Extreme Value Approach to CoVaR Estimation

by

Menglin Zhou

B.Sc., Sun Yat-sen University, 2017

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)

August 2019

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the thesis entitled:

Extreme Value Approach to CoVaR Estimation

submitted by **Menglin Zhou** in partial fulfillment of the requirements for the degree of **Master of Science** in **Statistics**.

Examining Committee:

Natalia Nolde, Statistics
Supervisor

Harry Joe
Additional Examiner

Abstract

The global financial crisis of 2007-2009 revealed the importance of systemic risk: the risk that may destabilize the global economy due to financial contagion. Accurate assessment of systemic risk would not only enable regulators to introduce suitable policies to mitigate the risk, but also allow individual institutions to monitor and mitigate their vulnerability. An effective measure of systemic risk should be able to capture the co-movements between a financial system (or market) and individual financial institutions. One popular measure of systemic risk is CoVaR. In this thesis, a methodology is proposed to compute dynamic forecasts of CoVaR semi-parametrically within the classical framework of multivariate extreme value theory (EVT). According to the definition, CoVaR can be viewed as a high quantile of a conditional distribution where the conditioning event corresponds to large losses of an institution. The idea of our methodology is to relate this conditional distribution to the tail dependence function. We develop an EVT-based framework to estimate CoVaR statically by combining parametric modelling of the tail dependence function to address the issue of data sparsity in the joint tail regions and semi-parametric univariate tail estimation techniques. The performance of the methodology is illustrated via simulation studies and real data examples.

Lay Summary

One of the popular measures of systemic risk is Conditional Value-at-Risk (CoVaR), which can capture co-movements between a financial system (or market) and individual financial institutions. CoVaR is defined as a high quantile of the conditional distribution of system proxy such as a market index conditional on the event that an institution is experiencing a large loss in excess of a high quantile or the so-called Value-at-Risk (VaR). In view of data sparsity and model uncertainty when dealing with extreme events, empirical estimates and fully parametric statistical inference are not reliable and an effective way is to rely on the asymptotic approximations in the spirit of extreme value theory (EVT). In this sense, this thesis develops a flexible EVT-based framework to estimate CoVaR semi-parametrically by combining parametric modelling of the tail dependence function to address the issue of data sparsity in the joint tail regions and nonparametric univariate tail estimation techniques.

Preface

This thesis is original, unpublished work by the author, Menglin Zhou, under the supervision of Professor Natalia Nolde. The research idea in Chapter 3 was proposed by Professors Natalia Nolde and Chen Zhou. All simulations and analyses were designed and carried out by the author.

Table of Contents

List of Tables

List of Figures

Glossary

ABias	*Asymptotic Bias*
AFL	*AFLAC INC*
ALL	*ALLSTATE CORP*
AMSE	*Asymptotic Mean Squared Error*
AVar	*Asymptotic Variance*
AXP	*AMERICAN EXPRESS CO*
BAC	*BANK OF AMERICA CORP*
BEN	*FRANKLIN RESOURCES INC*
BK	*BANK NEWYORK INC*
CEV	*Conditional Extreme Value*
CLT	*Central Limit Theorem*
CoVaR	*Conditional Value-at-Risk*
df	*distribution function*
DJUSFN	*Dow Jones US Financials Index*
DOA	*Domain of Attraction*
EST	*Extended Skew-t Distribution*
EVD	*Extreme Value Distribution*
EVT	*Extreme Value Theory*
GP	*Generalized Pareto Distribution*
GS	*GOLDMAN SACHS GROUP INC*
HR	*Hüsler-Reiss*

MES	*Marginal Expected Shortfall*
MEV	*Multivariate Extreme Value*
POT	*Peaks Over Threshold*
RMSE	*Root Mean Squared Error*
SES	*Systemic Expected Shortfall*
ST	*Skew-t Distribution*
TD	*Tail Dependence*
TROW	*T ROWE PRICE GROUP INC*
VaR	*Value-at-Risk*

Acknowledgments

First and foremost I would like to thank my supervisor Natalia Nolde. It is quite lucky to be her student, as she is not only knowledgable in statistics, but also patient with students. I appreciate all her contributions of time, ideas, and funding to my research. The joy and enthusiasm she has for her research inspire me to further pursue a Ph.D. degree. Natalia, many thanks for your kind support and guidance.

