there exists a correlation matrix T_0 such that $f(T_0) = 0$. With the correlation matrix of $(X_2, \cdots, X_n)$ being T_0, we define $X_1 = -Y + \mathbb{E}(Y) + \mu_1$. Hence $X_1 \sim N(\mu_1, \sigma_1^2)$ and $\mathrm{Var}(X_1 + \cdots + X_n) = 0$, which imply that $F_1, \cdots, F_n$ are JM.

Proof of Proposition 4.6.4.

1. By (a) in Section 3.1, for any $i \ne j$, $U_i \in [0, c] \Rightarrow U_j \in [1-(n-1)c, 1]$. Hence

$$A_i := \{U_i \in [0, c]\} \subseteq \{U_j \in [1-(n-1)c, 1]\} =: B_j$$

   and $\mathbb{P}(A_i \cap A_j) = 0$. As a consequence, $\bigcup_{i \ne j} A_i \subseteq B_j$. Note that $\mathbb{P}(\bigcup_{i \ne j} A_i) = (n-1)c = \mathbb{P}(B_j)$. Thus $I_{\bigcup_{i \ne j} A_i} = I_{B_j}$ a.s. and

$$I_{\bigcup_{i=1}^n A_i} = I_{A_j \cup B_j} = I_{\{U_j \in [0,c] \cup [1-(n-1)c, 1]\}} \quad \text{a.s.}$$

   which imply that $I_{\{U_j \in (c, 1-(n-1)c)\}} = I_{(\bigcup_{i=1}^n A_i)^c}$ a.s. for $j = 1, \cdots, n$.

2. We only prove the case when F has an increasing density. When $c_n = 0$, (4.49) follows from the definition of Q_n^F. Next we assume $c_n > 0$. Write $D_j = A_j \cup B_j$ and $X_j = F^{-1}(U_j)$, $U_j \sim$ U[0,1] for $j = 1, \cdots, n$. First note that by condition (b) in Section 3.1, for any $j = 1, \cdots, n$, $F^{-1}(U_1) + \cdots + F^{-1}(U_n)$ is a constant on the set D_j^c. This constant equals its expectation, which is

$$
\begin{aligned}
& \mathbb{E}(F^{-1}(U_1) + \cdots + F^{-1}(U_n)|D_j^c) \\
= \; & n\mathbb{E}(F^{-1}(U_1)|D_1^c) \\
= \; & \frac{n}{1-nc_n}\int_{c_n}^{1-(n-1)c_n} F^{-1}(x)dx \\
= \; & \frac{n}{1-nc_n}\int_{c_n}^{\frac{1}{n}} F^{-1}(x)dx + \frac{n}{1-nc_n}\int_{\frac{1}{n}}^{1-(n-1)c_n} F^{-1}(x)dx \\
= \; & \frac{n}{1-nc_n}\int_{c_n}^{\frac{1}{n}} F^{-1}(x)dx + \frac{n}{1-nc_n}\int_{c_n}^{\frac{1}{n}} F^{-1}(1-(n-1)t)d(n-1)t \\
= \; & \frac{n}{1-nc_n}\int_{c_n}^{\frac{1}{n}} H(x)dx \\
= \; & H(c_n).
\end{aligned}
$$

The last equality holds because (4.45) and

$$\int_{c_n}^{\frac{1}{n}} H(x)dx = (\frac{1}{n} - c_n)H(c_n) \quad \text{for} \quad c_n > 0.$$

Therefore, almost surely

$$
\begin{aligned}
S &= F^{-1}(U_1) + \cdots + F^{-1}(U_n) \\
&= \sum_{i=1}^{n} F^{-1}(U_i)I_{D_1} + \sum_{i=1}^{n} F^{-1}(U_i)I_{D_1^c} \\
&= \sum_{i=1}^{n} F^{-1}(U_i)I_{\bigcup_{j=1}^n A_j} + H(c_n)I_{D_1^c} \\
&= \sum_{i=1}^{n} F^{-1}(U_i)(\sum_{j=1}^{n} I_{A_j}) + H(c_n)I_{D_1^c} \\
&= \sum_{j=1}^{n} [F^{-1}(U_j) + (n-1)F^{-1}(1-(n-1)U_j)]I_{A_j} + H(c_n)I_{D_1^c} \\
&= \sum_{j=1}^{n} H(U_j)I_{A_j} + H(c_n)I_{D_1^c}.
\end{aligned}
$$

