

Conditional Extremes in Asymmetric Financial Markets

by

Jinyuan Zhang

B.Sc., The Chinese University of Hong Kong, 2009

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

in

The Faculty of Graduate and Postdoctoral Studies
(Statistics)

THE UNIVERSITY OF BRITISH COLUMBIA
(Vancouver)

July 2015

© Jinyuan Zhang 2015

Abstract

The project focuses on the estimation of the probability distribution of a bivariate random vector given that one of the components takes on a large value. These conditional probabilities can be used to quantify the effect of financial contagion when the random vector represents losses on financial assets and as a stress-testing tool in financial risk management. However, it is tricky to quantify these conditional probabilities when the main interest lies in the tails of the underlying distribution. Specifically, empirical probabilities fail to provide adequate estimates while fully parametric methods are subject to large model uncertainty as there is too little data to assess the model fit in the tails.

We propose a semi-parametric framework using asymptotic results in the spirit of extreme values theory. The main contributions include an extension of the limit theorem in Abdous et al. [Canad. J. Statist. 33 (2005)] to allow for asymmetry, frequently encountered in financial and insurance applications, and a new approach for inference. The results are illustrated using simulations and two applications in finance.

Preface

This work was prepared under the supervision of Professor Natalia Nolde. Chapter 3 and 4.1 were developed based on draft of Professor Natalia Nolde. The remaining chapters were independently written by the author. Professor Natalia Nolde provided many valuable suggestions on improving the material in these chapters.

Table of Contents

| | |
|---------------------------------------|------|
| Abstract | ii |
| Preface | iii |
| Table of Contents | iv |
| List of Tables | vii |
| List of Figures | x |
| List of Algorithms | xii |
| Acknowledgements | xiii |
| Dedication | xiv |
| 1 Introduction | 1 |
| 2 Preliminary Theory | 5 |
| 2.1 Extreme Value Theory | 5 |
| 2.1.1 Univariate Case | 5 |
| 2.1.2 Multivariate Case | 7 |
| 2.2 Regular Variation | 8 |
| 2.2.1 Definition | 9 |

Table of Contents

| | | |
|----------|---|----|
| 2.2.2 | Link to MVEVD | 11 |
| 2.3 | Inference | 11 |
| 2.3.1 | Tail Index Estimators | 11 |
| 2.3.2 | Spectral Measure | 14 |
| 3 | Review of Skewed Distributions | 17 |
| 3.1 | Elliptical Distributions | 17 |
| 3.2 | Skew-symmetric Distributions | 18 |
| 3.3 | Skew-elliptical Distributions | 20 |
| 3.3.1 | Skew-normal Distribution | 21 |
| 3.3.2 | Skew- t Distribution | 22 |
| 4 | Theoretical Results | 25 |
| 4.1 | Limit Results | 25 |
| 5 | Inference | 29 |
| 5.1 | Tail Index Estimation | 29 |
| 5.2 | Parametric EVT Estimation | 31 |
| 5.3 | Iterative Parametric EVT Estimation | 34 |
| 5.4 | Estimation of the Scale Matrix | 36 |
| 5.5 | Simulation Studies | 37 |
| 6 | Financial Applications | 42 |
| 6.1 | Financial Contagion | 42 |
| 6.2 | CoVaR | 48 |
| 7 | Conclusion | 52 |
| | Bibliography | 54 |

Table of Contents

| | |
|--|----|
| Appendices | |
| A Proofs | 61 |
| A.1 Proof of Lemma 3.3.2 | 61 |
| A.2 Proof of Proposition 4.1.1 | 61 |
| A.3 Proof of Lemma 4.1.2 | 63 |
| A.4 Proof of Theorem 4.1.3 | 64 |
| B Tables | 67 |
| B.1 Simulation Studies | 67 |
| B.2 CoVaR | 70 |

List of Tables

5.1 Performance of estimator $\hat{\nu}$ computed using Program 1. The bivariate skew- t distribution with tail index $\nu = 2$ is used in simulations to generate samples of four different sample sizes. Each cell presents the average value of estimates of ν based on 1000 simulated samples of a given size, and the corresponding standard error (in brackets). Two different parameter settings are considered. Case 1: $\alpha = (1, -3)$, $\rho = 0.5$, $\xi = (0, 0)$; Case 2: $\alpha = (1, -3)$, $\rho = 0.5$, $\xi = (3, 1)$ 30

5.2 Simulation results based on 1000 samples of size 1000 from a bivariate skew- t distribution with parameters $\xi = (3, 1)$, $\alpha = (1, -3)$, $\nu = 2$, $\rho = 0.5$ and a standardized scale matrix. Each cell provides the average (standard deviation) of the estimates of $\eta(x, y)$ under various methods; see Section 5.5 for details. For $\hat{\eta}_{AFG}(x, y)$, $\hat{\eta}_1(x, y)$, $\hat{\eta}_2(x, y)$ and $\eta_{\text{lim}}(x, y)$, we used $z = y/x - \rho$ in the limit results. Values of x and y are chosen as the theoretical marginal quantiles with probability p , where p labels columns and rows. 40

5.3 Simulation results based on 1000 samples of size 1000 from a bivariate skew- t distribution with parameters $\xi = (3, 1)$, $\alpha = (1, -3)$, $\nu = 2$, $\rho = 0.5$ and $\omega = \text{diag}(2, 3)$ for the scale matrix. Each cell provides the average (standard deviation) of the estimates of $\eta(x, y)$ under various methods; see Section 5.5 for details. For $\hat{\eta}_{AFG}(x, y)$, $\hat{\eta}_1(x, y)$, $\hat{\eta}_2(x, y)$ and $\eta_{\text{lim}}(x, y)$ we used $z = \omega_1 y / \omega_2 x - \rho$ in the limit results. Values of x and y are chosen as the theoretical marginal quantiles with probability p , where p labels columns and rows. . . . 41

List of Tables

| | | |
|-----|---|----|
| 6.1 | Point estimation of extreme conditional excess probability for 28 financial institutions. Daily losses are computed using log returns and are filtered by the AR(1)-GARCH(1,1) process. The threshold values of x and y are, respectively, 99.9% quantile of losses on DJUSFN and average of 99% quantile of losses on cross-sectional stocks. The sample period is from June 1, 2006 to May 31, 2008. The pre-crisis period is from June 1, 2006 to May 31, 2007, and the crisis period is from June 1, 2007 to May 31, 2008. | 46 |
| 6.2 | Summary statistics for cross-sectional $\Delta\text{CoVaR}_{q_1,t}^{s j}$ for all institutions during sample period from June 1, 2006 to May 31, 2007. Level q_1 is set to be 5%, and q_2 is 5%, 1% or 0.01%. “EVT” and “skew- t ”, respectively, refer to the use of the iterative EVT method and the bivariate skew- t distribution to model the sequence of standardized residuals $\{\hat{\mathbf{z}}_t\}$. Estimation based on the empirical distribution is reported under “Empirical”. Column “Std.TS” reports the average of the standard deviations of individual $\Delta\text{CoVaR}_{q_1,t}^{s j}$ and Column “Std.CS” reports the standard deviations of the mean of each individual $\Delta\text{CoVaR}_{q_1,t}^{s j}$ measure. | 51 |
| B.1 | Simulation results based on 1000 samples of size 1000 from a bivariate skew- t distribution with parameters $\boldsymbol{\xi} = (0,0)$, $\boldsymbol{\alpha} = (0,0)$, $\nu = 2$, $\rho = 0.5$ and $\omega = \text{diag}(1, 1)$ for the scale matrix. Each cell provides the average (standard deviation) of the estimates of $\eta(x, y)$ under various methods; see Section 5.5 for details. For $\hat{\eta}_{AFG}(x, y)$, $\hat{\eta}_1(x, y)$, $\hat{\eta}_2(x, y)$ and $\eta_{\text{im}}(x, y)$ we used $z = \omega_1 y / \omega_2 x - \rho$ in the limit results. Values of x and y are chosen as the theoretical marginal quantiles with probability p , where p labels columns and rows. . . . | 68 |

List of Tables

B.2 Simulation results based on 1000 samples of size 1000 from a bi-variate skew- t distribution with parameters $\xi = (3,1)$, $\alpha = (1,-3)$, $\nu = 20$, $\rho = 0.5$ and $\omega = \text{diag}(2,3)$ for the scale matrix. Each cell provides the average (standard deviation) of the estimates of $\eta(x, y)$ under various methods; see Section 5.5 for details. For $\hat{\eta}_{AFG}(x, y)$, $\hat{\eta}_1(x, y)$, $\hat{\eta}_2(x, y)$ and $\eta_{\text{im}}(x, y)$ we used $z = \omega_1 y / \omega_2 x - \rho$ in the limit results. Values of x and y are chosen as the theoretical marginal quantiles with probability p , where p labels columns and rows. . . . 69

B.3 Summary statistics for cross-sectional $\Delta\text{CoVaR}_{q_1,t}^{s|j}$ for all institutions during sample period from June 1, 2006 to May 31, 2007. $\text{VaR}_{q_2,t}^j$ is estimated by assuming that \hat{Z}_t^j follows the skew- t distribution, and $\text{CoVaR}_{q_1,t}^{s|b_j}$ is estimated by assuming that $\{\hat{\mathbf{Z}}_t\}$ follows a bivariate skew- t distribution. 70

B.4 Summary statistics for cross-sectional $\Delta\text{CoVaR}_{q_1,t}^{s|j}$ for all institutions during sample period from June 1, 2006 to May 31, 2007. $\text{VaR}_{q_2,t}^j$ is estimated using EVT method described in McNeil and Frey [2000], and $\text{CoVaR}_{q_1,t}^{s|b_j}$ is estimated by assuming that $\{\hat{\mathbf{Z}}_t\}$ follows a bivariate skew- t distribution. 71

B.5 Summary statistics for cross-sectional $\Delta\text{CoVaR}_{q_1,t}^{s|j}$ for all institutions during sample period from June 1, 2006 to May 31, 2007. $\text{VaR}_{q_2,t}^j$ is estimated by assuming that \hat{Z}_t^j follows the skew- t distribution, and $\text{CoVaR}_{q_1,t}^{s|b_j}$ is estimated empirically. 71

List of Figures

| | | |
|-----|---|----|
| 1.1 | Scatter plot of the daily log losses (negative returns) on the stock of Bank of America Corp (BAC) and JP Morgan Chase & Co (JPM) over a period of 500 days from June 1, 2005 to May 31, 2007 (left panel) and a period of 501 days from June 1, 2007 to May 31, 2009 (right panel). | 2 |
| 4.1 | Plot of the values of $\lim_{x \rightarrow \infty} \mathbb{P}(Y \leq (z + \rho)x \mid X > x)$ in terms of z for a bivariate skew- t distribution with $\boldsymbol{\xi} = (5, -5)$, $\boldsymbol{\alpha} = (1, -3)$, $\rho = 0.5$ and $\nu = 2$: a) limiting value from Theorem 4.1.3, and true value with x being (b) 99% marginal quantile, (c) 99.99% marginal quantile, and (d) 200. | 28 |
| 5.1 | Mean (left panel) and standard deviation (right panel) of absolute differences between exact value of $\eta(x, y)$ and estimated probability $\hat{\eta}(x, y)$ using Theorem 4.1.3 for different thresholds. Samples are drawn from the skew- t distribution with $\boldsymbol{\xi} = (0, 0)$, $\boldsymbol{\alpha} = (1, -3)$, $\nu = 2$, $\rho = 0.5$. Values of x and y are chosen as the theoretical marginal quantiles with probability 99.99%. | 33 |
| 5.2 | The contour plots $\{\mathbf{z} \in \mathbb{R}^2 \mid \psi_{st}(\mathbf{z}) < 1\}$ for spectral density ψ_{st} in (5.1). The parameters are set to $\nu = 2$, $\rho = 0.5$ and several values of $\boldsymbol{\alpha}$ (left panel); $\boldsymbol{\alpha} = (3, -1)$ and several values of ρ (right panel). | 33 |

List of Figures

| | | |
|-----|--|----|
| 5.3 | Comparison of the true level curve $\{\psi_{st}(\mathbf{z}) < 1\}$ with the estimated contours $\{\hat{\psi}_{st}(\mathbf{z}) < 1\}$ using the AFG method and the proposed iterative parametric EVT method. The data are generated from a bivariate skew- t distribution with parameters $\boldsymbol{\xi} = (3, 1)$, $\boldsymbol{\alpha} = (1, -3)$, $\rho = 0.5$, $\nu = 2$, and $\boldsymbol{\omega} = (1, 1)$ (left panel) or $\boldsymbol{\omega} = (2, 3)$ (right panel). | 39 |
| 6.1 | Estimated contours $\{\hat{\psi}_{st}(\mathbf{z}) < 1\}$ for daily losses on DJUSFN versus Hudson City Bancorp (HCBK, left panel) and Peoples Bank Bridgeport (PBCT, right panel) between June 1, 2006 and May 31, 2007 using the AFG method and the proposed iterative parametric EVT method. | 45 |

List of Algorithms

| | | |
|---|--|----|
| 1 | Estimation of tail index ν | 30 |
| 2 | Estimation using parametric EVT method | 32 |
| 3 | Estimation using iterative parametric EVT method | 36 |

Acknowledgements

First and foremost I would like to thank my supervisor Natalia Nolde. It has been an honour to be her first master student. I appreciate all her contributions of time, ideas, and funding to make my experience productive and stimulating. The joy and enthusiasm she has for her research inspire me. Natalia, many thanks for your kind support and guidance while encouraging me to move beyond my intellectual comfort zones.

My time at UBC was enjoyable thank to my dear friends. I am grateful for time spent with my friends and flatmates. Big thanks to Celina Liu and Kun Wang for being true inspirations and such an integral part of my graduate school experience. Doing projects with Celina have kept my mind sharp. I will always remember the passionate discussions with Kun Wang - work related or not. Their enthusiasm, curiosity and encouragement brought much fun to the study and research process. I have been very fortunate to be friend with so many incredible people in my department. Thanks for all of the wonderful conversations. They have enriched my life in graduate school.

Most of all, I would love to thank my family for all their love and encouragement. I am grateful for my mom. Her unfailing love and unconditional support in all of my pursuits are invaluable. I would like to thank Xi Kang for taking care of me when I suffer and for her understanding, faithful support, encouragement and patience.

Dedication

TO MY PARENTS AND GRANDPARENTS

Chapter 1

Introduction

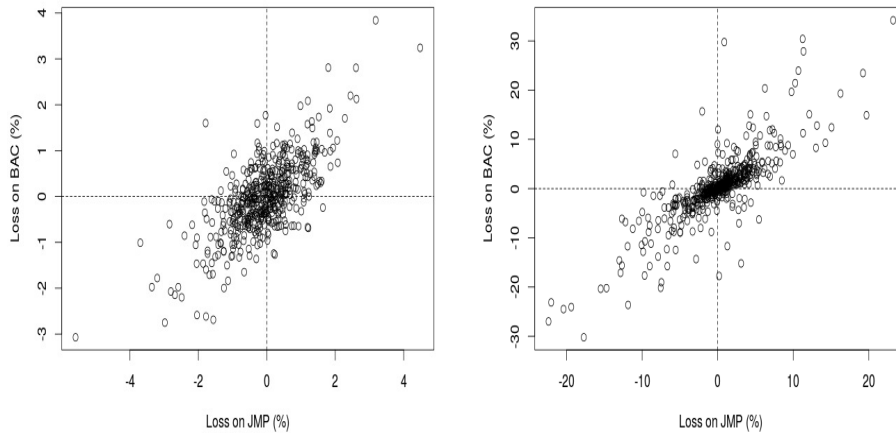
In a wide range of applications, it is of interest to quantify how large movements in one variable affect another variable. Particularly, in finance and insurance, one is often concerned with the impact of a stock market crash on individual stocks, or with the impact of a large loss on one financial asset on other assets. One common approach to assess such influences is by considering conditional probabilities in which the conditioning event is extreme. Let (X, Y) be a bivariate random vector, which can be interpreted as representing losses on two financial assets. The conditional probability

$$\eta(x, y) := \mathbb{P}(Y \leq y | X > x), \quad (1.1)$$

when x is large, can be seen as a measure of the effect the occurrence of a large loss on X exerts on the distribution of losses Y . When x and y in (1.1) are chosen as p -marginal quantiles, the limit of $1 - \eta(x, y)$, if it exists, as the quantile level p goes to one, is known as the (upper) tail dependence coefficient (Joe [1997]). It is a widely used way to quantify extremal dependence in the presence of tail dependence, i.e., when the coefficient is strictly positive (McNeil et al. [2005]). In finance, the tail dependence coefficient is often interpreted as a measure of financial contagion and plays a role in a variety of applications including risk management, derivative pricing and portfolio selection; see, e.g., Poon et al. [2004], Burtschell et al. [2009], Chan-Lau et al. [2004], Aloui et al. [2011] and DiTraglia and Gerlach [2013].

Figure 1.1 shows the scatter plot of the daily log losses (negative returns) on the stock of Bank of America Corp (BAC) and JP Morgan Chase & Co (JPM) over a period of 500 days from June 1, 2005 to May 31, 2007 and a period of 501 days from June 1, 2007 to May 31, 2009. It is obvious that these two stocks exhibit weaker dependence in the extreme movements over

Figure 1.1: Scatter plot of the daily log losses (negative returns) on the stock of Bank of America Corp (BAC) and JP Morgan Chase & Co (JPM) over a period of 500 days from June 1, 2005 to May 31, 2007 (left panel) and a period of 501 days from June 1, 2007 to May 31, 2009 (right panel).



the second period, which is widely recognized as financial crisis period. The difficulty to quantify this extreme dependence precisely via statistical estimation is various. Firstly, it is noted that for x large it becomes exceedingly difficult to estimate $\eta(x, y)$ empirically as there are too few, if any, data points falling in the associated extremal region. On the other hand, parametric models fitted using the entire dataset are subject to undue influence of the central observations possibly compromising the fit in the tail regions. Such a fully parametric approach naturally gives rise to a significant model risk. An effective approach in situations dealing with extreme events is to rely on asymptotic models in the spirit of extreme value theory (EVT). The study of conditional extreme value models was initiated by Heffernan and Tawn [2004], followed by Heffernan and Resnick [2007] and Das and Resnick [2011], among others.

Similar to Abdous et al. [2005a], we focus on the estimation of $\eta(x, y)$ for x large but extend their methodology from the class of elliptical distributions to a more general framework of asymmetric models describing the stochastic

behaviour of underlying tail risks.

Elliptical distributions are frequently used to model the stationary distribution of financial losses and returns; see, e.g., McNeil et al. [2005] for an overview. They are a natural generalization of the multivariate normal distribution by giving flexibility in modelling the tail behaviour. Typically, financial data exhibit slower decay in the tails that can be captured by the student- t distribution. However, elliptical distributions fail to address the likely tail asymmetry in the distribution of financial returns, both for the marginals and the (tail) dependence. Empirical evidence of skewness of financial returns has been documented by Longin and Solnik [2001], Rockinger and Jondeau [2002] and Chang et al. [2013], among others. Applications of skewed distributions in finance and insurance along with implications of asymmetry for asset pricing and risk management are reviewed in Adcock et al. [2012]. One example is that a stock's co-skewness with market could shed light on investor's behaviour, especially in a downside market (see e.g., Harvey and Siddique [2000], Alles and Murray [2013], Conrad et al. [2013]).

Our aim here is to relax the assumption of elliptical symmetry while still retaining a certain parametric structure which can be used to develop asymptotic approximations as well as subsequent inference procedures using EVT. First, assuming that (X, Y) comes from a sub-family of skew-elliptical distributions of Azzalini and Capitanio [2003], we derive a limit expression for $\eta(x, y)$ as $x \rightarrow \infty$ and y is chosen to grow at a suitable rate to ensure existence of a non-degenerate limit. The limit expression is then used as an approximation for large but finite values of x . Estimation of unknown parameters is carried out using a semi-parametric procedure, consisting of two main components. The tail behaviour is modelled under the assumption of regular variation and standard EVT techniques are employed to estimate the tail index parameter. Dependence in the joint tail region is represented by the spectral density for which we assume a specific parametric form based on the multivariate skew- t distribution. There exist other more general (semi-)parametric modelling approaches for spectral densities as, for instance, proposed in Peiro [1999], Boldi and Davison [2007] and Beran and Mainik [2014]. However, they tend to come at a significant computational or model-formulation costs, and will not be explored here.

The report is organized as follows. Chapter 2 gives basic overview of extreme value theory and regular variation to be used in the sequel. It also includes some inference methodologies. In Chapter 3 we review a general class of skew-symmetric distributions with a special attention to its skew-elliptical sub-family. These skewed distributions will form the basis for modelling asymmetry in the extremal behaviour of multivariate random vectors. Chapter 4.1 summarizes a number of limit results, which will be used to approximate the condition probability $\eta(x, y)$ in (1.1) as well as to justify subsequent inference procedures. In Chapter 5, we give details on model fitting and estimation of $\eta(x, y)$, and illustrate performance of the proposed methods in several simulation studies. Chapter 6 gives an application of the developed methodology to financial data. Finally, Chapter 7 presents some concluding remarks and outlines directions for future research. Proofs and additional tables are delegated to the Appendix.

Chapter 2

Preliminary Theory

2.1 Extreme Value Theory

This section briefly reviews classical EVT, which mainly deals with the asymptotic distribution of sample maxima. The presentation is given for upside gains but the same results holds for downside losses.

2.1.1 Univariate Case

The classical univariate EVT dates back to the work of Fisher and Tippett [1928] and Gnedenko [1943]. Let Z_1, \dots, Z_n be a sequence of independent and identically distributed (i.i.d.) random variables with population cumulative distribution function F , and M_n be the maximum of the sequence. Then the distribution for M_n is

$$\mathbb{P}(M_n < x) = \mathbb{P}(Z_1 \leq x, \dots, Z_n \leq x) = (F(x))^n.$$

If the distribution F is known, then the distribution of M_n can be determined; otherwise, it can be approximated by modelling through asymptotic theory of M_n . When $n \rightarrow \infty$, the distribution of M_n will degenerate to a point mass, but this problem, in most cases, can be eliminated by allowing some linear normalization of M_n .