I would also love to thank my family for all their love. I am grateful for my mom. Her unfailing love and unconditional support in all of my pursuits are invaluable. I would like to thank my boyfriend Xiaotian Zhan for taking care of me when I suffer. His understanding encourage me to move forward.

Finally, I am grateful to my friends in UBC and SYSU. Thanks for their inspirations when I came cross research problems. Thanks for their accompany when I felt lonely in this foreign land. I have been very fortunate to be friend with so many incredible people.

Chapter 1

Introduction

The global financial crisis of 2007-2009 alerted us to the importance of systemic risk: the risk or the possibility of breakdowns in an entire system, as opposed to breakdowns in individual parts or components, that can be contained without harming the entire system (Kaufman and Scott [2003]). Systemic risk is argued to be a particular feature of financial systems and may have significant adverse effects on the global economy due to financial contagion (De Bandt and Hartmann [2000]). The spillover losses to the U.S. and some foreign banks when the Herstatt Bank in Germany failed and was closed by the authorities in 1974 are typical examples. To mitigate the risk spillovers and keep the stability of the financial system, there is a clear need for effective measurement of systemic risk.

Kaufman and Scott [2003] point out that systemic risk is evidenced by the co-movements (correlation) among most or all the components in an entire system. It is followed that an effective measure of systemic risk should be able to capture co-movements between a financial system (or market) and individual financial institutions. Benoit et al. [2017] give a comprehensive survey of systemic risk, reviewing measures of systemic risk and connecting them to the current regulatory debate. Acharya et al. [2017] calculate Marginal Expected Shortfall (MES) and Systemic Expected Shortfall (SES) by using equity returns of a financial institution. MES is an institution's average loss when the financial system is in its left tail, and SES extends MES by calculating the weighted average of the institution's MES and its leverage. Acharya et al. [2012] propose SRISK measure, which corresponds to the expected capital shortfall of a given financial institution, conditional on a system crisis. The firms with the largest capital shortfall are considered the most systemically risky from the SRISK prospective.

In addition to the measures introduced above, this report considers another popular measure of systemic risk – the Conditional Value-at-Risk (CoVaR). Adrian and Brunnermeier [2011] define CoVaR as a high quantile of the conditional distribution of one institution conditional on the event that another institution is experiencing a large loss being at a high quantile or the so-called Value-at-Risk VaR. VaR is a widely-used risk measure by financial institutions. For a random variable X, VaR at confidence level $1-p$ is defined as the $(1-p)$-quantile of the underlying distribution:

$$\mathrm{VaR}_X(1-p) = \inf_x \{\mathrm{P}(X \leq x) \geq 1-p\}, \quad p \in (0,1). \tag{1.1}$$

By considering a specific case where the first institution is the financial system, it is able to assess the impact of a financial institution's large losses to systemic risk.

Girardi and Ergün [2013] modify the definition of CoVaR by specifying the conditional event as a loss in excess of the VaR, rather than being exactly at the VaR level. Let $X \sim F_1$ and $Y \sim F_2$ denote, respectively, losses for a financial institution and a system proxy. Following the modified definition, the $\mathrm{CoVaR}_{Y|X}$ at confidence level $1-p$ is defined as:

$$\mathrm{P}\left(Y \geq \mathrm{CoVaR}_{Y|X}(1-p) | X \geq \mathrm{VaR}_X(1-p)\right) = p, \quad p \in (0,1). \tag{1.2}$$

This modification allows us to consider more severe distress events of institutions and to back-test CoVaR estimates with the tests used for VaR (Girardi and Ergün [2013]). In addition, Mainik and Schaanning [2014] show that this change leads to the dependence consistency of CoVaR, as it allows CoVaR to hold a monotonic relationship with the Pearson correlation coefficient for elliptical distributions. In other words, as the losses of an institution become more correlated with the financial system, its systemic risk will increase.