Since $c_n \le 1/n$ and the sets $A_1, \cdots, A_n$ and D_1^c are disjoint, we have

$$\begin{aligned} & \mathbb{P}(\sum_{j=1}^{n} H(U_j) I_{A_j} + H(c_n) I_{D_1^c} < t) \\ = \; & n\mathbb{P}(H(U_1) I_{\{U_1 \le c_n\}} < t) + \mathbb{P}(H(c_n) I_{D_1^c} < t) \\ = \; & \mathbb{P}(H(U_1/n) I_{\{U_1 \le nc_n\}} < t) + \mathbb{P}(H(c_n) I_{\{U_1 > nc_n\}} < t) \\ = \; & \mathbb{P}(H(U_1/n) I_{\{U_1 \le nc_n\}} + H(c_n) I_{\{U_1 > nc_n\}} < t). \end{aligned}$$

Hence there exists a $U \sim$U[0,1] such that

$$\sum_{j=1}^{n} H(U_j) I_{A_j} + H(c_n) I_{D_j^c} = H(U/n) I_{\{U \le nc_n\}} + H(c_n) I_{\{U > nc_n\}}.$$

□

Proof of Lemma 4.6.5.

(i) Under the assumption of F, $F^{-1}(x)$ is convex and differentiable. Thus $H_a(x)$ is convex and differentiable. The definition of $c_n(a)$ shows that the average of $H_a(x)$ on $[c_n(a), \frac{1}{n}(1-a)]$ is $H_a(c_n(a))$ if $0 < c_n(a) < \frac{1-a}{n}$, namely

$$\frac{1}{(1-a) - c_n(a)} \int_{c_n(a)}^{\frac{1}{n}(1-a)} H_a(t) dt = H_a(c_n(a)).$$

With $H_a(x)$ being convex, we have $H_a'(c_n(a)) \le 0$ and so $H_a'(x) \le 0$ on $[0, c_n(a)]$. Here $H_a'(x)$ denotes $\partial H_a(x)/\partial x$. Note that for $n > 2$, $H_a'(\frac{1-a}{n}) = ((n-1)^2 - 1)(F^{-1})'(\frac{1-a}{n}) > 0$ implies

$$\int_{c}^{\frac{1}{n}(1-a)} H_a(t) dt \ge (\frac{1}{n}(1-a) - c) H_a(c)$$

for some $c < \frac{1-a}{n}$, thus $c_n(a) < \frac{1-a}{n}$ always holds. For $n = 2$, $H_a'(x) \le 0$ on $[0, \frac{1-a}{n}]$ since $H_a'(\frac{1-a}{n}) = 0$ and H is convex.

(ii) It follows from similar arguments as in (i).

(iii) Suppose $c_n(a) > 0$. By the continuity of $H_a(x)$ w.r.t. x and (4.51), we know that $c_n(a)$ satisfies

$$\int_{c_n(a)}^{\frac{1}{n}(1-a)} H_a(t) dt = (\frac{1}{n}(1-a) - c_n(a)) H_a(c_n(a)).$$

Note that for any $c \in [0, \frac{1}{n}(1-a)]$,

$$\begin{aligned}\int_c^{\frac{1}{n}(1-a)} H_a(t)dt &= \int_c^{\frac{1}{n}(1-a)} (n-1)F^{-1}(a+(n-1)t)dt + \int_c^{\frac{1}{n}(1-a)} F^{-1}(1-t)dt \\ &= \int_{a+(n-1)c}^{a+\frac{n-1}{n}(1-a)} F^{-1}(t)dt + \int_{1-\frac{1}{n}(1-a)}^{1-c} F^{-1}(t)dt \\ &= \int_{a+(n-1)c}^{1-c} F^{-1}(t)dt.\end{aligned}$$

Thus it follows from the definition of $c_n(a)$ that $H_a(c_n(a)) = n\mathbb{E}[F^{-1}(V_a)]$. For the case $c_n(a) = 0$, it is obvious that $\psi(a) = n\phi(a) = n\mathbb{E}[F^{-1}(V_a)]$.