Theorem 2.1.1. (*Fisher-Tippett Theorem*) *If there exist normalization constants $a_n > 0$ and $b_n \in \mathbb{R}$, such that as $n \rightarrow \infty$*

$$\mathbb{P}\left(\frac{M_n - b_n}{a_n} \leq x\right) \rightarrow H(x),$$

where $H(x)$ is some non-degenerate distribution function, then $H(x)$ belongs to one of the following three families of distributions:

2.1. Extreme Value Theory

- *Gumbel*: $H(x) = \exp\{-\exp(-x)\}$, $x \in \mathbb{R}$;
- *Fréchet*: $H(x) = \begin{cases} 0, & x \leq 0 \\ \exp\{-x^{-\nu}\}, & x > 0, \nu > 0 \end{cases}$,
- *Weibull*: $H(x) = \begin{cases} \exp\{-(-x)^{-\nu}\}, & x \leq 0, \nu < 0 \\ 1, & x > 0 \end{cases}$,

where ν is the shape parameter which describes the fatness of the tail.

A rigorous proof of the theorem can be found in Gnedenko [1943]. Von Mises [1936] and Jenkinson [1955] derive a one-parameter representation of the three limit distributions above, which is known as the standard generalized extreme value (GEV) distribution.

Definition 2.1.1. *The distribution function of the standard GEV distribution is given by*

$$H(x) = \begin{cases} \exp\{-(-(1 + \xi x))^{-\frac{1}{\xi}}\}, & \xi \neq 0 \\ \exp\{-\exp(-x)\}, & \xi = 0 \end{cases},$$

where $1 + \xi x > 0$.

The Gumbel ($\xi = 0$) distribution is related to light-tailed distributions such as normal, log-normal or exponential; the Fréchet distribution ($\xi > 0$) is related to heavy-tailed distributions such as Pareto, Cauchy, or Student distribution; the Weibull ($\xi < 0$) is associated to finite support distributions such as the uniform distribution. In practice, there are various ways to adopt the EVT to perform inference for the tail index $\nu = \frac{1}{\xi}$. For details, please refer to Section 2.3.1.

In practice, above results can be applied to perform inference for block (e.g., annual) maxima data using the maximum-likelihood method. However, this method can be wasteful of data due to the trade-off between the size of the blocks and the number of blocks to be constructed from a given dataset. In contrast, the peaks over threshold (POT) method, which employs the data of exceedances over a high threshold to estimate the tail of distribution, is a more efficient approach.

The generalized Pareto distribution (GPD) is the pivotal distribution for modelling the data of exceedances over a high threshold, see Embrechts et al. [1997] and McNeil et al. [2005], among others.

Definition 2.1.2. (*Generalized Pareto Distribution*). *The distribution function of the GPD is given by*

$$H_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi \frac{x}{\beta})^{-\frac{1}{\xi}}, & \xi \neq 0 \\ 1 - \exp\{-\frac{x}{\beta}\}, & \xi = 0 \end{cases},$$

where ξ is the shape parameter and β is an additional scale parameter.

When $\xi > 0$, $H_{\xi,\beta}(x)$ is a re-parameterized version of a heavy-tailed, ordinary Pareto distribution; when $\xi = 0$, we have a light-tailed, exponential distribution; when $\xi < 0$, $H_{\xi,\beta}(x)$ corresponds to a bounded (i.e. short-tailed), Pareto type II distribution.

For a random variables Z with the cumulative probability function F , the conditional excess distribution of Z over a certain threshold u is given by

$$F_u(x) = \mathbb{P}(Z - u \leq x | Z > u) = \frac{F(u+x) - F(u)}{1 - F(u)}.$$

A famous limit result by Balkema and De Haan [1974] and Pickands [1975] shows that the GPD is the limiting distribution for the data of exceedances over a high threshold. This results allows us to perform inference for the data of exceedances over a high threshold (the tail observations).

2.1.2 Multivariate Case

Multivariate EVT studies the limiting distribution of a vector of appropriately normalized coordinate-wise maxima. Assuming $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ are a sequence of i.i.d. d -dimensional random vectors, and

$$\mathbf{M}_n = (\max_{i=1}^n(Z_{i,1}), \dots, \max_{i=1}^n(Z_{i,d}))^T,$$

then \mathbf{M}_n can be modelled by the multivariate extreme value distribution (MVEVD).

2.2. Regular Variation

Theorem 2.1.2. *If there exist vectors of normalizing constants $\mathbf{a}_n > \mathbf{0}$ and $\mathbf{b}_n \in \mathbb{R}^d$, such that as $n \rightarrow \infty$*

$$\mathbb{P}\left(\frac{\mathbf{M}_n - \mathbf{b}_n}{\mathbf{a}_n} \leq \mathbf{x}\right) \rightarrow H(\mathbf{x}), \quad (2.1)$$

where $H(\mathbf{x})$ is some non-degenerate distribution function, then $H(\mathbf{x})$ is called the multivariate extreme value distribution with univariate GEV marginals.

The details of above proposition can be found in de Haan and Resnick [1977] and Resnick [1987]. A typical procedure to model multivariate extremes typically involves two steps: marginal estimation and dependence estimation. Marginal block maximum are first modelled using the GEV distribution and then transformed based on the fitted distribution to have a unit Fréchet distribution. The dependence structure can then be modelled via an existing parametric distribution or non-parametric distribution, see, e.g., Coles and Tawn [1994].

The POT method can also be applied to do inference for the multivariate data; however, the definition of a threshold exceedance is not as obvious as in the univariate case. One approach is to employ a multivariate GPD (Rootzén and Tajvidi [2006]), which leads to the study of data exceeding marginal thresholds. Another one is to define a multivariate threshold exceedance in terms of the norm of a random vector.

2.2 Regular Variation

It is widely accepted that high frequency financial return data comes from a distribution with heavy tails. We make this assumption throughout the project. Mathematically, heavy tails are often formalized by imposing the condition of regular variation on the tail of the underlying distribution. Multivariate regular variation provides a probability framework for modelling the joint tail of a random vector with threshold-exceedance data. Generally speaking, it decomposes data into polar coordinates, and then characterizes tail dependence by a limiting angular measure on the unit sphere under a fixed norm.

2.2. Regular Variation

In this section we review some results on regular variation, which will be used in the later parts of the project. For more details, see, e.g., Resnick [2006].

2.2.1 Definition

Definition 2.2.1. A positive measurable function h on $(0, \infty)$ is said to be regularly varying at infinity with index $\nu > 0$ (written $h \in RV_{-\nu}$) if $\lim_{t \rightarrow \infty} h(tx)/h(t) = x^{-\nu}$ for all $x > 0$.

Definition 2.2.2. A non-negative random variable Z is said to be regularly varying with index $\nu > 0$ if for every $x > 0$,

$$\lim_{t \rightarrow \infty} \frac{\mathbb{P}(Z > tx)}{\mathbb{P}(Z > t)} = x^{-\nu}.$$

Definition 2.2.3. A random vector \mathbf{Z} on \mathbb{R}^d and its distribution are said to be multivariate regularly varying with index $\nu > 0$ if

$$\lim_{t \rightarrow \infty} \frac{\mathbb{P}(\|\mathbf{Z}\| \geq tx, \mathbf{Z}/\|\mathbf{Z}\| \in D)}{\mathbb{P}(\|\mathbf{Z}\| \geq t)} = x^{-\nu} \Psi(D), \quad (2.2)$$

for every $x > 0$ and Borel set D in $\mathcal{S}^{d-1} = \{\mathbf{z} \in \mathbb{R}^d : \|\mathbf{z}\| = 1\}$ with $\Psi(\partial D) = 0$, where Ψ is a spectral probability measure on \mathcal{S}^{d-1} , and $\|\cdot\|$ denotes the L_2 -norm.

Condition (2.2) is equivalent to having

$$\lim_{t \rightarrow \infty} \frac{\mathbb{P}(\|\mathbf{Z}\| \geq tx)}{\mathbb{P}(\|\mathbf{Z}\| \geq t)} = x^{-\nu}, \quad x > 0, \quad (2.3)$$

and there must exist a measure μ such that

$$\lim_{t \rightarrow \infty} \frac{\mathbb{P}(\mathbf{Z} \in tE)}{\mathbb{P}(\|\mathbf{Z}\| \geq t)} = \mu(E) < \infty \quad (2.4)$$

for every Borel set E on \mathbb{R}^d , bounded away from the origin and satisfying $\mu(\partial E) = 0$. Measures Ψ in (2.2) and μ in (2.4) are related via

$$\Psi(D) = \mu(E_{1,D}) \quad \text{for } D \subset \mathcal{S}^{d-1} \text{ and } E_{1,D} = \{\mathbf{z} \in \mathbb{R}^d : \|\mathbf{z}\| \geq 1, \mathbf{z}/\|\mathbf{z}\| \in D\}.$$

2.2. Regular Variation

Definition 2.2.3 suggests a possibility to model multivariate tails by separating the (radial) tail behaviour and the extremal dependence structure expressed in the form of measure Ψ in (2.2). However, condition (2.2) may be cumbersome to work with as often multivariate distributions are specified in terms of their densities. The following result, due to de Haan and Resnick [1987], gives sufficient conditions on the density for (2.2) to hold; see also Cai et al. [2011].

Theorem 2.2.1. *Let f denote the density of random vector \mathbf{Z} on \mathbb{R}^d . Suppose f is positive and continuous, and for some regularly varying function V with index $\nu > 0$ the following two limit conditions hold:*

$$\lim_{t \rightarrow \infty} \frac{f(t\mathbf{z})}{t^{-d}V(t)} = q(\mathbf{z}) > 0 \quad \text{for every } \mathbf{z} \neq \mathbf{0}, \quad (2.5)$$

and

$$\lim_{t \rightarrow \infty} \sup_{\mathbf{z} \in \mathcal{S}^{d-1}} \left| \frac{f(t\mathbf{z})}{t^{-d}V(t)} - q(\mathbf{z}) \right| = 0. \quad (2.6)$$

Then \mathbf{Z} is multivariate regularly varying with index ν . Necessarily, q is homogeneous: $q(t\mathbf{z}) = t^{-\nu-d} q(\mathbf{z})$ for $\mathbf{z} \neq \mathbf{0}$. Moreover, if $V(t) = \mathbb{P}(\|\mathbf{Z}\| > t)$ then $\mu(E) = \int_E q(\mathbf{z}) d\mathbf{z}$, where $\mu(\cdot)$ is the measure in (2.4).

Remark 2.2.1. Condition (2.5) says that the density f is a *multivariate regularly varying* function with index $(\nu+d)$ and limit function q (Stam [1977]). However, this condition alone is not sufficient to ensure multivariate regular variation of the tail of the distribution as it only controls the behaviour along rays. The uniformity condition in (2.6) then guarantees (2.2).

Existence of density q of μ implies the existence of density ψ of Ψ . In particular, one can show that

$$q(\mathbf{z}) = q(r\mathbf{w}) = \nu r^{-\nu-d} \psi(\mathbf{w}), \quad \mathbf{w} = \mathbf{z}/\|\mathbf{z}\| \in \mathcal{S}^{d-1}, \quad r = \|\mathbf{z}\|. \quad (2.7)$$

Proposition 2.2.2. *(Karamata's Theorem)*

If $\nu > 1$ (or if $\nu = 1$ and $\int_t^\infty h(u) du < \infty$), then $h \in RV_{-\nu}$ implies that $\int_t^\infty h(u) du$ is finite, $\int_t^\infty h(u) du \in RV_{-\nu+1}$, and

$$\lim_{t \rightarrow \infty} \frac{th(t)}{\int_t^\infty h(u) du} = \nu - 1.$$

2.2.2 Link to MVEVD

The multivariate regular variation can be linked to MVEVD discussed in Section 2.1.2 in the following way.

Proposition 2.2.3. *(Resnick [1987], Corollary 5.18) Suppose that \mathbf{Z} , a sequence of i.i.d. d -dimensional random vector, is regularly varying with index $\nu > 0$, and that $\mathbf{M}_n = (\max_{i=1}^n(Z_{i,1}), \dots, \max_{i=1}^n(Z_{i,d}))^T$ is the coordinate-wise maxima of \mathbf{Z} , then there exist normalizing sequences of vectors $\mathbf{a}_n > \mathbf{0}$ and $\mathbf{b}_n \in \mathbb{R}^d$ such that Formula 2.1 holds with*

$$H(\mathbf{x}) = \exp\{-\mu([\mathbf{0}, \mathbf{x}]^c)\},$$

where the marginal distributions of $H(\mathbf{x})$ are GEV with tail index ν .

The above Corollary basically claims that if \mathbf{Z} is multivariate regularly varying, then its coordinate-wise maximum has a MVEVD with Fréchet marginal distributions and dependence structure described by the measure μ .

2.3 Inference

Two important components in modelling multivariate tail behaviour of regularly varying random vectors are tail index ν of the radial component and tail dependence structure, which can be modelled by the spectral measure Ψ . In this section, we review several existing methods to perform inference.

2.3.1 Tail Index Estimators

As we assume a heavy-tailed behaviour for the financial return data, we need to estimate the tail index ν . Under the assumption that each variable of a random vector \mathbf{Z} has the same tail index ν , its L_2 -norm $\|\mathbf{Z}\|$ has a regularly varying tail with index ν . This reduces the multivariate problem to a univariate problem. In this section, we briefly review a few methods that can be applied to estimate the tail index ν . This review is strongly influenced by the master's thesis of Trudel [2008], and I rephrased and reordered several parts to fit in this project.

Hill Estimator For a random variable Z , Hill [1975] derived the maximum-likelihood estimator for the tail index ν given by

$$\hat{\nu} = \left[\frac{1}{k} \sum_{i=1}^k \log \frac{Z_{(i)}}{Z_{(k)}} \right]^{-1}, \quad (2.8)$$

where k is the threshold that defines the tail of the distribution, and $Z_{(1)} \geq Z_{(2)} \geq \dots \geq Z_{(n)}$ are the order statistics for the sample of i.i.d. copies of random variable Z , and n is the number of observations. This estimator is asymptotically unbiased and easy to implement, but is biased for small samples.

CSN Estimator Clauset et al. [2009] proposed another way to estimate the tail index, which relaxes the assumption that observations should be i.i.d. in the maximum-likelihood method. This method relies on the minimization of the Kolmogorov-Smirnov statistic between the empirical distribution of the data and the assumed heavy-tailed distribution. Intuitively, the CSN estimator intends to find the point where the change between the empirical distribution and the assumed distribution will most likely happen. The estimator is given by

$$\hat{\nu} = \arg \min_{\nu} D_{\nu},$$

$$\text{where } D_{\nu} = \max_x |\mathbb{P}_{emp}(x) - \mathbb{P}_{\nu}(x)|;$$

$\mathbb{P}_{emp}(x)$ and $\mathbb{P}_{\nu}(x)$ denote the empirical (cumulative) distribution function and the assumed heavy-tailed distribution function with tail index ν , respectively.

HKKP Estimator The drawback of Hill estimator is the difficulty to choose a suitable tail threshold number k such that the MSE is minimized. Practically, the selection of k normally depends on experience. Huisman et al. [2001] developed an ordinary linear regression (OLS) tail estimator to correct the bias generated by Hill estimator for small samples. This method does not simply rely on one single threshold k to estimate the tail index, but exploits the information contained in a series of Hill estimators, each estimated using different tail thresholds.

The HKKP Estimator utilizes an important characteristic of the bias term; that is, when the threshold k is small enough, the bias term can be approximated by a linear function. This suggests that a linear regression or OLS can help estimate the tail index. The linear model adopted by Huisman et al. [2001] is stated as follows:

$$\frac{1}{\nu(k)} = \beta_0 + \beta_1 k + \epsilon(k), \quad k = 1, 2, \dots, \kappa, \quad (2.9)$$

where κ is chosen such that the inverse of function $\nu(k)$ for $k = 1, 2, \dots, \kappa$ is approximately linear, and Huisman et al. [2001] showed that the estimation for tail index is robust to the choice of κ .

Furthermore, to correct for the heteroscedasticity of the error term $\epsilon(k)$, a weighted least square (WLS) regression can be adopted here. Since the variance of Hill estimator is inversely related to k , a $(\kappa \times \kappa)$ weighting matrix that has $\{\sqrt{1}, \sqrt{2}, \dots, \sqrt{\kappa}\}$ as diagonal elements and zeros elsewhere can be chosen. Since an unbiased estimator of ν can be attained for k approaching 0, Equation 2.9 yields the estimation of ν equal to the inverse of estimated intercept β_0 .

BGST Estimator Another popular method to estimate tail index is to run the following OLS log-log rank regression with $\gamma = 0$ (see Beirlant et al. [2006])

$$\log(k - \gamma) = \beta_0 - \beta_1 \log Z_{(i)}, \quad k = 1, 2, \dots, n. \quad (2.10)$$

The estimate of β_1 is the approximation for tail index ν . This method is motivated by the approximate linear relationship of the distribution with a heavy tail

$$\log\left(\frac{k}{n}\right) \approx \log(C) - \nu \log Z_k, \quad k = 1, 2, \dots, n. \quad (2.11)$$

Gabaix and Ibragimov [2011] provide an improved tail index estimate for this method, which simply sets γ to be $\frac{1}{2}$ instead of 0. In this way, the bias due to small samples can be reduced substantially.

2.3.2 Spectral Measure

As discussed above, the spectral measure Ψ in (2.2) can be used to describe the extreme dependence structure of a regularly varying random vector. A major issue here is the estimation of the spectral measure Ψ . Firstly, it is hard to tell where the tail begins in a given sample so that the asymptotic approximation is sufficiently accurate. Secondly, there are simply very few extreme observations. In this section, we summarize two non-parametric methods to estimate the spectral measure. To simplify notation, we only consider a random sample from a bivariate distribution. Note that if the distribution is known, the spectral measure can be obtained via Definition 2.2.3 or Theorem 2.2.1, at least numerically.

ES Estimator Einmahl and Segers [2009] proposed a non-parametric maximum empirical likelihood estimator for the spectral measure Ψ . The estimator is based on moment condition as below:

$$\int_{[0, \pi/2]} f(\theta) \Psi(d\theta) = 0, \quad (2.12)$$

$$\text{where } f(\theta) = \frac{\sin \theta - \cos \theta}{\|\sin \theta - \cos \theta\|}, \quad \theta \in [0, \pi/2]. \quad (2.13)$$

Let $\{(Z_{i,1}, Z_{i,2}); i = 1, \dots, n\}$ be a bivariate random sample with n observations. The empirical marginal distribution function can be written as

$$\hat{F}_j(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(Z_{i,j} \leq x), \quad x \in \mathbb{R}, j = 1, 2$$

and $\hat{\bar{F}}_{i,j} = 1 - \hat{F}_j(x_{i,j})$ is defined as the upper tail empirical probability. Then the consistent empirical spectral probability measure $\hat{\Psi}$ is

$$\hat{\Psi}(\cdot) = \sum_{i \in I_n} \tilde{p}_{i,n} \mathbf{1}(\theta_{i,n} \in \cdot),$$

$$\text{where } \theta_{in} = \arctan\left(\frac{\hat{\bar{F}}_{i,2}}{\hat{\bar{F}}_{i,1}}\right), i = 1, 2, \dots, n,$$

$$I_n = \{i = 1, 2, \dots, n : \|(\hat{\bar{F}}_{i,1}^{-1}, \hat{\bar{F}}_{i,2}^{-1})\| \geq n/k\},$$

2.3. Inference

and the weight vector $(\tilde{p}_{i,n} : i \in I_n)$ solves the following optimization problem:

$$\begin{aligned} & \text{maximize} && \prod_i p_{i,n}, \\ & \text{constraints} && p_{i,n} \geq 0 \quad \text{for all } i \in I_n, \\ & && \sum_i p_{i,n} = 1, \\ & && \sum_i p_{i,n} f(\theta_{i,n}) = 0. \end{aligned}$$

To prove asymptotic results of the estimator, the condition that $k = k(n)$ is an intermediate sequence satisfying the condition that $k \rightarrow \infty$ and $k/n \rightarrow 0$ as $n \rightarrow \infty$ is required.

This method makes no assumptions on the marginal distribution functions, and allows for arbitrary norms when defining the actual representation of the spectral measure. As then moment condition is set to be a constraint, it succeeds to overcome the shortcomings of the empirical spectral measure proposed in Einmahl et al. [2001] that is itself not a proper spectral measure due to violation of moment constraints.

NS Estimator One drawback of the ES estimator is the difficulty to choose the tail threshold k . Nguyen and Samorodnitsky [2013] proposed a method that allows for systematic decision on what part of the sample corresponds to “tail observations”. This method is based on the rank method, where the rank statistics are defined as

$$r_i^{(j)} = \sum_{m=1}^n \mathbf{1}(Z_{m,j} \geq Z_{i,j}), \quad 1 \leq i \leq n, j = 1, 2.$$

To estimate the spectral measure, the data are firstly transformed by polar transformation as

$$(R_{i,k}, \theta_{i,k}) = \left(\left\| \left(\frac{k}{r_i^{(1)}}, \frac{k}{r_i^{(2)}} \right) \right\|, \frac{\left(\frac{k}{r_i^{(1)}}, \frac{k}{r_i^{(2)}} \right)}{\left\| \left(\frac{k}{r_i^{(1)}}, \frac{k}{r_i^{(2)}} \right) \right\|} \right), \quad i = 1, 2, \dots, n.$$

2.3. Inference

Then, the consistent estimator for Ψ based on Equation 2.2 can be written as

$$\hat{\Psi}(\cdot) = \frac{\sum_{i=1}^n \mathbf{1}(R_{i,k} > 1, \theta_{i,k} \in \cdot)}{\sum_{i=1}^n \mathbf{1}(R_{i,k} > 1, \theta_{i,k} \in [0, \pi/2])}.$$

Their choice of k used in tail estimation depends on the test for exponentiality of “tail part” of the data. Specifically, $k = \min(N_n^{(1)}, N_n^{(2)})$, where $N_n^{(j)}$ is the smallest k such that the null hypothesis of exponentiality is rejected for each marginal variable. The statistic used to test for exponentiality is:

$$Q_{k,j} = \frac{\sqrt{k}}{2} \left(\frac{\frac{1}{k} \sum_{i=0}^{k-1} \left(\log \frac{Z_{n-i,j}}{Z_{n-k,j}} \right)^2}{\left(\frac{1}{k} \sum_{i=0}^{k-1} \log \frac{Z_{n-i,j}}{Z_{n-k,j}} \right)^2} - 2 \right), \quad j = 1, 2,$$

which converges to the standard normal distribution under the null hypothesis that the variable is exponentially distributed (see Dahiya and Gurland [1972]). Their choice of the significance level of the test gives

$$N_n^{(j)} = \inf \left\{ k : 1 \leq k \leq n, |Q_{k,n}| \geq \omega_j \sqrt{\frac{\theta_n^{(j)}}{k}} \right\}, \quad j = 1, 2,$$

where $\omega_j > 0$ and $\theta_n^{(j)}$ is an increasing sequence such that $\theta_n^{(j)} = o(n^{\frac{2\nu}{1+2\nu}})$.