CoVaR is actually a high quantile of the conditional distribution where the conditioning event corresponds to large losses. From the definition of CoVaR, one possible way to estimate CoVaR is to estimate the conditional probability $\mathrm{P}\{Y \geq y | X \geq x\}$ directly. However, when x is large, empirical estimates of this conditional probability might not be sensible since there are too few, even no, observations falling in the region of interest. Moreover, the fully parametric statistical inference (see e.g. Girardi and Ergün [2013]) focuses on the central part of the data whereas actual interest is in the tails. In situations dealing with extreme events, such as risk management in insurance, finance, and hydrology, an effective approach is to rely on asymptotic models in the spirit of extreme value theory (EVT); see e.g., Embrechts et al. [1997].

A conditional extreme value (CEV) model was proposed by Heffernan and Tawn [2004], followed by Heffernan and Resnick [2007] and Das and Resnick [2011]. Classical multivariate extreme value (MEV) models have limitations in the application when interest is in a tail region rather than the joint tail region. The CEV model addresses this issue by conditioning on one component of the random vector and finding the limiting conditional distribution of the remaining components as the conditioning variable becomes large. Moreover, the CEV model relaxes the assumption of MEV model that the distribution of (X, Y) is in a multivariate domain of attraction (DOA) to that when only one of the marginal distributions is in the univariate DOA. Abdous et al. [2005] estimate $\mathrm{P}\{Y < y | X \geq x\}$ when x is large for the class of elliptical distribution; Nolde and Zhang [2018] extend their methodology to a more general class of skew-elliptical distributions, which can describe the asymmetry for asset pricing and risk management in finance and insurance, and put it into a semi-parametric EVT-based framework for CoVaR estimation.

Some limitations of the approach in Nolde and Zhang [2018] are the requirement of multivariate regular variation, which in particular imposes the restriction of the same tail index for both the institutional and system losses, and a somewhat restrictive parametric assumption on the extremal dependence structure. While in some cases these assumptions may be viable, in general, system losses tend to be lighter-tailed than those of individual institutions based on our empirical analysis, and exhibit a greater variety of dependence structures. Our aim is to explore a more flexible framework to address CoVaR estimation.

The idea of our methodology is to link the definition of CoVaR in (1.2) with the tail dependence function introduced in Nikoloulopoulos et al. [2009] and Joe et al. [2010]. Assume that the distribution function of the random pair (X, Y) is in the DOA of a bivariate extreme value distribution and random variable Y has a positive tail index. Following de Haan and Ferreira [2006], we have

$$\lim_{u \to 0} \frac{\mathrm{P}\left(Y > F_2^{-1}(1-uy), X > F_1^{-1}(1-ux)\right)}{u} = R(x, y), \quad x, y > 0. \tag{1.3}$$

where R is the upper tail dependence function. The background information on multivariate EVT including concept of domain of attraction and tail dependence function will be provided in Chapter 3. Moreover, we assume that there exists a constant such that

$$\mathrm{CoVaR}_{Y|X}(1-p) = \mathrm{VaR}_Y(1-p\ ), \tag{1.4}$$

which allows us to rewrite equation (1.2) as follows:

$$\frac{\mathrm{P}\left\{Y > \mathrm{VaR}_Y(1 - p\ \ ); X > \mathrm{VaR}_X(1 - p)\right\}}{p} = p: \tag{1.5}$$