(iv) Note that in a given probability space, for any measurable set B with $\mathbb{P}(B) > 0$ and continuous random variable Z with cdf G, we have

$$\mathbb{E}(Z|B) \leq \mathbb{E}[Z|Z \geq G^{-1}(1-\mathbb{P}(B))].$$

To see this, denote the conditional distribution of Z on B by G_1 and the conditional distribution on $\{Z \geq G^{-1}(1-\mathbb{P}(B))\}$ by G_2. Then we have

$$\begin{aligned}G_2(x) &= \frac{\mathbb{P}(Z \leq x, G(Z) \geq 1-\mathbb{P}(B))}{\mathbb{P}(B)} \\ &= \frac{\max\{G(x)-1+\mathbb{P}(B), 0\}}{\mathbb{P}(B)} \\ &\leq \frac{\mathbb{P}(Z \leq x, B)}{\mathbb{P}(B)} = G_1(x), \;\; x \in \mathbb{R}, \qquad (4.64)\end{aligned}$$

which implies that for $U \sim$U[0,1],

$$\mathbb{E}(Z|B) = \mathbb{E}[G_1^{-1}(U)] \leq \mathbb{E}[G_2^{-1}(U)] = \mathbb{E}[Z|Z \geq G^{-1}(1-\mathbb{P}(B))]. \qquad (4.65)$$

Since $A = \bigcap_{i=1}^n \{U_i \in [a, 1-c_n(b)]\}$, we have $\mathbb{P}(A) \geq 1 - \frac{nc_n(b)}{1-a} > 0$ and $U_i \leq 1 - c_n(b)$ on A. By defining $Z = F^{-1}(U_i)I_{\{U_i \leq 1-c_n(b)\}} + F^{-1}(a)I_{\{U_i > 1-c_n(b)\}}$, it follows from (4.65) that

$$\begin{aligned}
\mathbb{E}[F^{-1}(U_i)|A] &= \mathbb{E}[Z|A] \\
&\leq \mathbb{E}[Z|Z \geq F^{-1}(1-c_n(b)-(1-a)\mathbb{P}(A))] \\
&\leq \mathbb{E}[F^{-1}(U_i)|U_i \in [1-c_n(b)-(1-a)\mathbb{P}(A), 1-c_n(b)]] \\
&\leq \mathbb{E}[F^{-1}(U_i)|U_i \in [a+(n-1)c_n(b), 1-c_n(b)]] \\
&< \mathbb{E}[F^{-1}(U_i)|U_i \in [b+(n-1)c_n(b), 1-c_n(b)]] \\
&= \mathbb{E}(F^{-1}(V_b)). \qquad (4.66)
\end{aligned}$$

(v) It follows from (i), (ii) and the arguments in *Remark* 4.6.1.

(vi) We first prove the case when F has a decreasing density. Since $H_a(x)$ is convex w.r.t. x and differentiable w.r.t. a, the definition of $c_n(a)$ implies that $c_n(a)$ is continuous. Hence $\phi(a) = n\mathbb{E}[F^{-1}(V_a)]$ is continuous.

Suppose $U_{a,1}, \cdots, U_{a,n} \sim \mathrm{U}[a,1]$ with copula $Q_n^{\tilde{F}_a}$. Then $F^{-1}(U_{a,1}), \cdots, F^{-1}(U_{a,n}) \sim \tilde{F}_a$ and have copula $Q_n^{\tilde{F}_a}$ too. By (v), we have

$$F^{-1}(U_{a,1}) + \cdots + F^{-1}(U_{a,n}) \geq \phi(a). \qquad (4.67)$$

Thus from (4.66) and (4.67) we have

$$\phi(a) \leq \mathbb{E}[\sum_{i=1}^{n} F^{-1}(U_{a,i})|A] < n\mathbb{E}(F^{-1}(V_b)) = \phi(b).$$

Next we prove the case when F has an increasing density. The continuity of $c_n(a)$ comes from the same arguments as above. By definition, $H_a(0)$ and $\psi(a)$ are continuous and increasing functions of a. So we only need to show that when $c_n(a)$ approaches 0, $H_a(0) - \psi(a)$ approaches 0. Suppose that as $a \nearrow a_0$, $c_n(a) \to 0$ and $c_n(a) \neq 0$ for $a_0 - \epsilon < a < a_0$ and $\epsilon > 0$. Then

$$\int_0^{\frac{1}{n}(1-a)} H_a(t)dt \to \frac{1}{n}(1-a_0)H_{a_0}(0),$$

which implies that

$$\psi(a) = \int_a^1 \frac{1}{1-a} F^{-1}(a+t)dt = \frac{n}{1-a} \int_0^{\frac{1}{n}(1-a)} H_a(t)dt \to H_{a_0}(0)$$

as $a \nearrow a_0$. Together with the continuity of $H_a(0) - \psi(a)$ we know $H_a(0) - \psi(a) \to 0$ as $a \to a_0$. □

REFERENCES

[1] Anderson, T. W. (2003). *An Introduction to Multivariate Statistical Analysis.* Third edition. Wiley, New York.