Their estimate is conservative about deciding on where the “tail” starts. Other choices are possible catering to different needs, for example, $k = \max(N_n^{(1)}, N_n^{(2)})$. This method is easily extended to higher dimension of random vector \mathbf{Z} , and is fast and simple to automate.

Other methods to estimate spectral measure include Guillotte et al. [2011], Eastoe et al. [2014], among others. All these non-parametric methods are flexible to capture the behaviour of tail observations. However, these methods suffer from a common disadvantage of non-parametric methods, that is, there is no parameter to make quantitative statements about populations. In this project, the objective is to derive a semi-parametric estimator for spectral measure, which is flexible enough to describe different tail behaviour.

Chapter 3

Review of Skewed Distributions

3.1 Elliptical Distributions

Elliptical distributions are often used as a starting point for further generalizations beyond elliptical symmetry. Here we give a definition, establish notation and some useful facts for future reference. A complete treatment of the topic may be found in Fang et al. [1990].

Definition 3.1.1. *A continuous random vector \mathbf{Z} on \mathbb{R}^d is said to have an elliptical distribution if its density is of the form*

$$f(\mathbf{z}; \boldsymbol{\xi}, \Omega) = \frac{c_d}{|\Omega|^{1/2}} \tilde{f}\{(\mathbf{z} - \boldsymbol{\xi})^T \Omega^{-1} (\mathbf{z} - \boldsymbol{\xi})\}, \quad \mathbf{z} \in \mathbb{R}^d, \quad (3.1)$$

where $\boldsymbol{\xi} \in \mathbb{R}^d$ is a location parameter, $\Omega \in \mathbb{R}^{d \times d}$ is a positive-definite scale matrix, $\tilde{f} : [0, \infty) \rightarrow [0, \infty)$ is a continuous and integrable decreasing function, known as the density generator, and c_d is a normalizing constant. We write $\mathbf{Z} \sim \text{Ell}_d(\boldsymbol{\xi}, \Omega, \tilde{f})$.

The following stochastic representation for elliptically distributed random vectors is often useful. If $\mathbf{Z} \sim \text{Ell}_d(\boldsymbol{\xi}, \Omega, \tilde{f})$, then \mathbf{Z} can be written as

$$\mathbf{Z} = \boldsymbol{\xi} + RL^T S, \quad (3.2)$$

where $L^T L = \Omega$, S is a uniformly distributed random vector on $\mathcal{S}^{d-1} = \{\mathbf{z} \in \mathbb{R}^d : \|\mathbf{z}\| = 1\}$, the unit sphere in \mathbb{R}^d , and R is a positive scalar random variable independent of S .

It will sometimes be convenient to work with a standardized form of the scale matrix. Given a scale matrix $\Omega = (\omega_{ij}) \in \mathbb{R}^{d \times d}$, define the associated *standardized scale matrix* $\bar{\Omega}$ as $\bar{\Omega} = \omega^{-1} \Omega \omega^{-1}$, where

$$\omega = \text{diag}(\omega_1, \dots, \omega_d) = \text{diag}(\sqrt{\omega_{11}}, \dots, \sqrt{\omega_{dd}}). \quad (3.3)$$

3.2 Skew-symmetric Distributions

In this section we describe a very general class of skewed distributions. The construction is simple and is summarized in Proposition 3.2.1 below. Let \mathbf{Z} be a d -dimensional random vector which is *centrally symmetric* around a point $\boldsymbol{\xi} \in \mathbb{R}^d$; i.e. $\mathbf{Z} - \boldsymbol{\xi} \stackrel{d}{=} \boldsymbol{\xi} - \mathbf{Z}$. Assuming that \mathbf{Z} is continuous with density f , the condition of central symmetry implies that $f(\mathbf{z} - \boldsymbol{\xi}) = f(\boldsymbol{\xi} - \mathbf{z})$ for all $\mathbf{z} \in \mathbb{R}^d$, up to a set of measure zero. Note that the property of central symmetry is satisfied by a wide variety of densities including for example the class of elliptical densities. Skew-symmetric distributions as introduced in Wang et al. [2004] are then constructed by perturbing the symmetry of the density f of \mathbf{Z} by a so-called skewing function. A function $\pi : \mathbb{R}^d \rightarrow [0, 1]$ is a *skewing function* if $\pi(-\mathbf{z}) = 1 - \pi(\mathbf{z})$ for $\mathbf{z} \in \mathbb{R}^d$.

Proposition 3.2.1. *Suppose f is the density of a continuous random vector on \mathbb{R}^d which is centrally symmetric around $\mathbf{0}$, and $\pi : \mathbb{R}^d \rightarrow [0, 1]$ is a skewing function. Let G denote a scalar distribution function of a random variable that is symmetric about 0 such that $G(-s) = 1 - G(s)$ for all $s \in \mathbb{R}$ and let $w : \mathbb{R}^d \rightarrow \mathbb{R}$ be an odd function with $w(-\mathbf{x}) = -w(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^d$. Then the function from $\mathbb{R}^d \rightarrow \mathbb{R}_+$*

$$2f(\mathbf{z} - \boldsymbol{\xi})\pi(\mathbf{z} - \boldsymbol{\xi}), \quad \mathbf{z}, \boldsymbol{\xi} \in \mathbb{R}^d \quad (3.4)$$

or equivalently,

$$2f(\mathbf{z} - \boldsymbol{\xi})G(w(\mathbf{z} - \boldsymbol{\xi})), \quad \mathbf{z}, \boldsymbol{\xi} \in \mathbb{R}^d \quad (3.5)$$

is a density.

Proof A proof is given in Proposition 1 of Wang et al. [2004] and Azzalini and Capitanio [2003]. Equivalence of the two formulations (3.4) and (3.5) is shown in Proposition 2, Wang et al. [2004]. \blacksquare

3.2. Skew-symmetric Distributions

As remarked in Wang et al. [2004], representation (3.5) of a skewing function is not unique since for any strictly increasing distribution function G , one can find a suitable odd function w to obtain a given skewing function π .

Definition 3.2.1. *A random vector \mathbf{Z} on \mathbb{R}^d has a skew-symmetric distribution with location parameter $\boldsymbol{\xi} \in \mathbb{R}^d$ if its density is of the form (3.4) or (3.5).*

For a skew-symmetric random vector \mathbf{Z} with density (3.5), there exists a convenient stochastic representation:

$$\mathbf{Z} = \begin{cases} \mathbf{Y} + \boldsymbol{\xi} & \text{if } U < \pi(\mathbf{Y}) \text{ or } X < w(\mathbf{Y} + \boldsymbol{\xi}) \\ -\mathbf{Y} + \boldsymbol{\xi} & \text{if } U > \pi(\mathbf{Y}) \text{ or } X > w(\mathbf{Y} + \boldsymbol{\xi}) \end{cases}, \quad (3.6)$$

where \mathbf{Y} has density f , U is uniformly distributed on $(0, 1)$, and X is a random variable with distribution function G ; \mathbf{Y} , U and X are mutually independent. This representation gives a straightforward way to simulate from skew-symmetric distributions.

An interesting and in fact useful property of skew-symmetric distributions is their invariance under even functions, which is immediate from the stochastic representation in (3.6).

Proposition 3.2.2. *Consider random vectors \mathbf{Y} and \mathbf{Z} on \mathbb{R}^d with densities f and of the form (3.4) with $\boldsymbol{\xi} = \mathbf{0}$, respectively, satisfying the conditions of Proposition 3.2.1. If $\tau : \mathbb{R}^d \rightarrow \mathbb{R}^p$ for some $p > 0$ is an even function, i.e. $\tau(-\mathbf{x}) = \tau(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^d$, then*

$$\tau(\mathbf{Y}) \stackrel{d}{=} \tau(\mathbf{Z}).$$

The complete generality of the class of skew-symmetric distributions is shown in Wang et al. [2004] by proving that in fact any density has a skew-symmetric representation and this representation is unique:

Proposition 3.2.3. *For any density g on \mathbb{R}^d and any point $\boldsymbol{\xi} \in \mathbb{R}^d$, g can be represented as*

$$g(\mathbf{z}) = 2f(\mathbf{z} - \boldsymbol{\xi})\pi(\mathbf{z} - \boldsymbol{\xi}), \quad (3.7)$$

3.3. Skew-elliptical Distributions

where f is a density, centrally symmetric around $\mathbf{0}$ and π is a skewing function. This representation is unique for any $\boldsymbol{\xi}$, and

$$f(\mathbf{z}) = \frac{g(\boldsymbol{\xi} + \mathbf{z}) + g(\boldsymbol{\xi} - \mathbf{z})}{2} \quad \text{and} \quad \pi(\mathbf{z}) = \frac{g(\boldsymbol{\xi} + \mathbf{z})}{g(\boldsymbol{\xi} + \mathbf{z}) + g(\boldsymbol{\xi} - \mathbf{z})}.$$

Next we would like to take a closer look at several important examples of this very general family of skewed distributions.

3.3 Skew-elliptical Distributions

Skew-elliptical distributions constitute a fairly large subclass within the family of skew-symmetric distributions, and are obtained by replacing the centrally symmetric density f in (3.5) with a density of an elliptical distribution (see Definition 3.1.1).

Definition 3.3.1. *A random vector \mathbf{Z} on \mathbb{R}^d has a skew-elliptical distribution if its density is given by*

$$h(\mathbf{z}) = 2 f(\mathbf{z}; \boldsymbol{\xi}, \Omega) G(w(\mathbf{z} - \boldsymbol{\xi})), \quad \mathbf{z} \in \mathbb{R}^d, \quad (3.8)$$

where:

- $\boldsymbol{\xi} \in \mathbb{R}^d$ is a location parameter and $\Omega \in \mathbb{R}^{d \times d}$ is a positive-definite scale matrix;
- f is the elliptical density in (3.1) with density generator \tilde{f} ;
- G is a scalar distribution function such that $G(-x) = 1 - G(x)$ for all $x \in \mathbb{R}$;
- $w : \mathbb{R}^d \rightarrow \mathbb{R}$ is an odd function; i.e., $w(-\mathbf{x}) = -w(\mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}^d$.

We write $\mathbf{Z} \sim SE_d(\boldsymbol{\xi}, \Omega, \tilde{f}, G \circ w)$.

Azzalini and Capitanio [2003] have derived a stochastic representation for skew-elliptical random vectors similar to the one for elliptical vectors presented in (3.2).

3.3. Skew-elliptical Distributions

Proposition 3.3.1. *If $\mathbf{Z} \sim SE_d(\boldsymbol{\xi}, \bar{\Omega}, \tilde{f}, G \circ w)$ as in Definition 3.3.1, then it admits the following stochastic representation*

$$\mathbf{Z} = \boldsymbol{\xi} + RL^T S', \quad (3.9)$$

where $\bar{\Omega} = L^T L$; $R > 0$ has density

$$f_R(r) = \frac{2\pi^{d/2}}{\Gamma(d/2)} c_d \tilde{f}(r^2) r^{d-1}, \quad r > 0, \quad (3.10)$$

and S' has a non-uniform distribution on the unit sphere in \mathbb{R}^d . The density of S' in spherical coordinates is given by

$$\frac{\Gamma(d/2)}{\pi^{d/2}} \prod_{k=1}^{d-2} (\sin \theta_k)^{d-k-1} \mathbb{P}\{X \leq w_L^*(\theta_1, \dots, \theta_{d-1}, R)\},$$

$$\boldsymbol{\theta} = (\theta_1, \dots, \theta_{d-1}) \in [0, \pi)^{d-2} \times [0, 2\pi),$$

where $w_L^*(\boldsymbol{\theta}, r) = w_L(r \cos \theta_1, r \sin \theta_1 \cos \theta_2, \dots, r \sin \theta_1 \cdots \sin \theta_{d-1})$, $w_L(\mathbf{x}) = w(L^T \mathbf{x})$, and X is an independent random variable with distribution function G .

Lemma 3.3.2 (Linear transformation of SE). *Let $\mathbf{Z} \sim SE_d(\boldsymbol{\xi}, \Omega, \tilde{f}, G \circ w)$, then $C\mathbf{Z} \sim SE_d(C\boldsymbol{\xi}, C^T \Omega C, \tilde{f}, G \circ \omega_0)$ for any $k \times k$ non-singular matrix C , where $\omega_0(\mathbf{z} - C\boldsymbol{\xi}) = \omega(C^{-1}(\mathbf{z} - C\boldsymbol{\xi}))$.*

We next look at the form of this representation for two specific examples of skew-elliptical distributions.

3.3.1 Skew-normal Distribution

The density of random vector \mathbf{Z} on \mathbb{R}^d with a *skew-normal distribution* (denoted $\mathbf{Z} \sim SN_d(\boldsymbol{\xi}, \Omega, \boldsymbol{\alpha})$) is given by

$$f_{SN}(\mathbf{z}) = 2\phi_d(\mathbf{z} - \boldsymbol{\xi}; \Omega) \Phi(\boldsymbol{\alpha}^T(\mathbf{z} - \boldsymbol{\xi})), \quad \mathbf{z} \in \mathbb{R}^d, \quad (3.11)$$

where $\boldsymbol{\alpha} \in \mathbb{R}^d$ is the shape parameter controlling the skewness of the distribution; $\phi_d(\cdot; \Omega)$ is the density of the centred d -variate normal distribution

3.3. Skew-elliptical Distributions

with scale matrix Ω and $\Phi(\cdot)$ is the distribution function of the standard normal random variable. Comparing f_{SN} in (3.11) with representation (3.5), we see that $G(\cdot) = \Phi(\cdot)$ and $w(\mathbf{y}) = \boldsymbol{\alpha}^T \mathbf{y}$, $\mathbf{y} \in \mathbb{R}^d$. Hence, for $\mathbf{Z} \sim SN_d(\boldsymbol{\xi}, \bar{\Omega}, \boldsymbol{\alpha})$, Proposition 3.3.1 says that $\mathbf{Z} = \boldsymbol{\xi} + RL^T S'$ with $R^2 \sim \chi_d^2$ and the density of S' in spherical coordinates equal to

$$f_{\Theta}(\boldsymbol{\theta}) = \frac{\Gamma(d/2)}{\pi^{d/2}} \prod_{k=1}^{d-2} (\sin \theta_k)^{d-k-1} \mathbb{P}\{X \leq R(\boldsymbol{\alpha}^*)^T \boldsymbol{\beta}_{\theta}\}, \quad (3.12)$$

$$\boldsymbol{\theta} = (\theta_1, \dots, \theta_{d-1}) \in [0, \pi)^{d-2} \times [0, 2\pi),$$

where $\boldsymbol{\alpha}^* = L\boldsymbol{\alpha}$, $\boldsymbol{\beta}_{\theta} = (\cos \theta_1, \sin \theta_1 \cos \theta_2, \dots, \sin \theta_1 \cdots \sin \theta_{d-1})^T$ and $X \sim N(0, 1)$, independent of R . Using the fact that $X/\sqrt{R^2/d}$ has a Student t distribution with ν degrees of freedom and letting $T_1(\cdot; \nu)$ denote its distribution function, we obtain

$$f_{\Theta}(\boldsymbol{\theta}) = \frac{\Gamma(d/2)}{\pi^{d/2}} \prod_{k=1}^{d-2} (\sin \theta_k)^{d-k-1} T_1(\sqrt{d}(\boldsymbol{\alpha}^*)^T \boldsymbol{\beta}_{\theta}; \nu). \quad (3.13)$$

3.3.2 Skew- t Distribution

Following Azzalini and Capitanio [2003], we say that a d -dimensional vector \mathbf{Z} has a *skew- t distribution* with location parameter $\boldsymbol{\xi}$, scale matrix Ω , shape parameter $\boldsymbol{\alpha}$ and ν degrees of freedom (written as $\mathbf{Z} \sim St_d(\boldsymbol{\xi}, \Omega, \boldsymbol{\alpha}, \nu)$) if its density is given by

$$f_{St}(\mathbf{z}) = 2t_d(\mathbf{z}; \Omega, \nu) T_1\left(\boldsymbol{\alpha}^T \omega^{-1}(\mathbf{z} - \boldsymbol{\xi}) \left(\frac{\nu + d}{Q(\mathbf{z}) + \nu}\right)^{1/2}; \nu + d\right), \quad \mathbf{z} \in \mathbb{R}^d, \quad (3.14)$$

where $Q(\mathbf{z}) = (\mathbf{z} - \boldsymbol{\xi})^T \Omega^{-1}(\mathbf{z} - \boldsymbol{\xi})$, and $t_d(\mathbf{z}; \Omega, \nu)$ is the density of a standardized d -dimensional t distributed random vector with scale matrix Ω and ν degrees of freedom.

Application of Proposition 3.3.1 to $\mathbf{Z} \sim St_d(\boldsymbol{\xi}, \bar{\Omega}, \boldsymbol{\alpha}, \nu)$ gives that in the stochastic representation (3.9), the density of S' in spherical coordinates is equal to

$$f_{\Theta}(\boldsymbol{\theta}) = \frac{\Gamma(d/2)}{\pi^{d/2}} \prod_{k=1}^{d-2} (\sin \theta_k)^{d-k-1} \mathbb{P}\left\{X \leq R \sqrt{\frac{\nu + d}{R^2 + \nu}} (\boldsymbol{\alpha}^*)^T \boldsymbol{\beta}_{\theta}\right\}, \quad (3.15)$$

3.3. Skew-elliptical Distributions

where $\boldsymbol{\alpha}^*$ and β_θ are as in (3.12), the density of R is given by

$$f_R(r) \propto (1 + r^2/\nu)^{-(\nu+d)/2} r^{d-1}, \quad r > 0,$$

and X has a Student t distribution with $\nu + d$ degrees of freedom and is independent of R . Let $Y = (X/R)\sqrt{(R^2 + \nu)/(\nu + d)}$. The density of Y can be computed as follows

$$f_Y(y) = \int_0^\infty f_{Y|R}(y | r) f_R(r) dr \propto \int_0^\infty r^d (\nu + r^2(1 + y^2))^{-(\nu+d+1)/2} dr;$$

making change of variable $t = r\sqrt{1 + y^2}$ gives

$$f_Y(y) \propto (1 + y^2)^{-(d+1)/2} \int_0^\infty t^d (\nu + t^2)^{-(\nu+d+1)/2} dt \propto (1 + y^2)^{-(d+1)/2}, \quad y \in \mathbb{R},$$

from which one recognizes f_Y as the density of a t distribution with d degrees of freedom evaluated at the point $\sqrt{d}y$. So, $(X/R)\sqrt{d(R^2 + \nu)/(\nu + d)}$ has a t distribution with d degrees of freedom, which implies that the density of S' in (3.15) for the skew- t distribution is exactly the same as in (3.13) for the skew-normal distribution. This result reveals similarity between the skew-normal and skew- t distributions via the stochastic behaviour of the angular component S' ; the differences arise due to the tail behaviour as governed by the radial component R , and the dependence between S' and R .

In a similar way, it can be shown that the same result holds for a more general class of skew-elliptical densities (of which skew- t is a special case) based on the multivariate Pearson type VII distributions, whose generator and normalizing constant are

$$\tilde{f}(x) = (1 + x/\nu)^{-M}, \quad c_d = \frac{\Gamma(M)}{(\pi\nu)^{d/2} \Gamma(M - d/2)}, \quad \nu > 0, \quad M > d/2.$$

Proposition 4 in Azzalini and Capitanio [2003] says that the density of the above skew-elliptical distributions is of the type (3.5) and is given by

$$2f(\mathbf{z}; \bar{\Omega}) F_1(\boldsymbol{\alpha}^T \mathbf{z} (\nu + Q(\mathbf{z}))^{-1/2}; M, 1), \quad \mathbf{z} \in \mathbb{R}^d,$$

where f is the density of $PVII_d(\mathbf{0}, \bar{\Omega}, M - 1/2, \nu)$ distribution and $F_1(\cdot; M, 1)$ is the distribution function of a $PVII_1(0, 1, M, 1)$ distributed random variable. One recovers the skew- t case by setting $M = (\nu + d + 1)/2$.

3.3. *Skew-elliptical Distributions*

An interesting question is whether there are other examples of skew-elliptical distributions for which the density of S' in (3.9) is equal to (3.13), and if so what conditions on the family of skew-elliptical distributions will ensure this property.

Chapter 4

Theoretical Results

In this chapter we explore the asymptotic behaviour of skew-elliptical distributions under certain restrictions on the skewing function. Specifically, we derived a limit expression for the conditional probability given that one of the components of a bivariate random vector is extreme, assuming a sub-family of skew-elliptical distributions with regularly varying tails as the underlying model. This extends results in Abdous et al. [2005a] for elliptically symmetric vectors. All the proofs can be found in Appendix A.

4.1 Limit Results

The following results on the asymptotic behaviour of random vector \mathbf{Z} will be used to motivate and justify the proposed inference procedure for estimation of the conditional probability $\eta(x, y)$ in (1.1).

Proposition 4.1.1. *Let $\mathbf{Z} = (X, Y) \sim SE_2(\boldsymbol{\xi}, \bar{\Omega}, \tilde{f}, G \circ w)$, where $\bar{\Omega}_{ii} = 1$ and $\bar{\Omega}_{ij} = \rho \in (-1, 1)$ for $i \neq j$, $i, j \in \{1, 2\}$. Assume the following:*

- (i) *The density generator \tilde{f} of the underlying elliptical distribution varies regularly: $\tilde{f} \in RV_{-\frac{\nu+2}{2}}$ for some $\nu > 0$;*
- (ii) *$\lim_{t \rightarrow \infty} w(t\mathbf{z}) =: w_\infty(\mathbf{z}) \in \mathbb{R}$ for all $\mathbf{z} \in \mathbb{R}^2$, and $\lim_{t \rightarrow \infty} \sup_{\mathbf{z} \in \mathcal{S}^1} |G(w(t\mathbf{z})) - G(w_\infty(\mathbf{z}))| = 0$.*

Then \mathbf{Z} is a bivariate regularly varying random vector with index ν . The density of the associated spectral measure Ψ in (2.2) is given by

$$\psi(\mathbf{w}) = 2 |\bar{\Omega}|^{-(1/2)} G(w_\infty(\mathbf{w})) Q^*(\mathbf{w})^{-(\nu+2)/2} \left[\int_0^{2\pi} A(\theta)^{\nu/2} d\theta \right]^{-1}, \quad \mathbf{w} \in \mathcal{S}^1, \quad (4.1)$$

4.1. Limit Results

where $A(\theta) = 1 + \rho\sqrt{1 - \rho^2} \sin(2\theta) + \rho^2 \cos(2\theta)$, and $Q^*(\mathbf{w}) = \mathbf{w}^T \bar{\Omega}^{-1} \mathbf{w}$.