With the positive dependence relationship between X and Y, we have $0 < \quad < 1$ and $\quad = \quad_p$ depends on the value of p. In an extreme value setting where the confidence level p is close to zero, combining (1.3) and (1.5) implies

$$\begin{aligned}&\frac{\mathrm{P}\left\{Y > \mathrm{VaR}_Y(1 - p\ \ ); X > \mathrm{VaR}_X(1 - p)\right\}}{p} \\ =\ &\frac{\mathrm{P}\left\{F_2(Y) > 1 - p\ \ ; F_1(X) > 1 - p\right\}}{p} \approx R(1;\ \ ):\end{aligned}$$

Hence, we are looking for an $\quad$ such that $R(1;\ \ ) = p$. An estimator of CoVaR can then be obtained based on equation (1.4). Due to data sparsity, efficiency can be gained by assuming a parametric model for the tail dependence function $R(x; y) = R(x; y;\ \ )$. We apply a method of moment to estimate the unknown parameters in $R(x; y;\ \ )$; see e.g., Einmahl et al. [2008] and Einmahl et al. [2012]. This moment estimator is shown to be consistent and asymptotically normal under weak conditions. The tail index parameter $\quad > 0$ for the system proxy Y can be estimated using the Hill estimator (Hill [1975]), in which the sampling fraction is automatically selected with a two-step subsample bootstrap method of Danielsson et al. [2001].

This report is organized as follows. Chapter 2 reviews relevant definitions and results that will be used as a basis for the proposed methodology. Chapter 3 give details of CoVaR estimation, and illustrate performance of the proposed methods via several simulation studies. Chapter 4 gives an application of the developed estimator to financial data. Finally, Chapter 5 presents some concluding remarks and outlines directions for future research. Proofs and additional figures are delegated to the Appendixes.

Chapter 2

Background

The aim of this chapter is to review relevant definitions and results from the literature that will be used in the sequel as a basis for the proposed methodology for CoVaR estimation, including extreme value theory, Hill estimation for the tail index and extreme quantile estimation.

2.1 Multivariate Extreme Value Theory

Multivariate extreme value theory (EVT) provides the main probabilistic framework that we will adopt. Intuitively, EVT deals with extreme events which occur with very small probability. The central result of EVT is the Fisher-Tippett theorem (see Theorem 2.1.1), which is developed in parallel with the central limit theorem (CLT). Suppose we have a sequence of independent and identically distributed (i.i.d.) non-degenerated random variables $X_1, X_2, \ldots X_n$. The general CLT concerns the limit laws of sample sums $X_1 + X_2 + \ldots + X_n$ when properly normalized and centered, whereas the Fisher-Tippett theorem investigates the limit laws of sample maxima $\max\{X_1, X_2, \ldots, X_n\}$ or $\min\{X_1, X_2, \ldots, X_n\}$ when properly normalized and centered (for more details, see Embrechts et al. [1997]).

The EVT has been widely used in many fields, especially in financial risk management, which is closely related to tail probabilities and quantiles. The ability to manage extreme financial risks (e.g., currency crises, stock market crashes and large bond defaults) effectively is in fact the ability to assess the extreme probabilities and quantiles accurately (Diebold et al.

[2000]). For example, in the credit or operational risk management, the goal is to determine the risk capital we require as a cushion against irregular losses. In the problem of portfolio selection, a safe criterion is to select the assets which minimize the probability of a return below a prespecified threshold return level. Traditional statistical methods that based on the entire dataset produce a good fit in central region but is biased to the assessment of tail regions, where fewer or even no observations fall. In this sense, it seems that EVT can help us to build more appropriate statistical models describing extreme events in financial risk management.

2.1.1 Univariate EVT

We begin by recalling the key results from univariate EVT. For more details, please refer to Resnick [1987] or de Haan and Ferreira [2006].