[2] Arlot, S., Blanchard, G. and Roquain, E. (2010a). Some nonasymptotic results on resampling in high dimension I: confidence regions. *Ann. Statist.* **38**, 51–82.

[3] Arlot, S., Blanchard, G. and Roquain, E. (2010b). Some nonasymptotic results on resampling in high dimension II: multiple tests. *Ann. Statist.* **38**, 83–99.

[4] Artzner, P. (1999). Application of coherent risk measures to capital requirements in insurance. *North American Actuarial Journal* **3**, 11–29.

[5] Bai, Z. and Saranadasa, H. (1996). Effect of high dimension: by an example of a two sample problem. *Statist. Sinica* **6**, 311–329.

[6] Baiocchi, F. (1991). Un problema di ottimizzatione nella classe di Fréchet. *Unpublished thesis,* Facoltà di Scienze Statistiche, Università La Sapienza, Roma.

[7] Baysal, R.E. and Staum, J. (2008). Empirical likelihood for value-at-risk and expected shortfall. *Journal of Risk* **11**(1), 3–32.

[8] Bertino, S. (1994). The minimum of the expected value of the product of three random variables in the Frèchet class. *Journal of the Italian Statistical Society*, **6**, 201–211.

[9] Bickel, P. and Levina, E. (2008). Regularized estimation of large covariance matrices. *Ann. Statist.* **36**, 199–227.

[10] Bradic, J., Fan, J. and Wang, W. (2011). Penalized composite quasi-likelihood for ultrahigh dimensional variable selection. *J. R. Statist. Soc. B* **73**, 325–349.

[11] Cai, T. and Liu, W. (2011). Adaptive Thresholding for sparse covariance matrix estimation. *J. Amer. Statist. Assoc.* **106**, 672–684.

[12] Cai, T. and Jiang, T. (2011). Limiting laws of coherence of random matrices with applications to testing covariance structure and construction of compressed sensing matrices. *Ann. Statist.* **39**, 1496–1525.

[13] (Cai, T. and Shen, X. (2010). *High-dimensional Data Analysis.* World Scientific, New Jersey.

[14] Cai, T., Zhang, C. H. and Zhou, H. (2010). Optimal rates of convergence for covariance matrix estimation. *Ann. Statist.* **38**, 2118–2144.

[15] Chen, S.X., Peng, L. and Qin, Y. (2009). Empirical likelihood methods for high dimension. *Biometrika* **96**, 711–722.

[16] Chen, S. and Qin, Y. (2010). A two-sample test for high-dimensional data with applications to gene-set testing. *Ann. Statist.* **38**, 808–835.

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

[18] Chen, S., Zhang, L. and Zhong, P. (2010). Tests for High-Dimensional Covariance Matrices. *J. Amer. Statist. Assoc.* **105**, 810–819.

[19] Chen, J., Peng, L. and Zhao, Y. (2009). Empirical likelihood based confidence intervals for copulas. *Journal of Multivariate Analysis.* **100**, 137–151.

[20] Chen, J., Variyath, A. M. and Abraham, B. (2008). Adjusted empirical likelihood and its properties. *J. Comput. Graph. Statist.* **17**, 426–443.

[21] Claeskens, G., Jing, B., Peng, L. and Zhou, W. (2003). Empirical likelihood confidence regions for comparison distributions and ROC curves. *The Canadian Journal of Statistics.* **31**(2), 173–190.

[22] Csorgo, M. and Horvath, L. (1993). Weighted Approximations in Probability and Statistics. *Wiley, Chichester.*

[23] Dakovic, R. and Czado, C. (2011). Comparing point and interval estimates in the bivariate t-copula model with applocation to financial data. *Stat. Papers*, **52**, 709–731.