Remark 4.1.1. Assumption (i) above is standard, and assumption (ii) holds for popular members of the skew-elliptical family of distributions, including multivariate skew- t distribution and more generally the skew-Pearson type VII distribution, both mentioned in Chapter 3.3. The first part of assumption (ii) indicates that the function w is bounded in all directions, and the second part guarantees the uniformity condition (2.6) for density function of \mathbf{Z} .

Lemma 4.1.2. Let $\mathbf{Z} = \boldsymbol{\xi} + RL^T S' \sim SE_d(\boldsymbol{\xi}, \Omega, \tilde{f}, G)$. Then the following statements are equivalent:

(i) $\tilde{f} \in RV_{-(\nu+d)/2}$;

(ii) R is regularly varying with index ν ;

(iii) $\|\mathbf{Z}\|$ is regularly varying with index ν , where $\|\cdot\|$ is the L_2 norm.

The next result gives the limit expression for the conditional probability in (1.1).

Theorem 4.1.3. Let \mathbf{Z} have a bivariate skew-elliptical distribution satisfying assumptions of Proposition 4.1.1. Suppose the marginal densities of \mathbf{Z} have the form:

$$h_i(z) = 2 f_i(z - \xi_i) G_0(\bar{w}_i(z - \xi_i)), \quad i = 1, 2, \quad (4.2)$$

where $f_i(\cdot)$ is the symmetric density of the i th component of elliptical vector $\tilde{\mathbf{Z}}$ and $G_0(\bar{w}_i(\cdot))$ is a skewing function with existing limit $\lim_{t \rightarrow \infty} \bar{w}_i(tx) = \bar{w}_{i,\infty}(x) \in \mathbb{R}$ for all $x \in \mathbb{R}$, $i = 1, 2$. Then, for $y \sim (z + \rho)x$ as $x \rightarrow \infty$ and $z \in \mathbb{R}$,

$$\lim_{x, y \rightarrow \infty} \mathbb{P}(Y \leq y \mid X > x) = 1 - K(\rho + z) J(z), \quad (4.3)$$

where:

$$\bullet K(z) = \frac{\iint_{(1,z)}^{\infty} Q^*(\mathbf{u})^{-(\nu+2)/2} G(w_{\infty}(\mathbf{u})) d\mathbf{u}}{G_0(\bar{w}_{1,\infty}(1)) \iint_{(1,z)}^{\infty} Q^*(\mathbf{u})^{-(\nu+2)/2} d\mathbf{u}} \quad \text{with } Q^*(\mathbf{x}) = \mathbf{x}^T \bar{\Omega}^{-1} \mathbf{x};$$

•

$$J(z) = \bar{T}_1\left(\frac{z\sqrt{\nu+1}}{\sqrt{1-\rho^2}}; \nu+1\right) + \frac{\text{sign}(\rho+z)}{|\rho+z|^\nu} \bar{T}_1\left(\text{sign}(\rho+z) \frac{\sqrt{\nu+1}}{\sqrt{1-\rho^2}} \left(\frac{1}{\rho+z} - \rho\right); \nu+1\right);$$

- $T_1(x; \nu)$ is the cdf of a univariate Student t random variable with ν degrees of freedom;
- $\bar{T}_1(x; \nu) = 1 - T_1(x; \nu)$.

Remark 4.1.2. When the skewing function is identity, the factor $K(z)$ in (A.9) is equal to one, and one recovers Theorem 1(i) in Abdous et al. [2005a] for elliptical random vectors.

Remark 4.1.3. The assumption of closure under marginalization is valid for some restricted generating functions and skewing functions; for example, the skew-elliptical distributions proposed by Branco and Dey [2001] and by Fang [2003]¹. Skew-normal, skew- t , and skewed Pearson Type II distributions are all included in this distribution family.

Example 1. Suppose \mathbf{Z} has a bivariate skew- t distribution, $\mathbf{Z} \sim St_2(\mathbf{0}, \bar{\Omega}, \boldsymbol{\alpha}, \nu)$. The marginals satisfy

$$\mathbb{P}(Z_i > z) \sim 2 \bar{T}_1(z; \nu) T_1(\bar{\alpha}_i \sqrt{\nu+1}, \nu+1), \quad z \rightarrow \infty, \quad i = 1, 2,$$

with

$$\bar{\alpha}_1 = \frac{\alpha_1 + \rho\alpha_2}{\sqrt{1 + \alpha_2^2(1 - \rho^2)}} \quad \text{and} \quad \bar{\alpha}_2 = \frac{\alpha_2 + \rho\alpha_1}{\sqrt{1 + \alpha_1^2(1 - \rho^2)}}. \quad (4.4)$$

For the skewing function we have $G(\cdot) = T_1(\cdot; \nu+2)$, $w(\mathbf{z}) = \boldsymbol{\alpha}^T \mathbf{z} \left(\frac{\nu+2}{Q(\mathbf{z}) + \nu}\right)^{1/2}$

and $w_\infty(\mathbf{z}) = \boldsymbol{\alpha}^T \mathbf{z} \left(\frac{\nu+2}{Q^*(\mathbf{z})}\right)^{1/2}$, and so the factor $K(z)$ in (A.9) is given by

$$K(z; \boldsymbol{\alpha}, \rho, \nu) = \frac{\iint_{(1,z)}^\infty Q^*(\mathbf{u})^{-(\nu+2)/2} T_1\left(\boldsymbol{\alpha}^T \mathbf{u} \sqrt{\frac{\nu+2}{Q^*(\mathbf{u})}}; \nu+2\right) d\mathbf{u}}{T_1(\bar{\alpha}_1 \sqrt{\nu+1}; \nu+1) \iint_{(1,z)}^\infty Q^*(\mathbf{u})^{-(\nu+2)/2} d\mathbf{u}}. \quad (4.5)$$

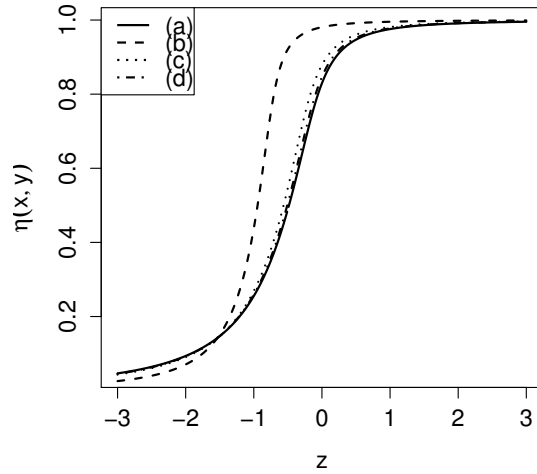
¹When $\lambda = 0$, skew-elliptical distribution in Fang [2003] is reduced to the form in Branco and Dey [2001].

4.1. Limit Results

In this case, the limit expression in (4.3) below can be evaluated numerically.

Figure 4.1 plots the limiting behaviours described in Theorem 4.1.3 for a bivariate skew- t distribution with shifted location. To show correctness of the derived formula, we firstly calculate the limiting value obtained from Theorem 4.1.3 (see (a)), and then calculate true values by numerical integration with different thresholds for x (see (b), (c), and (d)). As x increases, numerical values converge to the limiting value. Therefore, Theorem 4.1.3 works well even if the location parameter is not around $\mathbf{0}$.

Figure 4.1: Plot of the values of $\lim_{x \rightarrow \infty} \mathbb{P}(Y \leq (z + \rho)x \mid X > x)$ in terms of z for a bivariate skew- t distribution with $\boldsymbol{\xi} = (5, -5)$, $\boldsymbol{\alpha} = (1, -3)$, $\rho = 0.5$ and $\nu = 2$: a) limiting value from Theorem 4.1.3, and true value with x being (b) 99% marginal quantile, (c) 99.99% marginal quantile, and (d) 200.



Chapter 5

Inference

In practice, one often has to evaluate the probability $\eta(x, y)$ in (1.1) when x and y are large but finite. For example, $\eta(x, y)$ can represent the probability that a stock or a portfolio experiences a large loss when the market crashes. However, the empirical estimate of this probability may take a degenerate value of 0 or 1 whenever the sample contains no observations in the region of interest. The inference procedure we propose makes use of the asymptotic result in Theorem 4.1.3 as a finite sample approximation of the conditional probability $\eta(x, y)$ for large values of x and y . Evaluation of the limit in (4.3) requires estimation of the tail index ν , parameter ρ of the standardized scale matrix, and (parameters of) the asymptotic skewing function $G(w_\infty(\cdot))$. It is henceforth assumed that the bivariate sample $\{\mathbf{Z}_1, \dots, \mathbf{Z}_n\}$ on which inference is based comprises independent and identically distributed copies of random vector \mathbf{Z} satisfying assumptions of Theorem 4.1.3.

5.1 Tail Index Estimation

Under the assumptions of Theorem 4.1.3 on random vector \mathbf{Z} , its L_2 -norm $\|\mathbf{Z}\|$ has a regularly varying tail with index ν ; see Lemma 4.1.2. This reduces the problem of estimation of ν to a univariate setting, which has been widely studied in the literature. Similar to Abdous et al. [2005a], we adopt the method of Huisman et al. [2001], shown to have good performance in small samples. The estimation procedure is summarized in Program 1.

Table 5.1 re-examines performance of the above tail index estimator for two skewed distributions, in particular with the purpose to supplement the symmetric cases considered in the original paper of Huisman et al. [2001].

5.1. Tail Index Estimation

Algorithm 1 Estimation of tail index ν

- (1) Let $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$ represent a bivariate sample with location $\boldsymbol{\xi}$;
- (2) Re-center data $\tilde{\mathbf{z}}_j = \mathbf{z}_j - \hat{\boldsymbol{\xi}}$, $j = 1, \dots, n$, where $\hat{\boldsymbol{\xi}}$ is a robust estimate of $\boldsymbol{\xi}$;
- (3) Calculate L_2 -norm distances $\|\tilde{\mathbf{z}}_j\|$, $j = 1, \dots, n$;
- (4) Compute Hill estimates $\hat{\nu}_H(k)$ of the tail index for $k = 1, \dots, \frac{n}{2}$, where k is the number of upper order statistics used in estimation (Hill [1975]);
- (5) Fit a weighted linear regression $1/\hat{\nu}_H(k) = \beta_0 + \beta_1 k + \epsilon(k)$, $k = 1, \dots, \frac{n}{2}$ with weights $\{\sqrt{1}, \sqrt{2}, \dots, \sqrt{\frac{n}{2}}\}$;
- (6) Set $\hat{\nu} = 1/\hat{\beta}_0$, where $\hat{\beta}_0$ is the estimate of β_0 obtained in Step (5).

The analysis is based on 1000 simulated samples of sizes 5000, 1000, 500 and 200 from a bivariate skew- t distribution with different settings for the location parameter. As specified in Program 1, the generated samples are first re-centered by a robust estimate of location $\boldsymbol{\xi}$ (see Rousseeuw and Driessen [1999] and Maronna and Zamar [2002]). Re-centering helps reduce the bias in the tail index estimate caused by a non-zero location. Note that in the symmetric case, the sample mean is an unbiased estimator of location parameter $\boldsymbol{\xi}$; however, when the underlying distribution is skewed, estimation of $\boldsymbol{\xi}$ using the sample mean is subject to a bias. Based on Table 5.1, the performance of estimator $\hat{\nu}$ seems to be satisfactory even when the sample size is relatively small.

Table 5.1: Performance of estimator $\hat{\nu}$ computed using Program 1. The bivariate skew- t distribution with tail index $\nu = 2$ is used in simulations to generate samples of four different sample sizes. Each cell presents the average value of estimates of ν based on 1000 simulated samples of a given size, and the corresponding standard error (in brackets). Two different parameter settings are considered. Case 1: $\boldsymbol{\alpha} = (1, -3)$, $\rho = 0.5$, $\boldsymbol{\xi} = (0, 0)$; Case 2: $\boldsymbol{\alpha} = (1, -3)$, $\rho = 0.5$, $\boldsymbol{\xi} = (3, 1)$.

| Sample Size | Case 1 | Case 2 |
|-------------|-------------|-------------|
| 5000 | 2.00 (0.01) | 2.00 (0.09) |
| 1000 | 2.04 (0.22) | 2.05 (0.22) |
| 500 | 2.08 (0.33) | 2.08 (0.32) |
| 200 | 2.25 (0.61) | 2.23 (0.59) |

5.2 Parametric EVT Estimation

Estimation of the remaining terms in the limit expression (4.3) for $\eta(x, y)$ constitutes a more complex problem than estimation of the tail index ν , as the former requires inference about the multivariate tail behaviour. In the setting of multivariate regular variation, a standard procedure is to make use of the spectral measure Ψ in (2.2).

Suppose $\mathbf{Z}_1, \dots, \mathbf{Z}_n$ are independent and identically distributed random vectors from a bivariate regularly varying distribution; see Definition 2.2.3. Let $R_i = \|\mathbf{Z}_i\|$ and $\mathbf{W}_i = \mathbf{Z}_i/\|\mathbf{Z}_i\|$ denote the radius and direction of \mathbf{Z}_i , $i = 1, \dots, n$, respectively. Let $R_{n-k:n}$ denote the k^{th} largest observation in the sample of R_i 's; it is used to identify extreme or tail observations in the original bivariate sample of \mathbf{Z}_i 's. Then the limit in (2.2) suggests that, given \mathbf{Z}_i with $R_i > R_{n-k:n}$, the sample of \mathbf{W}_i 's can be modelled by measure Ψ . While various parametric and non-parametric procedures are possible to make inference about Ψ , here we build upon the class of skew-elliptical distributions.

Given a particular skewing function $\pi(\cdot) = G(w(\cdot))$, which satisfies assumptions of Theorem 4.1.3, spectral density $\psi(\mathbf{w})$ can be written down explicitly using (4.1). One can then estimate ρ as well as other parameters in $G(w_\infty(\cdot))$ by maximizing the log-likelihood based on $\psi(\mathbf{w})$ for the tail data projected on the unit sphere \mathcal{S}^1 , namely for \mathbf{W}_i 's with $R_i > R_{n-k:n}$.

In our implementation, we assume that the limiting skewing function $G(w_\infty(\cdot))$ has the same form as that for the skew- t distribution. The spectral density in this case is given by

$$\psi_{st}(\mathbf{w}) = 2 |\bar{\Omega}|^{-\frac{1}{2}} T_1\left(\boldsymbol{\alpha}^T \mathbf{w} \sqrt{\frac{\nu+2}{Q^*(\mathbf{w})}}; \nu+2\right) Q^*(\mathbf{w})^{-\frac{\nu+2}{2}} \left[\int_0^{2\pi} A(\theta)^{\frac{\nu}{2}} d\theta \right]^{-1},$$

$$\mathbf{w} \in \mathcal{S}^1; \quad (5.1)$$

cf. Example 1. The parameters to be estimated include shape parameter $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$ and parameter ρ of the standardized scale matrix $\bar{\Omega}$. The details of the resulting inference procedure are summarized in Program 2.

5.2. Parametric EVT Estimation

Algorithm 2 Estimation using parametric EVT method

- (1) Let $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$ represent a bivariate sample with location $\boldsymbol{\xi}$;
 - (2) Re-center data $\tilde{\mathbf{z}}_j = \mathbf{z}_j - \hat{\boldsymbol{\xi}}$, $j = 1, \dots, n$, where $\hat{\boldsymbol{\xi}}$ is a robust estimate of $\boldsymbol{\xi}$;
 - (3) Estimate tail index ν using Program 1;
 - (4) Calculate L_2 -norm distances $\|\tilde{\mathbf{z}}_j\|$ and set $\mathbf{w}_j = \tilde{\mathbf{z}}_j / \|\tilde{\mathbf{z}}_j\|$, $j = 1, 2, \dots, n$;
 Keep \mathbf{w}_j if corresponding $\|\tilde{\mathbf{z}}_j\|$ exceeds a specified high quantile threshold;
 Estimate parameters ρ and $\boldsymbol{\alpha}$ by fitting ψ_{st} in (5.1) to the retained sample of \mathbf{w}_j 's using the maximum-likelihood.
-

In Step (4) of Program 2, it is necessary to select a threshold to identify extreme observations based on their L_2 -norm distances. Such a choice always entails a bias-variance trade-off due to balancing between the model validity and estimation efficiency. To assess stability of our parametric EVT method, we simulated 1000 skew- t samples of size 1000. Absolute differences between the true value of $\eta(x, y)$ and estimated probability $\hat{\eta}(x, y)$ using (4.3) and Program 2 for parameter estimation with different thresholds were calculated, where x and y were taken as the theoretical 99.99% marginal quantiles. Figure 5.1 shows the mean and standard deviation of absolute differences for different thresholds. Based on these plots, in the subsequent analyses, we set the threshold at the 85% quantile of the distance variates $\|\tilde{\mathbf{z}}_j\|$ ($j = 1, \dots, n$).

Assuming a particular parametric form for the spectral density improves efficiency of the estimation procedure in comparison to non-parametric approaches, especially when the sample size is small. While this entails a certain degree of model risk, the proposed model can well capture asymmetric contour shapes and thus gives a more flexible alternative to elliptically symmetric models. Figure 5.2 shows several contour plots that can be modelled by ψ_{st} in (5.1). As only tail observations are used to do the estimation, our approach is fundamentally different from modelling data directly using the skew- t distribution; that is, skew- t distribution fitted using the entire data are subject to undue influence of the central observations, possibly compromising the fit in the tail regions.

5.2. Parametric EVT Estimation

Figure 5.1: Mean (left panel) and standard deviation (right panel) of absolute differences between exact value of $\eta(x, y)$ and estimated probability $\hat{\eta}(x, y)$ using Theorem 4.1.3 for different thresholds. Samples are drawn from the skew- t distribution with $\boldsymbol{\xi} = (0, 0)$, $\boldsymbol{\alpha} = (1, -3)$, $\nu = 2$, $\rho = 0.5$. Values of x and y are chosen as the theoretical marginal quantiles with probability 99.99%.

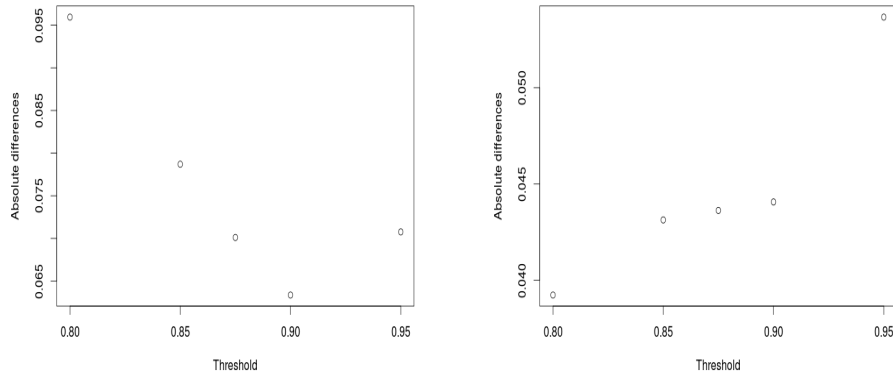
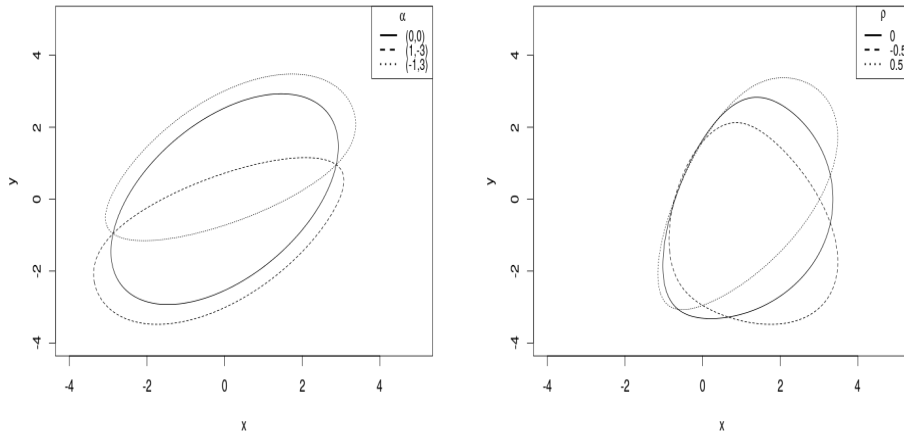


Figure 5.2: The contour plots $\{\mathbf{z} \in \mathbb{R}^2 \mid \psi_{st}(\mathbf{z}) < 1\}$ for spectral density ψ_{st} in (5.1). The parameters are set to $\nu = 2$, $\rho = 0.5$ and several values of $\boldsymbol{\alpha}$ (left panel); $\boldsymbol{\alpha} = (3, -1)$ and several values of ρ (right panel).



Our method shares a similar idea as in Cai et al. [2011], where the authors propose an asymptotically motivated estimator to approximate an extreme risk region of the form $\{\mathbf{z} \in \mathbb{R}^d : h(\mathbf{z}) \leq \beta\}$, where h is the underlying density function and $\beta > 0$ is a small number. They argue that a (semi)-parametric method with a pre-specified form of the spectral density does not perform as well as when the spectral density is estimated non-parametrically. However, in their comparative analysis an elliptically contoured model is used which indeed may not be flexible enough for an application at hand. Different non-parametric consistent estimators for spectral density have been proposed recently (see Nguyen and Samorodnitsky [2013], Eastoe et al. [2014], among others); however, these methods do not fit in our framework.

While asymptotically a non-zero location $\boldsymbol{\xi}$ does not affect the results, in practice when working with finite values the location does have influence on the performance of the estimation procedure. So far we have been using a robust estimate of location $\boldsymbol{\xi}$ to re-centre the data. In the next section, we propose a refinement of Program 2. Based on the subsequent simulation studies with the location parameter substantially different from zero, the proposed iterative procedure leads to more accurate estimation results.

5.3 Iterative Parametric EVT Estimation

To estimate conditional probability $\eta(x, y)$ when the location of the sample is far from the origin, one needs a good estimator for location parameter $\boldsymbol{\xi}$. Ma et al. [2005] propose a locally efficient semi-parametric method specifically for univariate skew-elliptical data. However, their method becomes tedious in the multivariate setting. Other methods including robust estimation used in the previous section cannot handle skewed data very well.