Definition 1. *Suppose $X_1, \ldots, X_n$ are i.i.d. random variables with distribution function (df) F. Let $M_n := \max\{X_1, \ldots, X_n\}$ denotes the sample maximum. The df F is said to belong to the (maximum) domain of attraction of df G (written $F \in \mathcal{D}(G)$) if*

(i) G is non-degenerate, and
(ii) There exist $a_n > 0$ and $b_n \in \mathbb{R}$ such that

$$P\left\{\frac{M_n - b_n}{a_n} \leq x\right\} = F^n(a_n x + b_n) \to G(x); \quad n \to \infty; \text{ for all } x \in \mathcal{C}(G);$$

where $\mathcal{C}(G)$ denotes the set of all continuity points of G.

In the CLT, with finite variance, the non-degenerate limit distribution is found to be the normal distribution. In the EVT, Fisher-Tippett theorem help us to identify the class of limit df G.

Theorem 2.1.1 (Fisher and Tippett [1928], Gnedenko [1943]). *Suppose $F \in \mathcal{D}(G)$. Then G is one of the following types[1]:*

1. *Type I, Gumbel:* $\Lambda(x) = \exp\{-e^{-x}\}; \quad x \in \mathbb{R};$
2. *Type II, Fréchet:* $\Phi_\alpha(x) = \begin{cases} 0; & x < 0 \\ \exp\{-x^{-\alpha}\}; & x \geq 0 \end{cases}; \quad \alpha > 0;$

[1] Two dfs $U(x)$ and $V(x)$ are of the same type if for some $A > 0$, $B \in \mathbb{R}$, we have $V(x) = U(Ax + B)$ for all x.

3. *Type III, Weibull:* $\Psi_\alpha(x) = \begin{cases} \exp\{-(-x)^\alpha\}, & x < 0 \\ 1, & x \geq 0 \end{cases}, \quad \alpha > 0.$

These three types of distributions are called the class of Extreme Value Distributions (EVD).

For statistical purposes, Von Mises [1936] and Jenkinson [1955] generalize and unify the above three types of extreme value distributions into one family, named generalized extreme value distribution ($GEV(\mu, \sigma, \xi)$), whose df is given by

$$G_\xi\left(\frac{x-\mu}{\sigma}\right) = \begin{cases} \exp\{-e^{-(x-\mu)/\sigma}\}, & \xi = 0, \\ \exp\left\{-\left(1+\xi\frac{x-\mu}{\sigma}\right)^{-1/\xi}\right\}, & \xi \neq 0, \quad 1+\xi\frac{x-\mu}{\sigma} > 0. \end{cases}$$

where $\mu \in \mathbb{R}$ is the location parameter, $\sigma > 0$ is the scale parameter and $\xi \in \mathbb{R}$ is termed as the shape parameter or tail index. It is clear that when $\xi > 0$, G_ξ is of the same type as the Fréchet df with $\alpha = 1/\xi$; when $\xi < 0$, G_ξ is of the same type as the Weibull df with $\alpha = -1/\xi$. And G_0 is of the same type as the Gumbel df.

Definition 2. *A non-degenerate df F is max-stable if for $X_1, \ldots, X_n$ i.i.d. with df F, there exist $a_n > 0$ and $b_n \in \mathbb{R}$ such that $M_n \stackrel{d}{=} a_n X_1 + b_n$, i.e., for each n, M_n and X_1 are of the same type.*

Clearly, the Gumbel df is max-stable with $a_n = 1$ and $b_n = -\log n$; the Fréchet df is max-stable with $a_n = n^{1/\alpha}$ $(\alpha > 0)$ and $b_n = 0$; the Weibull df is max-stable with $a_n = n^{1/\alpha}$ $(\alpha < 0)$ and $b_n = 0$.

In the application of risk management, the data often show a pattern of heavy tails, such as record-breaking insurance losses and financial log-returns. Heavy tail is a characteristic of phenomena where the probability of a huge value is relatively big. An effective tool for dealing with heavy-tail phenomena is the theory of regularly varying functions, so here we give some brief introductions about regular variation (for more details, see Resnick [2007]).