[24] Deelstra. G, Dhaene J., Vanmaele, M. (2011). An overview of comonotonicity and its applications in finance and insurance, in: G. Di Nunno, B. Oksendal (Eds.), *Advanced Mathematical Methods for Finance*, Springer.

[25] Denneberg, D. (1994). *Non-additive Measure and Integral.* Kluwer Academic Publishers, Boston.

[26] Denuit, M., Genest, C., Marceau, É. (1999). *Stochastic bounds on sums of dependent risks.* Insurance: Mathematics and Economics **25**, 85–104.

[27] Dharmadhikari, S. W., Fabian, V. and Jogdeo, K. (1968). Bounds on the moments of martingale. *Ann. Math. Statist.* **39**, 1719–1723.

[28] Dharmadhikari, S.W. and Jogdeo, K. (1969). Bounds on moments of certain random variables. *Ann. Math. Statist.* **40**, 1506–1508.

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

[30] Dupuis, D. and Jones, B.L. (2006). Multivariate extreme value theory and its usefulness in understanding risk. *North American Actuarial Journal* **10**(4), 1 – 27.

[31] Embrechts, P and Höing, G. (2006). Extreme VaR scenarios in higher dimensions. *Extremes* **9**, 177–192.

[32] Embrechts, P., Höing, A. and Juri, A. (2003). Using copulae to bound the value-at-risk for functions of dependent risks. *Finance and Stoch.* **7**, 145–167.

[33] Embrechts, P., Lindskog, F. and McNeil, A. (2001). Modelling dependence with copulas and applications to risk management, in: *Handbook of Heavy Tailed Distributions in Finance*. S.T. Rachev (Ed.). Elsevier Science B.V. 2003.

[34] Embrechts, P., McNeil, A. and Straumann, D. (2002). Correlation and Dependence in Risk Management: Properties and Pitfalls, in: *Risk Management: Value at Risk and Beyond*, M. Dempster (Ed.), Cambridge University Press, Cambridge, 176–223.

[35] Embrechts, P and Puccetti, G. (2006a). Bounds for functions of dependent risks. *Finance and Stoch.* **10**, 341–352.

[36] Embrechts, P. and Puccetti, G. (2006b), Bounds for functions of multivariate risks. *Journal of Multivariate Analysis* **97**, 526–547.

[37] Embrechts, P. and Puccetti, G. (2010). Risk Aggregation. In: *Copula Theory and Its Applications*, P.Jaworski et al. (eds.), Lecture Notes in Statistics 198, Springer, Berlin/Heidelberg.

[38] Fishman, G.S. (1972). Variance reduction in simulation studies. *J. Statist. Comp. and Simulation* **1**, 173–182.

[39] Frees, E.W. and Valdez, E.A. (1998). Understanding relationships using copulas. *North American Actuarial Journal* **2**, 1–25.

[40] Fréchet, M. (1951). Sur les tableaux de correlation dont les marges sont donnés. *Annales de l'Université de Lyon Sect.A*, **14**, 53–77.

[41] Gaffke, N. and Rüschendorf, L. (1981). On a class of extremal problems in statistics. *Math. Operationsforschung Statist., Series Optimization*, **12**, 123–135.

[42] Genest, C., Ghoudi, K. and Rivest, L.P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika* **82**, 543–552.

[43] Hammersley, I.M and Handscomb, D.C. (1964). *Monte Carlo Methods*. Methuen, London.

[44] Haug, S. Klüppelberg, C. and Peng, L. (2011). Statistical models and methods for dependence in insurance data. *Journal of Korean Statistical Society* **40**, 125–139.

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

[46] Hoeffding, W. (1940). Masstabinvariante Korrelationstheorie. *Schriften des mathematischen Instituts und des Instituts für angewandte Mathematik des Universitat Berlin*, **5** 179–233.

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

[48] Joe, H. (1997). *Multivariate models and dependence concepts*. Chapman & Hall, London, 1997.

[49] John, S. (1971). Some optimal multivariate tests. *Biometrika* **58**, 123–127.

[50] John, S. (1972). The distribution of a statistic used for testing sphericity of normal distributions. *Biometrika* **59**, 169–173.

[51] Jones, B.L. and Zitikis, R. (2003). Empirical estimation of risk measures and related quantities. *North American Actuarial Journal* **7**, 44–54.