Note that our estimation method relies on the tail data only given an initial estimate of the location. If the location of the shifted sample is close to zero, the accuracy of the estimation using the parametric EVT method of the previous section should not be significantly influenced by the location parameter. Tail data alone are unlikely to provide an accurate estimate of the location parameter which determines the center of the entire data cloud.

5.3. Iterative Parametric EVT Estimation

Based on these considerations, we introduce an iterative procedure to update the location parameter using the entire sample, followed by re-estimation of the other model parameters with the parametric EVT method.

In particular, we assume that the angular component S' in the stochastic representation (3.9) of skew-elliptical distributions for the bivariate case has the following density in polar coordinates:

$$f_{\Theta}(\theta) = \frac{1}{\pi} T_1 \left(\sqrt{2}(L\boldsymbol{\alpha})^T \begin{pmatrix} \cos \theta \\ \sin \theta \end{pmatrix}; 2 \right), \quad \theta \in [0, 2\pi). \quad (5.2)$$

This is precisely the form that was obtained in examples considered in Section 3.3. Now, based on (3.9), write

$$\tilde{\mathbf{Z}} = (\tilde{X}, \tilde{Y}) = (L^{-1})^T (\mathbf{Z} - \boldsymbol{\xi}) = R(\cos \Theta, \sin \Theta),$$

which implies that $\Theta = \tan^{-1}(\tilde{Y}/\tilde{X})$. Hence, given an initial estimate $\hat{\boldsymbol{\xi}}$ of $\boldsymbol{\xi}$ such as the robust estimate used previously and the estimate of ρ , the parameter of standardized scale matrix $\bar{\Omega} = L^T L$, computed using the parametric EVT method, we can obtain the initial sample of angles $\{\Theta_i^{(0)}; i = 1, \dots, n\}$ from the original data $\mathbf{Z}_1, \dots, \mathbf{Z}_n$. Then we can update $\hat{\boldsymbol{\xi}}$ by minimizing a distance between the theoretical density $f_{\Theta}(\theta; \boldsymbol{\xi} | \hat{\boldsymbol{\alpha}}, \hat{\rho})$ in (5.2) and empirical density of Θ_i 's, denoted $\hat{f}_{\Theta, n}(\theta; \boldsymbol{\xi} | \hat{\boldsymbol{\alpha}}, \hat{\rho})$ over values of $\boldsymbol{\xi}$ with the other parameters fixed at their estimates, i.e.,

$$\hat{\boldsymbol{\xi}} = \arg \min_{\boldsymbol{\xi}} \|f_{\Theta}(\theta; \boldsymbol{\xi} | \hat{\boldsymbol{\alpha}}, \hat{\rho}) - \hat{f}_{\Theta, n}(\theta; \boldsymbol{\xi} | \hat{\boldsymbol{\alpha}}, \hat{\rho})\|^2. \quad (5.3)$$

To improve stability of estimation of $\hat{\boldsymbol{\xi}}$, we also control the convergence of $\hat{\boldsymbol{\alpha}}$ and $\hat{\rho}$ by updating their values for the fixed value of $\hat{\boldsymbol{\xi}}$ from the previous iteration:

$$(\hat{\boldsymbol{\alpha}}, \hat{\rho}) = \arg \min_{\boldsymbol{\alpha}, \rho} \|f_{\Theta}(\theta; \boldsymbol{\alpha}, \rho | \hat{\boldsymbol{\xi}}) - \hat{f}_{\Theta, n}(\theta; \boldsymbol{\alpha}, \rho | \hat{\boldsymbol{\xi}})\|^2. \quad (5.4)$$

Let $\|(\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\alpha}}, \hat{\rho})^j - (\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\alpha}}, \hat{\rho})^{j-1}\|_{\infty}$ denote the largest absolute value of differences between two consecutive estimates. Once $\|(\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\alpha}}, \hat{\rho})^j - (\hat{\boldsymbol{\xi}}, \hat{\boldsymbol{\alpha}}, \hat{\rho})^{j-1}\|_{\infty} < \epsilon$ for a chosen tolerance level $\epsilon > 0$, we stop the above iterative procedure of updating parameter estimates. As a final step, given the value of $\hat{\boldsymbol{\xi}}$ from the

5.4. Estimation of the Scale Matrix

last iteration, the data are re-centred and α and ρ are re-estimated using the parametric EVT method. The details of this procedure are summarized in Program 3.

Algorithm 3 Estimation using iterative parametric EVT method

- (1) Let $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$ represent a bivariate sample with location ξ ;
 - (2) Re-center data $\tilde{\mathbf{z}}_j = \mathbf{z}_j - \hat{\xi}$, $j = 1, \dots, n$, where $\hat{\xi}$ is a robust estimate of ξ ;
 - (3) Estimate parameters ν , α and ρ using Program 1 and 2;
 - (4) Set $Diff = 1$
 - while** $Diff > \epsilon$ **do**
 - Compute $\hat{\xi}$ according to (5.3);
 - Given $\hat{\xi}$, update $\hat{\alpha}$ and $\hat{\rho}$ according to (5.4);
 - Set $Diff = \|(\hat{\xi}, \hat{\alpha}, \hat{\rho})^i - (\hat{\xi}, \hat{\alpha}, \hat{\rho})^{i-1}\|_\infty$
 - end while**
 - (5) With updated $\hat{\xi}$ from the last iteration in Step (4), apply Program 2 to re-compute $\hat{\alpha}$ and $\hat{\rho}$.
-

As a final remark, we note that by assuming a particular form of the density of the angular component in (5.2), we implicitly restrict the behaviour of the underlying density generator \tilde{f} due to dependence between the radial and angular components in representation (3.9). However, this restriction is used only to improve the estimation of the location parameter, while the other parameters are estimated as before using the parametric EVT method based on (re-centred) tail data.

5.4 Estimation of the Scale Matrix

So far, in Theorem 4.1.3 and inference methods discussed above, it was assumed that the scale matrix is given in the standardized form as $\bar{\Omega}$. However, estimation of scale matrix Ω assuming that the underlying distribution is skewed is not as straightforward as in the case of elliptical symmetry.

To address this problem, first recall $\Omega = \omega \bar{\Omega} \omega$, where ω is the diagonal matrix defined in (3.3). Note that if $\mathbf{Z} \sim SE_d(\xi, \Omega, \tilde{f}, G \circ w)$, then from (3.9)

it follows that $\omega^{-1}\mathbf{Z} \sim SE_d(\omega^{-1}\boldsymbol{\xi}, \bar{\Omega}, \tilde{f}, G \circ w)$. A simple approach is to first estimate ω by say a robust method mentioned before (see Rousseeuw and Driessen [1999] and Maronna and Zamar [2002]), and then directly apply previous inference results to the transformed vector $\omega^{-1}\mathbf{Z}$. However, the performance of this approach is not satisfactory especially when data is skewed and ω is far from the identity matrix. Other approaches include:

M1 Estimate ω via MLE with density $\psi(\mathbf{w})$ in (4.1);

M2 Estimate ω together with $\boldsymbol{\xi}$ using the iterative parametric EVT method, i.e.,

$$(\hat{\boldsymbol{\xi}}, \hat{\omega}) = \arg \min_{\boldsymbol{\xi}, \omega} \|f_{\Theta}(\theta; \boldsymbol{\xi}, \omega | \hat{\boldsymbol{\alpha}}, \hat{\rho}) - \hat{f}_{\Theta, n}(\theta; \boldsymbol{\xi}, \omega | \hat{\boldsymbol{\alpha}}, \hat{\rho})\|^2.$$

The above two approaches only require slight modifications of Program 2 and 3. It is worth to mentioning that method M1 requires simultaneous estimation of too many parameters making the optimization process hard to control. In Section 5.5, we demonstrate that the second approach (M2) works better than the first one.

5.5 Simulation Studies

In this section we report results of a simulation study to show performance of the proposed methods. The assessment is based on 1000 samples of size 1000 simulated from two bivariate skew- t distributions, one with a standardized scale matrix $\bar{\Omega}$ and the other with scale matrix $\Omega = \omega \bar{\Omega} \omega$, where $\omega = \text{diag}(2, 3)$. The other distribution parameters are set as $\boldsymbol{\xi} = (3, 1)$, $\boldsymbol{\alpha} = (1, -3)$, $\rho = 0.5$, and $\nu = 2$. In the first case, we do not estimate ω , in other words, we take $\bar{\Omega}$ as given; in the second case, ω is estimated using the two approaches discussed in the previous section. For the estimation of conditional probability $\eta(x, y)$, the values of x and y are taken as the theoretical 97.5%, 99.0%, 99.9% and 99.99% marginal quantiles. For each sample and various values of x and y , we evaluated $\hat{\eta}(x, y)$ using different methods:

- AFG ($\hat{\eta}_{AFG}(x, y)$): based on the limit result for elliptical distributions (see Abdous et al. [2005b]), and ω is estimated using robust method when necessary;
- Parametric EVT ($\hat{\eta}_1(x, y)$): based on the limit result in Theorem 4.1.3 and the parametric EVT method (Program 2), and ω is estimated using M1 when necessary;
- Iterative parametric EVT ($\hat{\eta}_2(x, y)$): based on the limit result in Theorem 4.1.3 and the iterative parametric EVT method (Program 3), and ω is estimated using M2 when necessary;
- Empirical ($\hat{\eta}_{emp}(x, y)$): based on the empirical distribution;
- True ($\eta(x, y)$): the true value computed by numerical integration using preset parameters;
- Limiting ($\eta_{lim}(x, y)$): the true value of the limit stated in Theorem 4.1.3 using preset parameters.

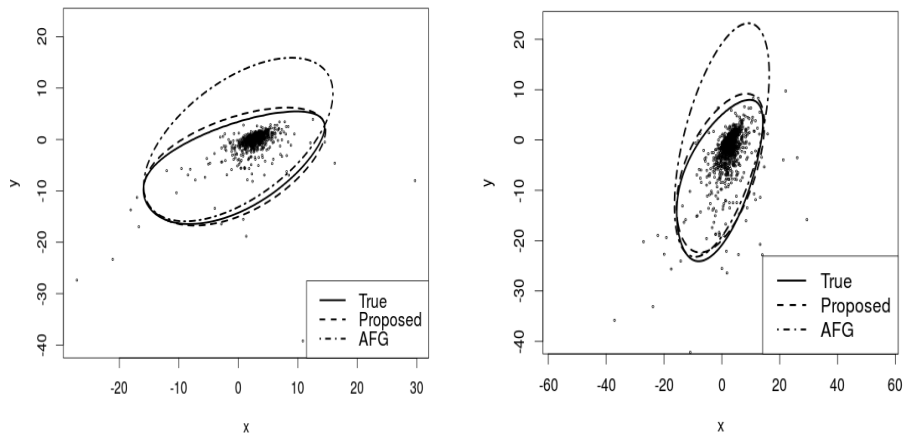
The simulation results are presented in Tables 5.2 and 5.3. They show that, under the considered simulation settings, the AFG method greatly underestimates condition probability $\eta(x, y)$. The empirical method provides good estimates of $\eta(x, y)$ for moderate quantile levels relative to the sample size, albeit a much larger standard deviation in comparison to the other methods. For the very extreme quantile levels, no proper empirical estimates could be obtained. The two methods proposed in this project produce accurate estimates of $\eta(x, y)$ with reasonable standard errors, and clearly provide an improvement over the AFG method in terms of accuracy. The iterative parametric EVT estimator $\hat{\eta}_2(x, y)$ does achieve gains in accuracy over the parametric EVT estimator $\hat{\eta}_1(x, y)$ due to better estimation of the location parameter. A similar simulation for elliptically symmetric distributions was also conducted (see Table B.1), in which our estimator $\hat{\eta}_2(x, y)$ showed a similar performance to $\hat{\eta}_{AFG}(x, y)$ ².

²For comparison, we also report simulation results for a bivariate skew- t distribution with a lighter tail ($\nu = 20$) in Table B.2. We can see that the limiting value $\eta_{lim}(x, y)$ is still close to the true value $\eta(x, y)$, but estimate $\hat{\eta}_2(x, y)$ is not. This is due to poor

5.5. Simulation Studies

Figure 5.3 compares the true contour $\{\mathbf{z} \in \mathbb{R}^2 \mid \psi_{st}(\mathbf{z}) < 1\}$ with estimated contours $\{\mathbf{z} \in \mathbb{R}^2 \mid \hat{\psi}_{st}(\mathbf{z}) < 1\}$ for one of the simulated scenarios using the iterative parametric EVT method and the method discussed in Abdous et al. [2005b]. It can be easily observed that the proposed method is more flexible in capturing asymmetric shapes of the data clouds.

Figure 5.3: Comparison of the true level curve $\{\psi_{st}(\mathbf{z}) < 1\}$ with the estimated contours $\{\hat{\psi}_{st}(\mathbf{z}) < 1\}$ using the AFG method and the proposed iterative parametric EVT method. The data are generated from a bivariate skew- t distribution with parameters $\boldsymbol{\xi} = (3, 1)$, $\boldsymbol{\alpha} = (1, -3)$, $\rho = 0.5$, $\nu = 2$, and $\boldsymbol{\omega} = (1, 1)$ (left panel) or $\boldsymbol{\omega} = (2, 3)$ (right panel).



estimation of tail index when the tail decays fast. Abdous et al. [2005a] also suffers the same problem.

Table 5.2: Simulation results based on 1000 samples of size 1000 from a bivariate skew- t distribution with parameters $\xi = (3,1)$, $\alpha = (1,-3)$, $\nu = 2$, $\rho = 0.5$ and a standardized scale matrix. Each cell provides the average (standard deviation) of the estimates of $\eta(x, y)$ under various methods; see Section 5.5 for details. For $\hat{\eta}_{AFG}(x, y)$, $\hat{\eta}_1(x, y)$, $\hat{\eta}_2(x, y)$ and $\eta_{\text{lim}}(x, y)$, we used $z = y/x - \rho$ in the limit results. Values of x and y are chosen as the theoretical marginal quantiles with probability p , where p labels columns and rows.

| Quantile y | | Quantile x | | | |
|------------|--------------------------------------|---------------|---------------|---------------|---------------|
| | | 97.5% | 99.0% | 99.9% | 99.99% |
| 97.5% | $\hat{\eta}_{AFG}(x, y)$ | 0.239 (0.019) | 0.215 (0.019) | 0.179 (0.018) | 0.165 (0.017) |
| | $\hat{\eta}_1(x, y)$ | 0.605 (0.046) | 0.562 (0.046) | 0.490 (0.044) | 0.460 (0.043) |
| | $\hat{\eta}_2(x, y)$ | 0.693 (0.048) | 0.641 (0.048) | 0.553 (0.046) | 0.517 (0.045) |
| | $\hat{\eta}_{emp}(x, y)$ | 0.653 (0.098) | 0.588 (0.168) | * | * |
| | $\eta(x, y)/\eta_{\text{lim}}(x, y)$ | 0.652/0.685 | 0.587/0.633 | 0.518/0.544 | 0.498/0.508 |
| 99.0% | $\hat{\eta}_{AFG}(x, y)$ | 0.273 (0.020) | 0.238 (0.019) | 0.186 (0.018) | 0.167 (0.017) |
| | $\hat{\eta}_1(x, y)$ | 0.659 (0.045) | 0.603 (0.046) | 0.506 (0.044) | 0.465 (0.043) |
| | $\hat{\eta}_2(x, y)$ | 0.755 (0.046) | 0.691 (0.048) | 0.573 (0.047) | 0.523 (0.045) |
| | $\hat{\eta}_{emp}(x, y)$ | 0.761 (0.089) | 0.658 (0.165) | * | * |
| | $\eta(x, y)/\eta_{\text{lim}}(x, y)$ | 0.761/0.749 | 0.661/0.683 | 0.540/0.564 | 0.505/0.514 |
| 99.9% | $\hat{\eta}_{AFG}(x, y)$ | 0.507 (0.024) | 0.399 (0.022) | 0.234 (0.019) | 0.181 (0.018) |
| | $\hat{\eta}_1(x, y)$ | 0.863 (0.030) | 0.794 (0.038) | 0.596 (0.046) | 0.495 (0.044) |
| | $\hat{\eta}_2(x, y)$ | 0.932 (0.022) | 0.884 (0.032) | 0.683 (0.048) | 0.559 (0.046) |
| | $\hat{\eta}_{emp}(x, y)$ | 0.963 (0.039) | 0.911 (0.095) | * | * |
| | $\eta(x, y)/\eta_{\text{lim}}(x, y)$ | 0.964/0.930 | 0.915/0.882 | 0.666/0.675 | 0.542/0.550 |
| 99.99% | $\hat{\eta}_{AFG}(x, y)$ | 0.880 (0.022) | 0.802 (0.026) | 0.439 (0.022) | 0.231 (0.019) |
| | $\hat{\eta}_1(x, y)$ | 0.977 (0.008) | 0.961 (0.012) | 0.823 (0.035) | 0.592 (0.046) |
| | $\hat{\eta}_2(x, y)$ | 0.991 (0.004) | 0.984 (0.007) | 0.905 (0.028) | 0.677 (0.048) |
| | $\hat{\eta}_{emp}(x, y)$ | 0.996 (0.012) | 0.991 (0.030) | * | * |
| | $\eta(x, y)/\eta_{\text{lim}}(x, y)$ | 0.996/0.990 | 0.991/0.983 | 0.916/0.903 | 0.667/0.670 |

Table 5.3: Simulation results based on 1000 samples of size 1000 from a bivariate skew- t distribution with parameters $\xi = (3,1)$, $\alpha = (1,-3)$, $\nu = 2$, $\rho = 0.5$ and $\omega = \text{diag}(2,3)$ for the scale matrix. Each cell provides the average (standard deviation) of the estimates of $\eta(x, y)$ under various methods; see Section 5.5 for details. For $\hat{\eta}_{AFG}(x, y)$, $\hat{\eta}_1(x, y)$, $\hat{\eta}_2(x, y)$ and $\eta_{\text{lim}}(x, y)$ we used $z = \omega_1 y / \omega_2 x - \rho$ in the limit results. Values of x and y are chosen as the theoretical marginal quantiles with probability p , where p labels columns and rows.

| Quantile y | Quantile x | | | |
|--|----------------|---------------|---------------|---------------|
| | 97.5% | 99.0% | 99.9% | 99.99% |
| $\hat{\eta}_{AFG}(x, y)$ | 0.236 (0.035) | 0.210 (0.027) | 0.178 (0.018) | 0.168 (0.017) |
| $\hat{\eta}_1(x, y)$ | 0.540 (0.049) | 0.491 (0.047) | 0.425 (0.046) | 0.402 (0.045) |
| 97.5% $\hat{\eta}_2(x, y)$ | 0.627 (0.050) | 0.577 (0.050) | 0.510 (0.049) | 0.485 (0.048) |
| $\hat{\eta}_{emp}(x, y)$ | 0.653 (0.098) | 0.588 (0.168) | * | * |
| $\eta(x, y) / \eta_{\text{lim}}(x, y)$ | 0.652/0.648 | 0.587/0.597 | 0.518/0.526 | 0.498/0.501 |
| $\hat{\eta}_{AFG}(x, y)$ | 0.286 (0.0491) | 0.241 (0.037) | 0.187 (0.020) | 0.170 (0.017) |
| $\hat{\eta}_1(x, y)$ | 0.623 (0.048) | 0.549 (0.049) | 0.444 (0.046) | 0.408 (0.045) |
| 99.0% $\hat{\eta}_2(x, y)$ | 0.708 (0.048) | 0.635 (0.049) | 0.529 (0.049) | 0.492 (0.048) |
| $\hat{\eta}_{emp}(x, y)$ | 0.761 (0.089) | 0.658 (0.165) | * | * |
| $\eta(x, y) / \eta_{\text{lim}}(x, y)$ | 0.761/0.733 | 0.661/0.657 | 0.540/0.547 | 0.505/0.508 |
| $\hat{\eta}_{AFG}(x, y)$ | 0.598 (0.100) | 0.457 (0.086) | 0.245 (0.038) | 0.186 (0.020) |
| $\hat{\eta}_1(x, y)$ | 0.891 (0.026) | 0.810 (0.037) | 0.556 (0.049) | 0.442 (0.046) |
| 99.90% $\hat{\eta}_2(x, y)$ | 0.929 (0.021) | 0.870 (0.033) | 0.642 (0.050) | 0.527 (0.049) |
| $\hat{\eta}_{emp}(x, y)$ | 0.963 (0.039) | 0.911 (0.095) | * | * |
| $\eta(x, y) / \eta_{\text{lim}}(x, y)$ | 0.964/0.945 | 0.915/0.893 | 0.666/0.665 | 0.542/0.544 |
| $\hat{\eta}_{AFG}(x, y)$ | 0.915 (0.055) | 0.846 (0.078) | 0.489 (0.090) | 0.246 (0.038) |
| $\hat{\eta}_1(x, y)$ | 0.985 (0.006) | 0.970 (0.010) | 0.832 (0.034) | 0.557 (0.048) |
| 99.99% $\hat{\eta}_2(x, y)$ | 0.991 (0.004) | 0.981 (0.007) | 0.887 (0.030) | 0.644 (0.049) |
| $\hat{\eta}_{emp}(x, y)$ | 0.996 (0.012) | 0.991 (0.030) | * | * |
| $\eta(x, y) / \eta_{\text{lim}}(x, y)$ | 0.996/0.994 | 0.991/0.987 | 0.916/0.908 | 0.667/0.666 |

Chapter 6

Financial Applications

6.1 Financial Contagion

In this section, a practical application is illustrated in the context of financial contagion at the domestic level, for which estimation of the extreme conditional excess probability is needed. We investigate the extreme conditional excess probability $1 - \eta(x, y) = \mathbb{P}(L^j > y | L^s > x)$, where L^s is the daily loss on a portfolio which is a proxy for the aggregate financial system, and L^j is the daily loss on a stock or a portfolio. This probability quantifies the tail risk of a stock when the overall financial market encounters a severe negative shock. It could also be interpreted as the extreme dependence of the stock and the aggregate financial system.

We consider a partial panel of financial institutions studied in Acharya et al. [2010], with complete data between June 1, 2006 and May 31, 2008. The time period from June 1, 2006 to May 31, 2007 is defined as the pre-crisis window (same as in Girardi and Tolga [2013]), and the remaining one-year is the crisis window. The panel contains U.S. financial firms in the depositories industry with a market capitalization greater than 5 billion USD as of the end of June 2007. The Dow Jones US Financials Index (DJUSFN) is used as a proxy for the aggregate financial system. The daily prices and capitalization information are extracted from Yahoo Finance.