Definition 3. *A positive measurable function h on $(0, \infty)$ is regularly varying at infinity with index $\rho \in \mathbb{R}$ (written $h \in RV_\rho$) if for all $x > 0$,*

$$\lim_{t \to \infty} \frac{h(tx)}{h(t)} = x^\rho,$$

where ρ is the exponent of variation.

If $\alpha = 0$, we call the function slowly varying. For example, function $L(x) = \log x$ is a slowly varying function, as $\lim_{x \to \infty} \log(tx)/\log x = 1$. Moreover, if function $h \in RV_\alpha$, then $h(x) = L(x) \cdot x^\alpha$, where L is a slowly varying function.

With the definition above, we can see that the distributions we concerned in application of financial risk management are actually the distributions whose tails are regularly varying.

Definition 4. *A random variable X with df F is said to be regularly varying with index $\alpha > 0$ (written $1 - F \in RV_{-\alpha}$) if for every $x > 0$;*

$$\lim_{t \to \infty} \frac{P(X > tx)}{P(X > t)} = x^{-\alpha}.$$

Intuitively, a random variable is regularly varying when its associated distribution has a heavy tail which decays according to a power law with exponent $-\alpha$. In probability theory, the parameter $\alpha > 0$ is often called the tail index.

Example 2.1.1. The Student's t distribution with degree of freedom ν has a density function:

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma(\nu/2)} \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}}, \quad x \in \mathbb{R}.$$

It is easy to verify that $f \in RV_{-\nu-1}$. For every $x \in \mathbb{R}$, we have

$$\lim_{t \to \infty} \frac{1 - F_T(tx; \nu)}{1 - F_T(t; \nu)} = \lim_{t \to \infty} \frac{xf(tx)}{f(t)} = x^{-\nu},$$

which means that the tail of Student's t distribution with degree of freedom ν is regularly varying with tail index ν.

Theorem 2.1.2 (Gnedenko [1943])**.** *$F \in \mathcal{D}(\Phi_\alpha)$ if and only if $1 - F \in RV_{-\alpha}$. In this case, $F^n(a_n x) \to \Phi_\alpha(x)$ with $a_n = (1/(1-F))^{-1}(n)$.*

Theorem 2.1.2 identifies that Frchet domain of attraction characterizes heavy-tailed distributions with a regularly-varying upper tail. To be specific, a random variable that has a heavy-tailed df F should be regularly varying and the df F is in the Fréchet domain of attraction. This connection will help us to construct our semi-parametric framework of CoVaR estimation. Here, we give four examples that are in the Fréchet domain of attraction in Table 2.1[Embrechts et al. [1997]].

Table 2.1: Domain of attraction of the Fréchet distribution

Cauchy distribution	$f(x) = (\pi(1+x^2))^{-1}; \quad x \in \mathbb{R}.$ $a_n = n/\pi.$
Pareto distribution	$f(x) = \frac{\alpha k^{\alpha}}{x^{\alpha+1}}; \quad x \geq k; \quad k, \alpha > 0.$ $a_n = (kn)^{1/\alpha}.$
Burr distribution	$f(x) = \frac{ckx^{c-1}}{(1+x^c)^{k+1}}; \quad x > 0; \quad c, k > 0.$ $a_n = (n^{1/k} - 1)^{1/c}.$
Loggamma distribution	$f(x) = \frac{\alpha^{\beta}}{\Gamma(\beta)}(\ln x)^{\beta-1}x^{-\alpha-1}; \quad x > 1; \quad \alpha, \beta > 0.$ $a_n = \left((\Gamma(\beta))^{-1}(\ln n)^{\beta-1}n\right)^{1/\alpha}.$

2.1.2 Hill estimator for the tail index

In this section, we introduce the widely used Hill estimator (Hill [1975]) of the tail index in the quantile view (for detail, see Beirlant et al. [2006]). Let $X_1, X_2, \ldots, X_n$ be independent random variables with common df F.