[52] Jones, B.L. and Zitikis, R. (2005). Testing for the order of risk measures: an application of L-statistics in an actuarial science. *Metron LXIII*, 193–211.

[53] Jones, B.L. and Zitikis, R. (2007). Risk measures, distortion parameters, and their empirical estimation. *Insurance: Mathematics and Economics* **41**, 279–297.

[54] Kaas, R., Goovaerts, M., Dhaene, J. and Denuit, M. (2001). *Modern Actuarial Risk Theory*. Kluwer Academic Publishers, Boston.

[55] Kaas, R., Laeven, R. and Nelsen, R. (2009). Worst VaR scenarios with given marginals and measures of association. *Insurance: Mathematics and Economics* **44**, 146-158.

[56] Kaiser, T. and Brazauskas, V. (2007). Interval estimation of actuarial risk measures. *North American Actuarial Journal* **10**(4), 249–268.

[57] Klugman, S.A., Panjer, H.H. and Willmot, G.E. (2008). *Loss Distributions: From Data to Decisions. Wiley.*

[58] Klugman, S.A. and Parsa, R. (1999). Fitting bivariate loss distributions with copulas. *Insurance: Mathematics and Economics* **24**, 139–148.

[59] Kuelbs, J. and Vidyashankar, A.N. (2010). Asymptotic inference for high dimensional data. *Ann. Statist.* **38**, 836–869.

[60] Knott, M. and Smith, C.S. (1998). On multivariate variance reduction. *Unpublished results.*

[61] Knott, M. and Smith, C.S. (2006). Choosing joint distributions so that the variance of the sum is small. *Journal of Multivariate Analysis* **97**, 1757–1765.

[62] Ledoit, O., and Wolf, M. (2002). Some Hypothesis Tests for the Covariance Matrix When the Dimension Is Large Compare to the Sample Size. *Ann. Statist.*, **30**, 1081–1102.

[63] Makarov, G.D. (1981). Estimates for the distribution function of the sum of two randm variables with given marginal distributions. *Theory. Probab. Appl.* **26**, 803C806.

[64] Mason, D.M. (1981). Bounds for weighted empirical distribution functions, *Ann. Probab.* **9**, 881–884.

[65] McNeil, A.J., Frey, R. and Embrechts, P. (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.

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

[67] Nagao, H. (1973). On some test criteria for covariance matrix. *Ann. Statist.* **1**, 700–709.

[68] Necir, A. and Meraghni, D. (2009). Empirical estimation of the proportional hazard premium for heavy-tailed claim amounts. *Insurance: Mathematics and Economics* **45**, 49–58.

[69] Nelsen, R. (2006). *An Introduction to Copulas*. Second Edition. Springer, New York.

[70] Nelsen, R. and Úbeda-Flores, M. (2012). Directional dependence in multivariate distributions. *Annals of the Institute of Statistical Mathematics* **64**(3), 677–685

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

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

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

[74] Peng, L. (2008). Estimating the probability of a rare event via elliptical copulas. *North American Actuarial Journal* **12**(2), 116–128.

[75] Peng, L., Qi, Y. and Van Keilegom, I. (2012). Jackknife empirical likelihood methods for copulas. *Test*, to appear.

[76] Peng, L., Qi, Y. and Wang, R. (2012). Empirical likelihood test for high-dimensional linear models. *Preprint*.

[77] Peng, L., Qi, Y., Wang, R. and Yang, J. (2012). Jackknife empirical likelihood methods for risk measures and related quantities. *Insurance: Mathematics and Economics*, to appear.

[78] Peng, L., Qi, Y. and Wang, F. (2011). Test for a mean vector with fixed or divergent dimension. *Technical report*.

[79] Portnoy, S. (1984). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large; I, consistency. *Ann. Statist.* **12**, 1298–1309.

[80] Portnoy, S. (1985). Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large; II, normal approximation. *Ann. Statist.* **13**, 1403–1417.

[81] Puccetti, G., Wang, B. and Wang, R. (2012). Advances in complete mixability. *Preprint*.

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

[83] Rachev, S.T. and Rüschendorf, L. (1998) *Mass Transportation Problems*. Volume I: Theory. Springer.

[84] Rüschendorf, L. (1980). Inequalities for the expectation of Λ-monotone functions. *Z. Wahrsch.* **54**, 341–554.

[85] Rüschendorf, L. (1982). Random variables with maximum sums. *Adv. Appl. Probab.* **14**, 623–632.