Daily losses are calculated as negative log returns. Due to the presence of serial dependence and volatility clustering in daily financial returns, it is common to first filter the data using the AR(1)-GARCH(1,1) model. Then residuals can be treated as a sequence of independent and identically distributed random variables.

Table 6.1 reports estimation results for the filtered data. For each institution in pre-crisis period, we estimate $1 - \eta(x, y)$ using the AFG method and the iterative parametric EVT method (denoted as before $1 - \hat{\eta}_{AFG}(x, y)$ and $1 - \hat{\eta}_2(x, y)$, respectively), together with the estimated marginal survival probability $\hat{\mathbb{P}}(Y > y)$. The latter is estimated by fitting the Generalized Pareto distribution to threshold excesses using the maximum-likelihood method (see, e.g., Coles [2001] for details). The values of x and y are, respectively, the 99.9% quantile of the losses on DJUSFN and the average of the 99% quantile of the losses on cross-sectional stocks³. The right columns in the panel “pre-crisis” present parameter estimates used to evaluate $1 - \hat{\eta}_2(x, y)$. The “crisis” column reports the estimated extreme conditional excess probability during the crisis period.

As indicated by the estimates of the tail index ν , the bivariate joint distributions of losses on stocks and the DJUSFN index exhibit a fairly slow radial decay, with most of $\hat{\nu}$ in the range from 2.5 to 4.5⁴. Based on the estimates of the shape parameter $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$, there is evidence of skewness in the joint distribution of the index and stock returns over the considered one-year estimation window. It can be noted that the estimates of the skewness parameter α_1 , corresponding to the DJUSFN index, tend to fluctuate considerably, contrary to the intuition that these estimates should be fairly stable and reflect the skewness in the index data. Partly, this can be explained by the fact that α_1 is not the marginal parameter. If the data of stock and the index losses were to come from a bivariate skew- t distribution, one could convert the bivariate skewness parameter $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)$ to the marginal parameters $\bar{\alpha}_1$ and $\bar{\alpha}_2$ using Equation 4.4. Indeed, $\hat{\alpha}_1$ should be fairly constant as the marginal corresponding to the index is the same for all pairs of data; $\hat{\alpha}_1$ is determined by $\hat{\alpha}_2$ and the parameters of the scale matrix. Another reason for a large variation in the values of $\hat{\alpha}_1$ is the level of sampling variability when estimating the tail of the multivariate distribution. However, majority of the estimates of α_1 do tend to be positive,

³The values of x and y are 3.487% and 2.722% for pre-crisis period, and 4.587% and 5.761% for crisis period.

⁴The Hill estimators (Hill [1975]) for the linear combinations of losses on stocks and the DJUSFN index have a similar tail index range; hence, the assumption of multivariate regular variation is reasonable in this example.

thus indicating that the losses on the index are likely to be right-skewed. This is in line with the skewness estimate calculated using the index data alone. Finally, it is worth pointing out that a similar issue of large variability in skewness parameter estimates was present in the simulation studies. At the same time, the estimates of the conditional probability, the ultimate quantity of interest here, were fairly stable and close to the true values.

Comparing $1 - \hat{\eta}_{AFG}$ and $1 - \hat{\eta}_2$, we see that under the assumption of elliptical symmetry, the estimates of the conditional exceedance probability are, in most cases, bigger than those obtained when the skewness in the joint distribution of losses is accounted for. We select Hudson City Bancorp which has a small difference between $1 - \hat{\eta}_{AFG}$ and $1 - \hat{\eta}_2$ and Peoples Bank Bridgeport which has a large difference, and plot their fitted contours $\{\mathbf{z} \in \mathbb{R}^2 \mid \hat{\psi}_{st}(\mathbf{z}) < 1\}$ in Figure 6.1. It can be observed that the elliptical contour under the AFG method provides a poor fit to the data cloud. Hence, in this particular example, ignoring the asymmetry leads to inaccurate estimation of $\eta(x, y)$. We also note that, since conditional exceedance probabilities are greater than the corresponding marginal exceedance probabilities for Y , the data clearly exhibit contagion from the DJUSFN index to each institution. Lastly, the results also show that extremal dependence cannot be modelled by the AR(1)-GARCH(1,1) process. This is in line with the findings in Bae et al. [2003], where the counts of coincident extreme daily returns across international equity markets could not be explained by the AR(1)-GARCH(1,1) filter.

To compare financial contagion between the pre-crisis and crisis periods, we report the estimated extreme conditional excess probability in the crisis period using the iterative parametric EVT method (the last column of Table 6.1). It is interesting to observe that the extreme conditional excess probability drops substantially during the financial crisis, indicating that extreme dependence between the market and individual stocks decreases in a downside market. One possible explanation is that firms start to watch their risk exposures when they realize they are in a dangerous position. This result is in accordance with the findings in Adrian and Brunnermeier [2011] that contemporaneous risk measures⁵ are procyclical.

⁵They measure systemic risk using CoVaR.

6.1. Financial Contagion

Figure 6.1: Estimated contours $\{\hat{\psi}_{st}(\mathbf{z}) < 1\}$ for daily losses on DJUSFN versus Hudson City Bancorp (HCBK, left panel) and Peoples Bank Bridgeport (PBCT, right panel) between June 1, 2006 and May 31, 2007 using the AFG method and the proposed iterative parametric EVT method.

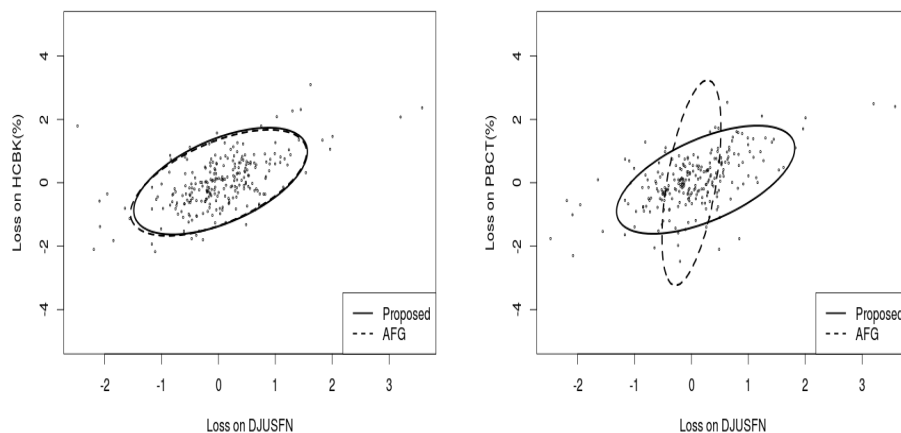


Table 6.1: Point estimation of extreme conditional excess probability for 28 financial institutions. Daily losses are computed using log returns and are filtered by the AR(1)-GARCH(1,1) process. The threshold values of x and y are, respectively, 99.9% quantile of losses on DJUSFN and average of 99% quantile of losses on cross-sectional stocks. The sample period is from June 1, 2006 to May 31, 2008. The pre-crisis period is from June 1, 2006 to May 31, 2007, and the crisis period is from June 1, 2007 to May 31, 2008.

| Company | Pre-crisis | | | | | | | Crisis |
|-------------------------------|----------------------|------------------|-----------------------------|-------------|------------------|------------------|--------------|---------------------------|
| | $1-\hat{\eta}_{AFG}$ | $1-\hat{\eta}_2$ | $\hat{\mathbb{P}}(L^j > y)$ | $\hat{\nu}$ | $\hat{\alpha}_1$ | $\hat{\alpha}_2$ | $\hat{\rho}$ | $1-\hat{\eta}_2$ (crisis) |
| 1 BANK NEW YORK | 0.856 | 0.722 | 0.008 | 2.815 | 0.177 | -0.385 | 0.790 | 0.224 |
| 2 BANK OF AMERICA CORP | 0.675 | 0.778 | 0.006 | 3.444 | -0.448 | 0.346 | 0.785 | 0.198 |
| 3 B B & T CORP | 0.703 | 0.859 | 0.003 | 3.113 | -0.524 | 0.568 | 0.811 | 0.291 |
| 4 CITIGROUP | 0.790 | 0.515 | 0.008 | 4.248 | 0.648 | -0.770 | 0.604 | 0.437 |
| 5 COMERICA | 0.781 | 0.508 | 0.012 | 4.042 | 0.052 | 0.056 | 0.509 | 0.330 |
| 6 COMMERCE BANCORP | 0.814 | 0.583 | 0.039 | 2.751 | 0.623 | -0.500 | 0.564 | 0.042 |
| 7 HUDSON CITY BANCORP | 0.522 | 0.527 | 0.003 | 3.319 | 0.028 | 0.082 | 0.555 | 0.154 |
| 8 HUNTINGTON BANCSHARES | 0.718 | 0.678 | 0.008 | 2.850 | 0.136 | -0.413 | 0.829 | 0.325 |
| 9 J P MORGAN CHASE & CO | 0.871 | 0.838 | 0.011 | 3.853 | 0.262 | -0.280 | 0.809 | 0.096 |
| 10 KEYCORP | 0.777 | 0.649 | 0.010 | 3.513 | 0.413 | -0.250 | 0.700 | 0.512 |
| 11 MARSHALL & ILSLEY CORP | 0.749 | 0.737 | 0.007 | 2.809 | -0.608 | 0.708 | 0.676 | 0.469 |
| 12 M & T BANK CORP | 0.818 | 0.630 | 0.014 | 3.448 | 0.125 | 0.023 | 0.697 | 0.518 |
| 13 NATIONAL CITY CORP | 0.728 | 0.649 | 0.018 | 3.161 | 0.030 | 0.049 | 0.596 | 0.020 |
| 14 NEW YORK COMMUNITY BANCORP | 0.521 | 0.364 | 0.000 | 6.051 | -0.681 | 0.344 | 0.357 | 0.094 |
| 15 NORTHERN TRUST CORP | 0.827 | 0.733 | 0.014 | 3.360 | 0.179 | -0.052 | 0.725 | 0.300 |
| 16 PEOPLES BANK BRIDGEPORT | 0.815 | 0.530 | 0.028 | 2.469 | 0.548 | -0.145 | 0.593 | 0.043 |
| 17 PNC FINANCIAL SERVICES GRP | 0.643 | 0.428 | 0.010 | 4.459 | 0.332 | -0.189 | 0.493 | 0.122 |

Continued on next page

| Company | Pre-crisis | | | | | | | Crisis |
|---------------------------|----------------------|------------------|-----------------------------|-------------|------------------|------------------|--------------|---------------------------|
| | $1-\hat{\eta}_{AFG}$ | $1-\hat{\eta}_2$ | $\hat{\mathbb{P}}(L^j > y)$ | $\hat{\nu}$ | $\hat{\alpha}_1$ | $\hat{\alpha}_2$ | $\hat{\rho}$ | $1-\hat{\eta}_2$ (crisis) |
| 18 REGIONS FINANCIAL CORP | 0.700 | 0.528 | 0.006 | 3.767 | -0.294 | -0.297 | 0.693 | 0.332 |
| 19 SOVEREIGN BANCORP | 0.650 | 0.590 | 0.025 | 3.596 | -0.044 | 0.300 | 0.461 | 0.365 |
| 20 STATE STREET CORP | 0.846 | 0.718 | 0.023 | 2.827 | 0.735 | -0.737 | 0.718 | 0.265 |
| 21 SUNTRUST BANKS | 0.629 | 0.448 | 0.008 | 3.943 | -0.014 | -0.000 | 0.507 | 0.372 |
| 22 SYNOVUS FINANCIAL CORP | 0.693 | 0.583 | 0.001 | 3.522 | 0.009 | 0.008 | 0.683 | 0.324 |
| 23 UNIONBANCAL CORP | 0.692 | 0.668 | 0.021 | 2.795 | 0.001 | 0.003 | 0.684 | 0.296 |
| 24 U S BANCORP DEL | 0.435 | 0.468 | 0.000 | 3.182 | -0.145 | -0.157 | 0.735 | 0.190 |
| 25 WACHOVIA CORP | 0.720 | 0.813 | 0.006 | 4.373 | 0.005 | 0.007 | 0.773 | 0.457 |
| 26 WASHINGTON MUTUAL | 0.830 | 0.730 | 0.022 | 3.380 | -0.019 | 0.264 | 0.564 | 0.540 |
| 27 WELLS FARGO & CO | 0.873 | 0.593 | 0.024 | 2.035 | 0.806 | -1.059 | 0.734 | 0.270 |
| 28 ZIONS BANCORP | 0.652 | 0.410 | 0.010 | 3.859 | 0.206 | -0.086 | 0.532 | 0.387 |

6.2 CoVaR

Conditional Value-at-Risk (CoVaR), introduced by Adrian and Brunnermeier [2011], is defined as Value-at-Risk (VaR) of one portfolio conditional on one institution being in financial distress. Given losses (negative returns) L^j (or L^s) of an institution (or a portfolio) and the confidence level q_1 , $\text{VaR}_{q_1}^j$ is defined as q_1 -quantile of loss distribution:

$$\mathbb{P}(L^j \geq \text{VaR}_{q_1}^j) = q_1,$$

and $\text{CoVaR}_{q_1}^{s|j}$ is defined by the q_1 -quantile of the conditional loss distribution

$$\mathbb{P}(L^s \geq \text{CoVaR}_{q_1}^{s|j} | L^j \geq \text{VaR}_{q_2}^j) = q_1. \quad (6.1)$$

Different from VaR, which only considers individual risk faced by an institution, CoVaR accounts for the possible contribution of each institution to the overall system risk. Additionally, it provides a way to capture the risk spillovers among institutions.

Girardi and Tolga [2013] define the systemic risk of an institution using $\Delta \text{CoVaR}_{q_1}^{s|j}$, the change between its CoVaR in benchmark state (defined as a one-standard deviation event) and its CoVaR under financial distress:

$$\Delta \text{CoVaR}_{q_1}^{s|j} = 100 * \frac{\text{CoVaR}_{q_1}^{s|j} - \text{CoVaR}_{q_1}^{s|b^j}}{\text{CoVaR}_{q_1}^{s|b^j}}, \quad (6.2)$$

where b^j is the benchmark state, which is one standard deviation about the mean: $L^j \in (\mu^j - \sigma^j, \mu^j + \sigma^j)$, where μ^j and σ^j are, respectively, the conditional mean and the standard deviation of institution j 's losses⁶.

In their examples, q_1 and q_2 are set to be 5%, and the joint dynamic of losses is estimated using a bivariate GARCH model with Engle [2002] DCC specification. To take skewness and kurtosis into consideration, they report results for both Gaussian and skew- t innovations. Despite that, their approach still suffers from common shortcomings of estimating tail probability using parametric models. Specifically, when one is interested in CoVaR with a very small q_2 , that is, the impact of one institution's bankruptcy on

⁶Specifically, $\mathbb{P}(L^s \geq \text{CoVaR}_{q_1}^{s|b^j} | \mu^j - \sigma^j \leq L^j \leq \mu^j + \sigma^j) = q_1$.

overall financial networks, their specification could not handle tail estimation well. However, the fact that 2008 financial crisis was partially triggered by the bankruptcy of Lehman Brothers makes this question essential for macroprudential supervision and regulation.

Our method could help resolve this problem. We continue considering financial institutions listed in Table 6.1⁷. The objective is to calculate the ΔCoVaR of the DJUSFN index (denoted as L^s) when one of the institutions (denoted as L^j) is being in financial distress. The sample period is taken from June 1, 2006 to May 31, 2007.

To fit in with our methodology, we adopt estimation procedure described below. First, losses L_t^j are filtered by the AR(1)-GARCH(1,1) model:

$$L_t^j = \mu_t^j + \sigma_{j,t} Z_t^j,$$

where Z_t^j are i.i.d. random variables with a univariate standard Gaussian distribution, and μ_t and σ_t are measurable with respect to sigma algebra \mathcal{F}_{t-1} , representing the information about the process $\{L_t^j\}$ available up to time $t-1$. In order to capture typical time dynamics of financial time series, one possibility is to assume that the conditional mean μ_t follows an AR(1) process

$$\mu_t^j = \alpha_0 + \alpha_1 L_{t-1}^j,$$

while the condition variance σ_t^2 evolves according to a GARCH(1,1) model specification

$$\sigma_{j,t}^2 = \beta_0^j + \beta_1^j (\sigma_{j,t-1} Z_{t-1}^j)^2 + \beta_2^j \sigma_{j,t-1}^2.$$

Since it is generally agreed that loss series exhibit both skewness and (excess) kurtosis, the (normalized) residuals $\hat{Z}_t^j = \frac{L_t^j - \hat{\mu}_t^j}{\hat{\sigma}_t^j}$, where $\hat{\mu}_t^j$ and $\hat{\sigma}_t^j$ are estimates of μ_t^j and σ_t^j , do not follow the Gaussian distribution perfectly. Hence, we compute $\text{VaR}_{q_2,t}^j$ of each institution using EVT approach introduced by McNeil and Frey [2000]. This approach is slightly different from that in Girardi and Tolga [2013], where they directly adopted the skew- t distribution when filtering the parameters of the AR(1)-GARCH((1,1) filter and then took the quantile of skew- t as the estimate of $\text{VaR}_{q_2,t}^j$. The

⁷This group of institutions is the same as “Depositories” group in Girardi and Tolga [2013].

6.2. CoVaR

rationale to separate the estimation process is that the conditional variance is the feature of the whole distribution, while $\text{VaR}_{q_2,t}^j$ is related to tail observations. A similar two-stage estimation is adopted by McNeil and Frey [2000] and Diebold et al. [2000].

Next, instead of specifying a bivariate parametric GARCH model, we simply assume that the joint dynamic of standardized residuals $\hat{\mathbf{Z}}_t = (\hat{Z}_t^s, \hat{Z}_t^j)$ follows a certain skew-elliptical distribution, and then model it using the iterative parametric EVT method. Specifically, $\text{CoVaR}_{q_1,t}^{Z^s|Z^j}$ can be estimated by solving equation below, where Theorem 4.1.3 can be applied:

$$\mathbb{P}(\hat{Z}_t^s \geq \text{CoVaR}_{q_1,t}^{Z^s|Z^j} | \hat{Z}_t^j \geq \text{VaR}_{q_2,t}^j) = q_1. \quad (6.3)$$

Then the forecast of $\text{CoVaR}_{q_1,t}^{s|j}$ is calculated with predictions of $\hat{\mu}_{t+1}^s$ and $\hat{\sigma}_{t+1}^s$ from AR(1)-GARCH(1) model:

$$\text{CoVaR}_{q_1,t}^{s|j} = \hat{\mu}_{t+1}^s + \hat{\sigma}_{t+1}^s \widehat{\text{CoVaR}}_{q_1,t}^{Z^s|Z^j}.$$

For the benchmark case, we adopt empirical approach to compute $\widehat{\text{CoVaR}}_{q_1,t}^{Z^s|Z^{b_j}}$ and forecast $\text{CoVaR}_{q_1,t}^{s|b_j}$ using above equation.

Table 6.2 reports the summary statistics⁸ for cross-sectional daily conditional $\Delta \text{CoVaR}_{q_1,t}^{s|j}$ over the sample period for $q_1 = 5\%$ and $q_2 = 5\%, 1\%$ and 0.01% . “EVT”⁹ indicates that bivariate \hat{Z}_t is modelled using the iterative parametric EVT method; for comparison, we also report results when $\text{CoVaR}_{q_1,t}^{s|j}$ is estimated assuming that \hat{Z}_t follows a bivariate skew- t distribution, as well as the empirical distribution. The numbers in column “Mean” give the increase, on average across all the institutions, in the CoVaR at level q_1 of the aggregate financial system given that an institution experiences a loss in excess of its VaR at level q_2 in comparison to when the institution is in its benchmark state. For example, when institutions are subject to large

⁸The summary statistics are robust to different methods adopted to model $\text{VaR}_{q_2,t}^j$ and the benchmark case. Please find other results in the Appendix B.2.

⁹ When the tail index ν estimated by the EVT method is greater than 10, we drop the estimate results of “EVT” method. Totally, there are 40 out of 7000 values dropped. This is because the performance of the iterative parametric EVT method is not satisfactory when the tail decays fast.

6.2. CoVaR

losses exceeding their 5% VaR, the 5% CoVaR of the aggregate financial system increases on average by 139% over its 5% CoVaR when the institutions are in their benchmark state, based on the "EVT" method. "Std.TS", which gives the average of the standard deviations of individual $\Delta\text{CoVaR}_{q_1,t}^{s|j}$, is a proxy for volatility of systemic risk contribution over time. "Std.CS", which reports the standard deviation of the mean of each individual $\Delta\text{CoVaR}_{q_1,t}^{s|j}$, is a proxy for the dispersion of the average systemic risk contribution.

Summary statistics for $q_2 = 5\%$ in row "skew- t " are quite similar to those reported in Table 6 of Girardi and Tolga [2013]. Comparing these with results in row "EVT", it is clear that $\Delta\text{CoVaR}_{q_1,t}^{s|j}$ estimated using the EVT method exhibits much higher standard derivation and much wider value range; for instance, Std.TS goes from 31.1 to 41.9. This suggests that the impact of one financial institution's distress on the aggregate financial system is estimated more conservatively by the parametric (skew- t) model in comparison to the proposed EVT method. When q_2 decreases from 0.05 to 0.01, both Std.TS and Std.CS increase. This reveals that the uncertainty of the risk spillovers among institutions rises when one institution faces a severe financial distress.