[86] Rüschendorf, L. (1983). Solution of a statistical optimization problem by rearrangement methods. *Metrika* **30**, 55–61.

[87] Rüschendorf, L. and Uckelmann, L. (2002). Variance minimization and random variables with constant sum, in: *Distributions with given marginals.* Cuadras, et al. (Eds.), Kluwer, 211–222.

[88] Salmon, F. (2009). Recipe for disaster: the formula that killed Wall street. *The Market Movers financial blog at Portfolio.com.*

[89] Sebastiani, P., Gussoni, E., Kohane, I. S. and Ramoni, M. F. (2003). Statistical Challenges in Functional Genomics. *Statist. Sci.* **18**, 33–70.

[90] Segers, J. (2010). Weak convergence of empirical copula processes under nonrestrictive smoothness assumptions. http://arxiv.org/abs/1012.2133

[91] Schmid, F. and Schmidt, R. (2007). Multivariate extensions of pearman's rho and related statistics. *Statistics and Probability Letters* **77**, 407–416.

[92] Shaked, M., Shanthikumar, J.G. (2007). *Stochastic orders.* Springer Series in Statistics.

[93] Shao, J. and Tu, D. (1995). *The Jackknife and Bootstrap.* Springer, New York.

[94] Shorack, G.R. and Wellner, J.A. (1986). *Empirical Processes with Applications to Statistics.* New York: Wiley.

[95] Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Statist. Univ. Paris* **8**, 229–231.

[96] Srivastava, M.S. (2009). A test for the mean vector with fewer observations than the dimension under non-normality. *Journal of Multivariate Analysis* **100**, 518–532.

[97] Srivastava, M.S. and Du, M. (2008). A test for the mean vector with fewer observations than the dimension. *Journal of Multivariate Analysis* **99**, 386–402.

[98] von Bahr, B. and Esseen, C.G. (1965). Inequality for the rth absolute moment of a sum of random variables, $1 \leq r \leq 2$. *Ann. Math. Statist.* **36**, 299–393.

[99] Wang, B. and Wang, R. (2011). The complete mixability and convex minimization problems for monotone marginal distributions. *Journal of Multivariate Analysis*, **102** 1344–1360.

[100] Wang, S. (1995). Insurance pricing and increased limits ratemaking by proportional hazards transforms. *Insurance: Mathematics and Economics* **17**, 43–54.

[101] Wang, S. (1996). Premium calculation by transforming the layer premium density. *ASTIN Bulletin* **26**, 71–92.

[102] Wang, S. (1998). An actuarial index of the right-tail risk. *North American Actuarial Journal* **2**, 88–101.

[103] Wang, S. and Young, V.R. (1998). Ordering risks: expected utility theory versus Yaari's dual theory of risk. *Insurance: Mathematics and Economics* **22**, 145–161.

[104] Wang, S., Young, V.R. and Panjer, H.H. (1997). Axiomatic characterization of insurance prices. *Insurance: Mathematics and Economics* **21**, 173–183.

[105] Wang, R. and Peng, L. (2011). Jackknife empirical likelihood for Spearman's rho. *North American Actuarial Journal*, **15**(4), 475-48.

[106] Wang, R., Peng, L. and Qi, Y. (2012) Jackknife empirical likelihood test for equality of two high dimensional means. *Preprint*.

[107] Wang, R., Peng, L. and Yang, J. (2012). Jackknife empirical likelihood for parametric copulas. *Scandinavian Actuarial Journal*, to appear.

[108] Wang, R., Peng, L. and Yang, J. (2012). Bounds for the sum of dependent risks and worst Value-at-Risk with monotone marginal densities. *Preprint*.

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

[110] Wirch, J.L. and Hardy, M.R. (1999). A synthesis of risk measures for capital adequacy. *Insurance: Mathematics and Economics* **25**, 337–347.

[111] Yaari, M.E. (1987). The dual theory of choice under risk. *Econometrica* **55**, 95–115.

[112] Yang, J., Qi, Y. and Wang, R. (2009). A class of multivariate copulas with bivariate Fréchet marginal copulas. *Insurance: Mathematics and Economics* **45**(1), 139-147.

[113] Zhang, R., Peng, L. and Wang, R. (2012). Tests for covariance matrix with fixed or divergent dimension. *Preprint*.