Table 6.2: Summary statistics for cross-sectional $\Delta\text{CoVaR}_{q_1,t}^{s|j}$ for all institutions during sample period from June 1, 2006 to May 31, 2007. Level q_1 is set to be 5%, and q_2 is 5%, 1% or 0.01%. "EVT" and "skew- t ", respectively, refer to the use of the iterative EVT method and the bivariate skew- t distribution to model the sequence of standardized residuals $\{\hat{\mathbf{z}}_t\}$. Estimation based on the empirical distribution is reported under "Empirical". Column "Std.TS" reports the average of the standard deviations of individual $\Delta\text{CoVaR}_{q_1,t}^{s|j}$ and Column "Std.CS" reports the standard deviations of the mean of each individual $\Delta\text{CoVaR}_{q_1,t}^{s|j}$ measure.

| q_2 | | Mean(%) | Std.TS | Std.CS | Max(%) | Min(%) |
|-------|-----------|---------|--------|--------|---------|--------|
| 5% | Skew- t | 178.3 | 31.1 | 36.5 | 416.7 | 70.0 |
| | EVT | 139.0 | 41.9 | 41.8 | 534.1 | -21.4 |
| | Empirical | 161.4 | 79.4 | 40.4 | 491.5 | -17.8 |
| 1% | Skew- t | 303.3 | 59.7 | 70.2 | 748.5 | 109.0 |
| | EVT | 306.9 | 80.4 | 96.9 | 1172.5 | 48.3 |
| 0.01% | EVT | 1409.1 | 897.6 | 1332.6 | 14333.6 | 89.4 |

Chapter 7

Conclusion

On the theoretical side, we derived a limit expression for the conditional probability given that one of the components of a bivariate random vector is extreme, assuming a sub-family of skew-elliptical distributions with regularly varying tails as the underlying model. This extends results in Abdous et al. [2005a] for elliptically symmetric vectors. We also developed a semi-parametric EVT method to estimate this conditional probability using the above-mentioned asymptotic result. The main advantage of EVT-based estimators is that they preserve useful information in the tail data without restricting the behaviour of the central part. This methodology allows to assess extreme risk in asymmetric financial markets. Through two financial applications, we demonstrated that how our method can be applied flexibly in different contents.

In this project, we assume that the underlying distribution has a multivariate regularly varying tail. It would be of interest to develop a methodology to handle light-tailed data with rapidly-varying tails. For financial time series, this would apply to less frequently sampled data such as weekly or monthly returns. Additionally, one assumption for Theorem 4.1.3 is closure under marginalization. However, it is still not clear what conditions on skew-elliptical distributions could guarantee this assumption.

Regarding financial applications, several more topics can be explored. Firstly, DiTraglia and Gerlach [2013] claim that the extreme conditional exceedance probability contains important information for risk-averse investors which can be a valuable tool to select portfolios. Our method provides a more accurate estimation of the extreme conditional exceedance probability, and hence can be used to further verify their conclusion. Secondly, it is interesting to investigate the financial contagion problem in international

equity markets, similar to Kenourgios et al. [2011]. Instead of adopting all observations, our method focuses more on the tail data and is more suitable for analysing extreme events such as financial crises.

Bibliography

- B. Abdous, A.L. Fougères, and K. Ghoudi. Extreme behaviour for bivariate elliptical distributions. *Canadian Journal of Statistics*, 33:317–334, 2005a.
- B. Abdous, C. Genest, and B. Rémillard. Dependence properties of meta-elliptical distributions. In P. Duchesne and B. RAMillard, editors, *In Statistical Modeling and Analysis for Complex Data Problems*, pages 1–15. Springer, 2005b.
- V.V. Acharya, L.H. Pedersen, T. Philippon, and M. Richardson. Measuring systemic risk. *NYU Working Paper*, 2010.
- C. Adcock, M. Eling, and N. Loperfido. Skewed distributions in finance and actuarial science: a review. *The European Journal of Finance*, iFirst: 1–29, 2012.
- T. Adrian and M.K. Brunnermeier. CoVaR. Technical report, National Bureau of Economic Research, 2011.
- L. Alles and L. Murray. Rewards for downside risk in Asian markets. *Journal of Banking & Finance*, 37:2501–2509, 2013.
- R. Aloui, M.S.B.N. Aissa, and D.K. Nguyen. Global financial crisis, extreme interdependences, and contagion effects: The role of economic structure? *Journal of Banking & Finance*, 35:130–141, 2011.
- A. Azzalini and A. Capitanio. Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t distribution. *Journal of the Royal Statistical Society: Series B*, 65:367–389, 2003.
- K.H. Bae, G.A. Karolyi, and R.M. Stulz. A new approach to measuring financial contagion. *Review of Financial Studies*, 16:717–763, 2003.

Bibliography

- A.A. Balkema and L. De Haan. Residual life time at great age. *Annals of Probability*, 2:792–804, 1974.
- B. Basrak, R.A. Davis, and T. Mikosch. A characterization of multivariate regular variation. *Annals of Applied Probability*, 12:908–920, 2002.
- J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of Extremes: Theory and Applications*. John Wiley & Sons, 2006.
- J. Beran and G. Mainik. On estimating extremal dependence structures by parametric spectral measures. *Statistical Methodology*, 21:1–22, 2014.
- N.H. Bingham, C.M. Goldie, and J. L. Teugels. *Regular Variation (Encyclopedia of Mathematics and its Applications)*. Cambridge University Press, 1987.
- M.O. Boldi and A.C. Davison. A mixture model for multivariate extremes. *Journal of the Royal Statistical Society: Series B*, 69:217–229, 2007.
- M.D. Branco and D.K. Dey. A general class of multivariate skew-elliptical distributions. *Journal of Multivariate Analysis*, 79:99–113, 2001.
- X. Burtschell, J. Gregory, and J.P. Laurent. A comparative analysis of CDO pricing models under the factor copula framework. *Journal of Derivatives*, 16:9–37, 2009.
- J.J. Cai, J.H.J. Einmahl, and L. De Haan. Estimation of extreme risk regions under multivariate regular variation. *The Annals of Statistics*, 39:1803–1826, 2011.
- J.A. Chan-Lau, D.J. Mathieson, and J.Y. Yao. Extreme contagion in equity markets. *IMF Staff Papers*, 51:386–408, 2004.
- B.Y. Chang, P. Christoffersen, and K. Jacobs. Market skewness risk and the cross section of stock returns. *Journal of Financial Economics*, 107:46–68, 2013.
- A. Clauset, C.R. Shalizi, and M. EJ. Newman. Power-law distributions in empirical data. *SIAM review*, 51:661–703, 2009.

- S.G. Coles. *An Introduction to Statistical Modelling of Extreme Values*. Springer Series in Statistics, 2001.
- S.G. Coles and J. Tawn. Statistical methods for multivariate extremes: an application to structural design (with discussion). *Journal of Applied Statistics*, 43:1–48, 1994.
- J. Conrad, R.F. Dittmar, and E. Ghysels. Ex ante skewness and expected stock returns. *The Journal of Finance*, 68:85–124, 2013.
- R.C Dahiya and J. Gurland. Goodness of fit tests for the gamma and exponential distributions. *Technometrics*, 14:791–801, 1972.
- B. Das and S.I. Resnick. Conditioning on an extreme component: Model consistency and regular variation on cones. *Bernoulli*, 17:226–252, 2011.
- L. de Haan and A. Ferreira. *Extreme Value Theory. An Introduction*. Springer-Verlag, 2006.
- L. de Haan and E. Omey. Integrals and derivatives of regularly varying functions in \mathbb{R}^d and domains of attraction of stable distributions II. *Stochastic Processes and their Applications*, 16:157–170, 1983.
- L. de Haan and S.I. Resnick. Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie verw. Gebiete*, 40:317–337, 1977.
- L. de Haan and S.I. Resnick. On regular variation of probability densities. *Stochastic Processes and their Applications Applications*, 25:83–93, 1987.
- F.X. Diebold, T. Schuermann, and J.D. Stroughair. Pitfalls and opportunities in the use of extreme value theory in risk management. *Journal of Risk Finance*, 1:30–35, 2000.
- F.J. DiTraglia and J.R. Gerlach. Portfolio selection: An extreme value approach. *Journal of Banking & Finance*, 37:305–323, 2013.
- E.F. Eastoe, J.E. Heffernan, and J.A. Tawn. Nonparametric estimation of the spectral measure, and associated dependence measures, for multivariate extreme values using a limiting conditional representation. *Extremes*, 17:25–43, 2014.

- J.H.J. Einmahl and J. Segers. Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution. *The Annals of Statistics*, 37:2953–2989, 2009.
- J.H.J. Einmahl, L. de Haan, and V.I. Piterbarg. Nonparametric estimation of the spectral measure of an extreme value distribution. *The Annals of Statistics*, 29:1401–1423, 2001.
- P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer, 1997.
- R. Engle. Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20:339–350, 2002.
- B.Q. Fang. The skew elliptical distributions and their quadratic forms. *Journal of Multivariate Analysis*, 87:298–314, 2003.
- K.T. Fang, S. Kotz, and K.W. Ng. *Symmetric Multivariate and Related Distributions*. Chapman and Hall, 1990.
- R.A. Fisher and L.H.C Tippet. Limiting forms of the frequency distribution of the largest or smallest member of a sample. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 24, pages 180–190. Cambridge Univ Press, 1928.
- X. Gabaix and R. Ibragimov. Rank- $1/2$: a simple way to improve the OLS estimation of tail exponents. *Journal of Business & Economic Statistics*, 29:24–39, 2011.
- G. Girardi and E.A. Tolga. Systemic risk measurement: Multivariate GARCH estimation of CoVaR. *Journal of Banking & Finance*, 37:3169–3180, 2013.
- B. Gnedenko. Sur la distribution limite du terme maximum d’une serie aleatoire. *Annals of Mathematics*, 44:423–453, 1943.
- S. Guillotte, F. Perron, and J. Segers. Non-parametric Bayesian inference on bivariate extremes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73:377–406, 2011.

Bibliography

- C.R. Harvey and A. Siddique. Conditional skewness in asset pricing tests. *The Journal of Finance*, 55:1263–1295, 2000.
- J. Heffernan and S.I. Resnick. Limit laws for random vectors with an extreme component. *Annals of Applied Probability*, 17:537–571, 2007.
- J. Heffernan and J. Tawn. A conditional approach for multivariate extreme values. *Journal of the Royal Statistical Society, Series B*, 66:497–546, 2004.
- B. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3:1163–1174, 1975.
- R. Huisman, K.G. Koedijk, C.L.M. Kool, and F. Palm. Tail-index estimates in small samples. *Journal of Business and Economic Statistics*, 19:208–216, 2001.
- H. Hult and F. Lindskog. Multivariate extremes, aggregation and dependence in elliptical distributions. *Advances in Applied Probability*, 34:587–608, 2002.
- A.F. Jenkinson. The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81:158–171, 1955.
- H. Joe. *Multivariate Models and Dependence Concepts*. Chapman & Hall, 1997.
- D. Kenourgios, A. Samitas, and N. Paltalidis. Financial crises and stock market contagion in a multivariate time-varying asymmetric framework. *Journal of International Financial Markets, Institutions and Money*, 21:92–106, 2011.
- F. Longin and B. Solnik. Extreme correlation of international equity markets. *The Journal of Finance*, 56:649–676, 2001.
- Y. Ma, .M.G. Genton, and A.A. Tsiatis. Locally efficient semiparametric estimators for generalized skew-elliptical distributions. *Journal of the American Statistical Association*, 100:980–989, 2005.

Bibliography

- R.A. Maronna and R.H. Zamar. Robust estimates of location and dispersion for high-dimensional datasets. *Technometrics*, 44:307–317, 2002.
- A. J. McNeil and R. Frey. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7:271–300, 2000.
- A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management: Concepts, Techniques, Tools*. Princeton University Press, Princeton, 2005.
- T. Nguyen and G. Samorodnitsky. Multivariate tail estimation with application to analysis of CoVaR. *Astin Bulletin*, 43:245–270, 2013.
- A. Peiro. Skewness in financial returns. *Journal of Banking & Finance*, 23:847–862, 1999.
- J. Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3:119–131, 1975.
- S.H. Poon, M. Rockinger, and J. Tawn. Extreme value dependence in financial markets: diagnostics, models, and financial implications. *Review of Financial Studies*, 17:581–610, 2004.
- S.I. Resnick. *Extreme Values, Regular Variation, and Point Processes*. Springer, 1987.
- S.I. Resnick. *Heavy-tail Phenomena. Probabilistic and Statistical Modeling*. Springer, 2006.
- M. Rockinger and E. Jondeau. Entropy densities with an application to autoregressive conditional skewness and kurtosis. *Journal of Econometrics*, 106:119–142, 2002.
- H. Rootzén and N. Tajvidi. Multivariate generalized pareto distributions. *Bernoulli*, 12:917–930, 2006.
- P.J. Rousseeuw and K.V. Driessen. A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41:212–223, 1999.

- A. J. Stam. *Regular variation in \mathbb{R}^d and the Abel-Tauber theorem*. Mathematisch Instituut, 1977.
- D. Trudel. Tail dependence of hedge funds. Master's thesis, Swiss Federal Institute of Technology Zurich, 2008.
- R. Von Mises. La distribution de la plus grande de n valeurs. *Reviews in Mathematical Union Interbalcanique*, 1, 1936.
- J. Wang, J. Boyer, and M.G. Genton. A skew-symmetric representation of multivariate distributions. *Statistica Sinica*, 14:1259–1270, 2004.

Appendix A

Proofs

A.1 Proof of Lemma 3.3.2

Proof Denoting $\bar{\mathbf{Z}} = C^{-1}\mathbf{Z}$, the density function \bar{h} of $\bar{\mathbf{Z}}$ can be derived from transformation of the density function h of \mathbf{Z} :

$$\begin{aligned}\bar{h}(\mathbf{z}) &= 2f(C^{-1}\mathbf{z}; \boldsymbol{\xi}, \Omega)G(\omega(C^{-1}\mathbf{z} - \boldsymbol{\xi}))|C|^{-1} \\ &= 2\frac{c_d}{|\Omega|^{1/2}|C|}\tilde{f}((C^{-1}\mathbf{z} - \boldsymbol{\xi})^T\Omega^{-1}(C^{-1}\mathbf{z} - \boldsymbol{\xi}))G(\omega(C^{-1}\mathbf{z} - \boldsymbol{\xi})) \\ &= 2\frac{c_d}{|C^T\Omega C|^{1/2}}\tilde{f}((\mathbf{z} - C\boldsymbol{\xi})^T(C^T\Omega C)^{-1}(\mathbf{z} - C\boldsymbol{\xi}))G(\omega_0(\mathbf{z} - C\boldsymbol{\xi})),\end{aligned}$$

where $\omega_0(\mathbf{z} - C\boldsymbol{\xi}) = \omega(C^{-1}(\mathbf{z} - C\boldsymbol{\xi}))$, and it is obvious that ω_0 is also an odd function. \blacksquare

A.2 Proof of Proposition 4.1.1

Proof We begin by showing that condition (2.5) holds, and then derive the form of limit function $q(\mathbf{z})$. The validity of uniformity condition (2.6) is discussed in the end.

It suffices to consider the case $\boldsymbol{\xi} = \mathbf{0}$ as a shift in location does not affect the index and spectral measure of regularly varying random vectors; see Lemma 2.2 in Hult and Lindskog [2002]. Letting $\tilde{\mathbf{Z}} \sim Ell_2(\mathbf{0}, \bar{\Omega}, \tilde{f})$ with density \tilde{f} , Proposition 3.2.2 says that $\|\tilde{\mathbf{Z}}\| \stackrel{d}{=} \|\mathbf{Z}\|$. From (3.2), $\tilde{\mathbf{Z}} = RL^TS$, where R has density (3.10), $L = \begin{pmatrix} 1 & \rho \\ 0 & \sqrt{1 - \rho^2} \end{pmatrix}$ using Cholesky

A.2. Proof of Proposition 4.1.1

decomposition, and S has uniform distribution on \mathcal{S}^1 . We then have

$$\begin{aligned}
V(t) &:= \mathbb{P}(\|\mathbf{Z}\| > t) = \mathbb{P}(\|\tilde{\mathbf{Z}}\| > t) = \mathbb{P}(R(S^T L L^T S)^{1/2} > t) \\
&= \mathbb{P}(R(1 + \rho\sqrt{1 - \rho^2} \sin(2\Theta) + \rho^2 \cos(2\Theta))^{1/2} > t), \quad \Theta \sim Unif(0, 2\pi) \\
&= \int_0^{2\pi} \int_{\frac{t}{\sqrt{A(\theta)}}}^{\infty} \frac{1}{2\pi} f_R(r) dr d\theta, \quad \text{where } A(\theta) := 1 + \rho\sqrt{1 - \rho^2} \sin(2\theta) + \rho^2 \cos(2\theta) \\
&= \int_0^{2\pi} \int_{\frac{t}{\sqrt{A(\theta)}}}^{\infty} c_d \tilde{f}(r^2) r dr d\theta \\
&= \int_0^{2\pi} \int_{(\frac{t}{\sqrt{A(\theta)}})^2}^{\infty} \frac{c_d}{2} \tilde{f}(s) ds d\theta \\
&= \int_0^{2\pi} \frac{c_d}{2} \bar{\bar{F}}(t^2/A(\theta)) d\theta, \quad \text{where } \bar{\bar{F}}(x) := \int_x^{\infty} \tilde{f}(u) du.
\end{aligned}$$

With density h of \mathbf{Z} in (3.8), the limit in (2.5) can be computed as

$$\begin{aligned}
q(\mathbf{z}) &= \lim_{t \rightarrow \infty} \frac{h(t\mathbf{z})}{t^{-2}V(t)} = \lim_{t \rightarrow \infty} \frac{2 c_d |\bar{\Omega}|^{-1/2} \tilde{f}(Q(t\mathbf{z})) G(w(t\mathbf{z}))}{t^{-2} \int_0^{2\pi} (c_d/2) \bar{\bar{F}}(t^2/A(\theta)) d\theta} \\
&= 4 |\bar{\Omega}|^{-1/2} G(w_{\infty}(\mathbf{z})) \lim_{t \rightarrow \infty} \frac{t^2 \tilde{f}(t^2 Q(\mathbf{z}))}{\int_0^{2\pi} \bar{\bar{F}}(t^2/A(\theta)) d\theta} \\
&= 4 |\bar{\Omega}|^{-1/2} G(w_{\infty}(\mathbf{z})) Q^*(\mathbf{z})^{-(\nu+2)/2} \lim_{t \rightarrow \infty} \frac{t^2 \tilde{f}(t^2)/\bar{\bar{F}}(t^2)}{\int_0^{2\pi} \bar{\bar{F}}(t^2/A(\theta))/\bar{\bar{F}}(t^2) d\theta} \\
&= 4 |\bar{\Omega}|^{-1/2} G(w_{\infty}(\mathbf{z})) Q^*(\mathbf{z})^{-(\nu+2)/2} \lim_{t \rightarrow \infty} \frac{t^2 \tilde{f}(t^2)/\bar{\bar{F}}(t^2)}{\int_0^{2\pi} \bar{\bar{F}}(t^2/A(\theta))/\bar{\bar{F}}(t^2) d\theta} \\
&= 2\nu |\bar{\Omega}|^{-1/2} G(w_{\infty}(\mathbf{z})) Q^*(\mathbf{z})^{-(\nu+2)/2} \left[\int_0^{2\pi} A(\theta)^{\nu/2} d\theta \right]^{-1}.
\end{aligned}$$

The final line above follows from the assumption $\tilde{f} \in RV_{-(\nu+2)/2}$, which implies $\bar{\bar{F}} \in RV_{-\nu/2}$ and $t\tilde{f}(t)/\bar{\bar{F}}(t) \rightarrow \nu/2$ as $t \rightarrow \infty$ by Karamata's Theorem (see de Haan and Ferreira [2006]).

Based on Potter's Theorem (see Theorem 1.5.6 in Bingham et al. [1987]), for any chosen $c > 1$ and $\epsilon > 0$, there exist $C = C(c, \epsilon)$ such that $\bar{\bar{F}}(t^2/A(\theta))/\bar{\bar{F}}(t^2) \leq c \cdot \max\{(\frac{1}{A(\theta)})^{-\nu/2+\epsilon}, (\frac{1}{A(\theta)})^{-\nu/2-\epsilon}\}$ when $t^2/A(\theta) > C$ and $t^2 > C$. The limit in the denominator holds by the Dominated Convergence Theorem.

Finally, from (2.7) and using homogeneity of function q , we obtain

$$\begin{aligned}\psi(\mathbf{w}) &= \nu^{-1} q(\|\mathbf{z}\|\mathbf{w}) \|\mathbf{z}\|^{\nu+2} = \nu^{-1} q(\mathbf{w}), \quad \mathbf{w} = \mathbf{z}/\|\mathbf{z}\| \in \mathcal{S}^1 \\ &= 2 |\bar{\Omega}|^{-(1/2)} G(w_\infty(\mathbf{z})) Q^*(\mathbf{z})^{-(\nu+2)/2} \left[\int_0^{2\pi} A(\theta)^{\nu/2} d\theta \right]^{-1}.\end{aligned}$$

It remains to show validity of uniformity condition (2.6) for density h . Following the arguments in de Haan and Resnick [1987] (Section 3), the uniformity condition is fulfilled by the elliptical density f of $\tilde{\mathbf{Z}}$. In particular, we have for $\mathbf{z} \neq \mathbf{0}$

$$\lim_{t \rightarrow \infty} \frac{f(t\mathbf{z})}{t^{-2}V(t)} = q_0(\mathbf{z}) > 0 \quad \text{and} \quad \lim_{t \rightarrow \infty} \sup_{\mathbf{z} \in \mathcal{S}^1} \left| \frac{f(t\mathbf{z})}{t^{-2}V(t)} - q_0(\mathbf{z}) \right| = 0. \quad (\text{A.1})$$

It is straightforward to show that $q(\mathbf{z}) = 2G(\omega_\infty(\mathbf{z})) q_0(\mathbf{z})$. Now write

$$\begin{aligned}\sup_{\mathbf{z} \in \mathcal{S}^1} \left| \frac{h(t\mathbf{z})}{t^{-2}V(t)} - q(\mathbf{z}) \right| &= \sup_{\mathbf{z} \in \mathcal{S}^1} \left| \frac{2f(t\mathbf{z}) G(\omega(t\mathbf{z}))}{t^{-2}V(t)} - q(\mathbf{z}) \right| \\ &\leq \sup_{\mathbf{z} \in \mathcal{S}^1} \left| 2G(\omega(t\mathbf{z})) \left| \frac{f(t\mathbf{z})}{t^{-2}V(t)} - q_0(\mathbf{z}) \right| \right| \\ &\quad + \sup_{\mathbf{z} \in \mathcal{S}^1} 2q_0(\mathbf{z}) \left| G(\omega(t\mathbf{z})) - G(\omega_\infty(\mathbf{z})) \right|.\end{aligned}$$

Letting $t \rightarrow \infty$, the first term vanishes by (A.1) and since $G(\cdot)$ is bounded by one; Assumption (ii) gives convergence of the second term. \blacksquare

A.3 Proof of Lemma 4.1.2

Proof (i) \implies (ii). For any $z > 0$, if $\lim_{t \rightarrow \infty} \frac{\tilde{f}(tz)}{f(t)} = z^{-(\nu+d)/2}$, then

$$\begin{aligned}\lim_{t \rightarrow \infty} \frac{\mathbb{P}(R > tz)}{\mathbb{P}(R > t)} &= \lim_{t \rightarrow \infty} \frac{\int_{tz}^{\infty} \tilde{f}(r^2) r^{d-1} dr}{\int_t^{\infty} \tilde{f}(r^2) r^{d-1} dr} \\ &\stackrel{L'Hospital's\ rule}{=} \lim_{t \rightarrow \infty} \frac{z \tilde{f}(t^2 z^2) (tz)^{d-1}}{\tilde{f}(t^2) t^{d-1}} = z^d z^{-(\nu+d)} = z^{-\nu}.\end{aligned}$$

A.4. Proof of Theorem 4.1.3

(ii) \implies (i). Regular variation of R implies its density function $f_R \in RV_{-(\nu+1)}$ based on Karamata's Theorem. For $z > 0$,

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{\tilde{f}(t^2 z^2)}{\tilde{f}(t^2)} &= \lim_{t \rightarrow \infty} \frac{f_R(tz)(tz)^{1-d}}{f_R(t)t^{1-d}} = z^{-\nu-1} z^{1-d} = z^{-(\nu+d)} \\ &\implies \lim_{t \rightarrow \infty} \frac{\tilde{f}(tz)}{\tilde{f}(t)} = z^{-\frac{(\nu+d)}{2}}. \end{aligned}$$

(ii) \iff (iii). Let $\tilde{\mathbf{Z}} = RL^T S \sim Ell_2(\boldsymbol{\xi}, \bar{\Omega}, \tilde{f})$, then $\|\mathbf{Z}\| \stackrel{d}{=} \|\tilde{\mathbf{Z}}\|$. Based on Theorem 4.3 in Hult and Lindskog [2002], R is regularly varying with index ν is equivalent to $\tilde{\mathbf{Z}}$ being regularly varying with index ν . This directly implies that

$$\lim_{t \rightarrow \infty} \frac{\mathbb{P}(\|\mathbf{Z}\| > tz)}{\mathbb{P}(\|\mathbf{Z}\| > t)} = \lim_{t \rightarrow \infty} \frac{\mathbb{P}(\|\tilde{\mathbf{Z}}\| > tz)}{\mathbb{P}(\|\tilde{\mathbf{Z}}\| > t)} = z^{-\nu}, \quad z > 0.$$

¶

A.4 Proof of Theorem 4.1.3

Proof

1. **(Multivariate regular variation of h)** Let h denote the density of \mathbf{Z} ; see (3.8). Let $\mathbf{1} = (1, 1)$ denote a bivariate vector of ones. We have

$$\lim_{t \rightarrow \infty} \frac{h(t\mathbf{z})}{h(t\mathbf{1})} = \lim_{t \rightarrow \infty} \frac{\tilde{f}(Q(t\mathbf{z})) G(w(t\mathbf{z} - \boldsymbol{\xi}))}{\tilde{f}(Q(t\mathbf{1})) G(w(t\mathbf{1} - \boldsymbol{\xi}))} \quad (\text{A.2})$$

$$= \left(\frac{Q^*(\mathbf{z})}{Q^*(\mathbf{1})} \right)^{-(\nu+2)/2} \frac{G(w_\infty(\mathbf{z}))}{G(w_\infty(\mathbf{1}))} =: \lambda(\mathbf{z}), \quad \mathbf{z} \neq \mathbf{0}. \quad (\text{A.3})$$

Note $\lambda(\mathbf{z}) > 0$ for $\mathbf{z} \neq \mathbf{0}$ and $\lambda(a\mathbf{z}) = a^{-(\nu+2)}\lambda(\mathbf{z})$ for $a > 0$. Hence density h is bivariate regularly varying with index $(\nu + 2) > 2$ and limit function λ . The first factor in the expression for λ comes from the underlying elliptical density, whereas the second factor is due to the skewing function.

2. **(Joint tail behaviour)** We next relate the tail behaviour of the skew-elliptical random vector \mathbf{Z} to that of the associated elliptical random

vector $\tilde{\mathbf{Z}} = (\tilde{X}, \tilde{Y}) \sim Ell_2(\boldsymbol{\xi}, \bar{\Omega}, \tilde{f})$ with density f . Let $V(t) = t^2 h(t\mathbf{1})$. Analogous to the arguments in Proposition 4.1.1, one can show that h satisfies conditions of Theorem 2.2.1 with limit functions λ . Hence, Theorem 2.2.1 and (2.4) with $B = [\mathbf{z}, \infty)$ (cf. de Haan and Omey [1983], Theorem 1) imply

$$\begin{aligned} \mathbb{P}(\mathbf{Z} > t\mathbf{z}) &\sim t^2 h(t\mathbf{1}) \iint_{\mathbf{z}}^{\infty} \lambda(\mathbf{u}) d\mathbf{u}, \quad t \rightarrow \infty, \quad \mathbf{z} \geq \mathbf{0}, \quad \mathbf{z} \neq \mathbf{0} \\ &= 2 f(t\mathbf{1}) G(w(t\mathbf{1} - \boldsymbol{\xi})) t^2 \iint_{\mathbf{z}}^{\infty} \lambda(\mathbf{u}) d\mathbf{u}. \end{aligned} \quad (\text{A.4})$$

Similarly, for the elliptical random vector, the limit function is given by $\lambda_0(\mathbf{z}) = (Q^*(\mathbf{z})/Q^*(\mathbf{1}))^{-(\nu+2)/2}$ and

$$\mathbb{P}(\tilde{\mathbf{Z}} > t\mathbf{z}) \sim t^2 f(t\mathbf{1}) \iint_{\mathbf{z}}^{\infty} \lambda_0(\mathbf{u}) d\mathbf{u}, \quad t \rightarrow \infty, \quad \mathbf{z} \geq \mathbf{0}, \quad \mathbf{z} \neq \mathbf{0}. \quad (\text{A.5})$$

Combining (A.4) and (A.5), and plugging in expressions for the limit functions λ and λ_0 gives

$$\begin{aligned} \frac{\mathbb{P}(\mathbf{Z} > t\mathbf{z})}{\mathbb{P}(\tilde{\mathbf{Z}} > t\mathbf{z})} &\sim \frac{2 G(w(t\mathbf{1} - \boldsymbol{\xi})) \iint_{\mathbf{z}}^{\infty} \lambda(\mathbf{u}) d\mathbf{u}}{\iint_{\mathbf{z}}^{\infty} \lambda_0(\mathbf{u}) d\mathbf{u}}, \quad t \rightarrow \infty, \quad \mathbf{z} \geq \mathbf{0}, \quad \mathbf{z} \neq \mathbf{0} \\ &\rightarrow \frac{2 \iint_{\mathbf{z}}^{\infty} Q^*(\mathbf{u})^{-(\nu+2)/2} G(w_{\infty}(\mathbf{u})) d\mathbf{u}}{\iint_{\mathbf{z}}^{\infty} Q^*(\mathbf{u})^{-(\nu+2)/2} d\mathbf{u}}. \end{aligned} \quad (\text{A.6})$$

3. (Marginals)

Since \mathbf{Z} is multivariate regularly varying in the sense of Definition 2.2.3, it follows from Theorem 1.1(i) in Basrak et al. [2002] that the marginal tails are regularly varying with index ν . Under the assumption that the marginals of \mathbf{Z} are also skew-elliptical with densities of the form (4.2), we have

$$\begin{aligned} \lim_{x \rightarrow \infty} \frac{\mathbb{P}(X > x)}{\mathbb{P}(\tilde{X} > x)} &\stackrel{\text{L'Hospital's rule}}{=} \lim_{x \rightarrow \infty} \frac{h_1(x)}{f_1(x - \xi_1)} \\ &\stackrel{(4.2)}{=} \lim_{x \rightarrow \infty} 2 G_0(\bar{w}_1(x - \xi_1)) = 2 G_0(\bar{w}_{1,\infty}(1)), \end{aligned}$$

and hence

$$\mathbb{P}(X > x) \sim 2 \mathbb{P}(\tilde{X} > x) G_0(\bar{w}_{1,\infty}(1)), \quad x \rightarrow \infty. \quad (\text{A.7})$$

A.4. Proof of Theorem 4.1.3

4. **(Conditional probability)** Consider $\mathbb{P}(Y > y \mid X > x)$ for $y \sim zx$ as $x \rightarrow \infty$ and $z > 0$. Setting $\mathbf{z} = (1, z)$ in (A.6), and putting the joint and marginal tail behaviour together gives

$$\lim_{x,y \rightarrow \infty} \mathbb{P}(Y > y \mid X > x) = K(z) \lim_{x,y \rightarrow \infty} \mathbb{P}(\tilde{Y} > y \mid \tilde{X} > x), \quad (\text{A.8})$$

where

$$K(z) = \frac{\iint_{(1,z)}^{\infty} Q^*(\mathbf{u})^{-(\nu+2)/2} G(w_{\infty}(\mathbf{u})) d\mathbf{u}}{G_0(\bar{w}_{1,\infty}(1)) \iint_{(1,z)}^{\infty} Q^*(\mathbf{u})^{-(\nu+2)/2} d\mathbf{u}}, \quad (\text{A.9})$$

and (\tilde{X}, \tilde{Y}) follows a bivariate elliptical distribution with density f . The proof is completed by using Theorem 1(i) in Abdous et al. [2005b] for the limit of $\mathbb{P}(\tilde{Y} > y \mid \tilde{X} > x)$. ¶

Appendix B

Tables

B.1 Simulation Studies

In this section, we further report two simulation results for different bivariate skew- t distributions as in Section 5. The objective is to show that 1) our method could produce similar result as in Abdous et al. [2005a] for the elliptical distribution, and 2) our estimation method may not work well when the tail decays fast due to poor estimation of the tail index, the same problem as in Abdous et al. [2005a].

Table B.1: Simulation results based on 1000 samples of size 1000 from a bivariate skew- t distribution with parameters $\xi = (0,0)$, $\alpha = (0,0)$, $\nu = 2$, $\rho = 0.5$ and $\omega = \text{diag}(1, 1)$ for the scale matrix. Each cell provides the average (standard deviation) of the estimates of $\eta(x, y)$ under various methods; see Section 5.5 for details. For $\hat{\eta}_{AFG}(x, y)$, $\hat{\eta}_1(x, y)$, $\hat{\eta}_2(x, y)$ and $\eta_{\text{lim}}(x, y)$ we used $z = \omega_1 y / \omega_2 x - \rho$ in the limit results. Values of x and y are chosen as the theoretical marginal quantiles with probability p , where p labels columns and rows.

| Quantile y | Quantile x | | | |
|--|----------------|---------------|---------------|---------------|
| | 97.5% | 99.0% | 99.9% | 99.99% |
| $\hat{\eta}_{AFG}(x, y)$ | 0.617 (0.089) | 0.448 (0.072) | 0.254 (0.028) | 0.207 (0.021) |
| $\hat{\eta}_1(x, y)$ | 0.641 (0.052) | 0.472 (0.051) | 0.278 (0.037) | 0.231 (0.032) |
| 97.5% $\hat{\eta}_2(x, y)$ | 0.640 (0.057) | 0.472 (0.057) | 0.279 (0.043) | 0.232 (0.038) |
| $\eta_{emp}(x, y)$ | 0.599 (0.102) | 0.437 (0.168) | * | * |
| $\eta(x, y) / \eta_{\text{lim}}(x, y)$ | 0.598/0.609 | 0.438/0.440 | 0.257/0.257 | 0.213/0.213 |
| $\hat{\eta}_{AFG}(x, y)$ | 0.790 (0.0801) | 0.617 (0.089) | 0.303 (0.040) | 0.220 (0.022) |
| $\hat{\eta}_1(x, y)$ | 0.810 (0.041) | 0.641 (0.052) | 0.328 (0.041) | 0.243 (0.034) |
| 99.0% $\hat{\eta}_2(x, y)$ | 0.808 (0.044) | 0.640 (0.057) | 0.329 (0.048) | 0.244 (0.040) |
| $\eta_{emp}(x, y)$ | 0.774 (0.088) | 0.602 (0.171) | * | * |
| $\eta(x, y) / \eta_{\text{lim}}(x, y)$ | 0.775/0.786 | 0.605/0.609 | 0.302/0.302 | 0.225/0.225 |
| $\hat{\eta}_{AFG}(x, y)$ | 0.973 (0.027) | 0.933 (0.045) | 0.617 (0.089) | 0.304 (0.041) |
| $\hat{\eta}_1(x, y)$ | 0.977 (0.010) | 0.942 (0.019) | 0.641 (0.052) | 0.330 (0.042) |
| 99.9% $\hat{\eta}_2(x, y)$ | 0.977 (0.010) | 0.942 (0.020) | 0.640 (0.057) | 0.330 (0.048) |
| $\eta_{emp}(x, y)$ | 0.972 (0.034) | 0.934 (0.082) | * | * |
| $\eta(x, y) / \eta_{\text{lim}}(x, y)$ | 0.970/0.972 | 0.930/0.932 | 0.609/0.609 | 0.304/0.304 |
| $\hat{\eta}_{AFG}(x, y)$ | 0.997 (0.007) | 0.993 (0.013) | 0.932 (0.045) | 0.617 (0.089) |
| $\hat{\eta}_1(x, y)$ | 0.998 (0.002) | 0.994 (0.003) | 0.941 (0.020) | 0.641 (0.052) |
| 99.99% $\hat{\eta}_2(x, y)$ | 0.998 (0.002) | 0.994 (0.004) | 0.940 (0.020) | 0.640 (0.057) |
| $\eta_{emp}(x, y)$ | 0.997 (0.011) | 0.993 (0.028) | * | * |
| $\eta(x, y) / \eta_{\text{lim}}(x, y)$ | 0.997/0.997 | 0.992/0.992 | 0.930/0.931 | 0.609/0.609 |

Table B.2: Simulation results based on 1000 samples of size 1000 from a bivariate skew- t distribution with parameters $\xi = (3,1)$, $\alpha = (1,-3)$, $\nu = 20$, $\rho = 0.5$ and $\omega = \text{diag}(2, 3)$ for the scale matrix. Each cell provides the average (standard deviation) of the estimates of $\eta(x, y)$ under various methods; see Section 5.5 for details. For $\hat{\eta}_{AFG}(x, y)$, $\hat{\eta}_1(x, y)$, $\hat{\eta}_2(x, y)$ and $\eta_{\text{lim}}(x, y)$ we used $z = \omega_1 y / \omega_2 x - \rho$ in the limit results. Values of x and y are chosen as the theoretical marginal quantiles with probability p , where p labels columns and rows.

| Quantile y | Quantile x | | | |
|--|---------------|---------------|---------------|---------------|
| | 97.5% | 99.0% | 99.9% | 99.99% |
| $\hat{\eta}_{AFG}(x, y)$ | 0.206 (0.027) | 0.178 (0.025) | 0.139 (0.023) | 0.119 (0.021) |
| $\hat{\eta}_1(x, y)$ | 0.475 (0.056) | 0.434 (0.054) | 0.370 (0.049) | 0.332 (0.045) |
| 97.5% $\hat{\eta}_2(x, y)$ | 0.568 (0.057) | 0.527 (0.056) | 0.460 (0.054) | 0.419 (0.052) |
| $\eta_{emp}(x, y)$ | 0.692 (0.094) | 0.604 (0.162) | * | * |
| $\eta(x, y) / \eta_{\text{lim}}(x, y)$ | 0.691/0.457 | 0.598/0.363 | 0.408/0.243 | 0.286/0.182 |
| $\hat{\eta}_{AFG}(x, y)$ | 0.276 (0.029) | 0.234 (0.028) | 0.176 (0.025) | 0.145 (0.023) |
| $\hat{\eta}_1(x, y)$ | 0.563 (0.060) | 0.513 (0.058) | 0.431 (0.054) | 0.380 (0.050) |
| 99.0% $\hat{\eta}_2(x, y)$ | 0.655 (0.057) | 0.606 (0.058) | 0.524 (0.056) | 0.471 (0.054) |
| $\eta_{emp}(x, y)$ | 0.821 (0.079) | 0.741 (0.153) | * | * |
| $\eta(x, y) / \eta_{\text{lim}}(x, y)$ | 0.821/0.684 | 0.739/0.553 | 0.535/0.355 | 0.381/0.260 |
| $\hat{\eta}_{AFG}(x, y)$ | 0.469 (0.036) | 0.394 (0.033) | 0.281 (0.030) | 0.220 (0.027) |
| $\hat{\eta}_1(x, y)$ | 0.739 (0.056) | 0.679 (0.059) | 0.569 (0.060) | 0.494 (0.057) |
| 99.9% $\hat{\eta}_2(x, y)$ | 0.812 (0.048) | 0.761 (0.053) | 0.660 (0.057) | 0.587 (0.058) |
| $\eta_{emp}(x, y)$ | 0.971 (0.036) | 0.943 (0.080) | * | * |
| $\eta(x, y) / \eta_{\text{lim}}(x, y)$ | 0.970/0.974 | 0.942/0.919 | 0.815/0.698 | 0.641/0.504 |
| $\hat{\eta}_{AFG}(x, y)$ | 0.650 (0.039) | 0.560 (0.038) | 0.404 (0.034) | 0.309 (0.031) |
| $\hat{\eta}_1(x, y)$ | 0.852 (0.043) | 0.800 (0.050) | 0.687 (0.059) | 0.600 (0.060) |
| 99.99% $\hat{\eta}_2(x, y)$ | 0.903 (0.034) | 0.862 (0.041) | 0.768 (0.052) | 0.689 (0.056) |
| $\eta_{emp}(x, y)$ | 0.997 (0.012) | 0.993 (0.027) | * | * |
| $\eta(x, y) / \eta_{\text{lim}}(x, y)$ | 0.996/0.999 | 0.992/0.995 | 0.955/0.929 | 0.857/0.772 |

B.2 CoVaR

In this section, we check robustness of the summary statistics for cross-sectional $\Delta\text{CoVaR}_{q_1,t}^{s|j}$. Firstly, instead adopting EVT method to calculate $\text{VaR}_{q_2,t}^j$ of each institution, we directly model \hat{Z}_t^j using the skew- t distribution and take the quantile of skew- t as the estimate of $\text{VaR}_{q_2,t}^j$. Secondly, we assume that $\{\hat{\mathbf{Z}}_t\}$ follows a bivariate skew- t distribution for the benchmark case, and then apply the same procedure to compute CoVaR for the innovations $\{\hat{\mathbf{Z}}_t\}$ using the fitted model.

Generally speaking, the summary statistics are similar under different measurement approaches. A more detailed look reveals that $\text{VaR}_{q_2,t}^j$ estimated by EVT methods leads to, on average, a higher $\Delta\text{CoVaR}_{q_1,t}^{s|j}$, indicating that the parametric model (skew- t) is conservative in modelling tail events when quantile q_2 is small.

Table B.3: Summary statistics for cross-sectional $\Delta\text{CoVaR}_{q_1,t}^{s|j}$ for all institutions during sample period from June 1, 2006 to May 31, 2007. $\text{VaR}_{q_2,t}^j$ is estimated by assuming that \hat{Z}_t^j follows the skew- t distribution, and $\text{CoVaR}_{q_1,t}^{s|b_j}$ is estimated by assuming that $\{\hat{\mathbf{Z}}_t\}$ follows a bivariate skew- t distribution.

| q_2 | | Mean(%) | Std.TS | Std.CS | Max(%) | Min(%) |
|-------|-----------|---------|--------|--------|--------|--------|
| 5% | Skew- t | 172.8 | 29.8 | 25.9 | 366.2 | 85.7 |
| | EVT | 133.0 | 39.5 | 39.9 | 510.1 | -14.2 |
| | Empirical | 164.3 | 74.1 | 30.3 | 370.6 | 18.0 |
| 1% | Skew- t | 285.6 | 61.7 | 55.5 | 717.2 | 126.8 |
| | EVT | 285.2 | 80.4 | 81.3 | 1182.3 | 26.4 |
| 0.01% | EVT | 1110.3 | 457.8 | 573.9 | 8557.5 | 153.2 |

B.2. CoVaR

Table B.4: Summary statistics for cross-sectional $\Delta\text{CoVaR}_{q_1,t}^{s|j}$ for all institutions during sample period from June 1, 2006 to May 31, 2007. $\text{VaR}_{q_2,t}^j$ is estimated using EVT method described in McNeil and Frey [2000], and $\text{CoVaR}_{q_1,t}^{s|b_j}$ is estimated by assuming that $\{\hat{\mathbf{Z}}_t\}$ follows a bivariate skew- t distribution.

| q_2 | | Mean(%) | Std.TS | Std.CS | Max(%) | Min(%) |
|-------|-----------|---------|--------|--------|---------|--------|
| 5% | Skew- t | 178.7 | 25.6 | 28.9 | 360.0 | 87.5 |
| | EVT | 140.5 | 38.4 | 40.9 | 487.1 | -9.9 |
| | Empirical | 161.1 | 76.1 | 32.0 | 370.6 | -15.6 |
| 1% | Skew- t | 304.1 | 54.6 | 63.7 | 681.8 | 120.2 |
| | EVT | 308.8 | 77.0 | 96.6 | 1278.3 | 53.8 |
| 0.01% | EVT | 1437.0 | 918.7 | 1368.2 | 14297.4 | 91.7 |

Table B.5: Summary statistics for cross-sectional $\Delta\text{CoVaR}_{q_1,t}^{s|j}$ for all institutions during sample period from June 1, 2006 to May 31, 2007. $\text{VaR}_{q_2,t}^j$ is estimated by assuming that \hat{Z}_t^j follows the skew- t distribution, and $\text{CoVaR}_{q_1,t}^{s|b_j}$ is estimated empirically.

| q_2 | | Mean(%) | Std.TS | Std.CS | Max(%) | Min(%) |
|-------|-----------|---------|--------|--------|--------|--------|
| 5% | Skew- t | 172.2 | 34.3 | 31.8 | 383.0 | 79.1 |
| | EVT | 131.9 | 42.3 | 39.4 | 559.0 | -9.2 |
| | Empirical | 164.5 | 77.9 | 38.8 | 491.5 | 14.3 |
| 1% | Skew- t | 284.5 | 66.4 | 60.1 | 709.6 | 120.3 |
| | EVT | 283.0 | 83.9 | 79.3 | 1285.0 | 33.9 |
| 0.01% | EVT | 1099.5 | 458.5 | 558.4 | 9251.0 | 168.1 |