Consider the mean excess function defined as:

$$e(t) = E(X - t | X > t).$$

The mean excess function can be estimated empirically by:

$$\hat{e}_n(t) = \frac{\sum_{i=1}^{n} X_i 1_{(t,\infty)}(X_i)}{\sum_{i=1}^{n} 1_{(t,\infty)}(X_i)} - t,$$

where $1_{(t,\infty)}(X_i)$ equals 1 if $X_i > t$, and 0 otherwise. Let $X_{n,1} \leq X_{n,2} \leq \ldots \leq X_{n,n}$ be the order statistics and take $t = X_{n,n-k}$, $k = 1, \ldots, n-1$. The estimator of the mean excess function at $X_{n,n-k}$ can be rewritten as

$$e_{k,n} = \hat{e}_n(X_{n,n-k}) = \frac{1}{k}\sum_{i=1}^{k} X_{n,n-i+1} - X_{n,n-k}. \tag{2.1}$$

In addition to being the estimator of $e(t)$ at some specific values of t, $e_{k,n}$ can also be interpreted as an estimator of the slope of the exponential Q-Q plot to the right of a reference point with coordinates $(-\log(k/n), X_{n-k,n})$.

Suppose the df F has a regularly varying tail, that is

$$1 - F(x) = x^{-1/\gamma} L_1(x), \quad \gamma > 0, \tag{2.2}$$

where L_1 is a slowly varying function and $1/\gamma$ is the tail index. In this case, the distribution is identified as Pareto-type distribution, which includes Burr distribution, Fréchet distribution and log-gamma distribution. Define $Q(p)$ as the quantile function: $Q(p) := \inf\{x : F(x) \geq p\}$. An equivalent formula of (2.2) is:

$$Q(1 - 1/x) = x^{\gamma} L_2(x), \quad \gamma > 0, \tag{2.3}$$

where $L_2(x)$ is also a slowly varying function and is linked with L_1 via de Bruyn conjugation[2] (see Proposition 2.5 in Beirlant et al. [2006]). Since for every slowly varying function $L_2(x)$, we have $\log L_2(1/p)/\log(1/p) \to 0$ as $p \to 0$ (see Proposition 2.4 in Beirlant et al. [2006]), then the Pareto-type distribution would satisfy

$$\lim_{p \to 0} \frac{\log Q(1-p)}{-\log p} = \gamma.$$

It means that the Pareto Q-Q plot, which is the exponential Q-Q plot based on the log-transformed data, is ultimately linear with slope γ near the largest observations and we could use the mean excess values introduced in equation (2.1) of the log-transformed data to estimate the slope. Then we could obtain the well-known Hill estimator as:

$$\hat{\gamma}_n(k) = \frac{1}{k} \sum_{i=1}^{k} \log X_{n,n-i+1} - \log X_{n,n-k}, \tag{2.4}$$

where $k = k_n \in \{1, \ldots, n\}$ is an intermediate sequence; that is, $k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$.

Assume the df F satisfies the second-order condition (de Haan and Resnick [1996]): there exists a function A_γ, not changing sign near infinity, such that for $x > 0$

$$\lim_{t \to \infty} \left(\frac{1 - F(tx)}{1 - F(t)} - x^{-1/\gamma} \right) \Big/ A_\gamma(t) = x^{-1/\gamma} \frac{x^{\rho/\gamma} - 1}{\rho/\gamma},$$

[2] If $l(x)$ is slowly varying function, then there exists an slowly varying function $l^{\#}(x)$, the de Bruyn conjugate of l, such that $l(x)l^{\#}(xl(x)) \to 1$, as $x \to \infty$. The de Bruyn conjugate is asymptotically unique in the sense that if also $\tilde{l}$ is slowly varying and $l(x)\tilde{l}(xl(x)) \to 1$, then $l^{\#} \sim \tilde{l}$. Furthermore $(l^{\#})^{\#} \sim l$.

where ≤ 0 is the second-order parameter. A reformulated version of this condition with the inverse function U of $1/(1-F)$ is: there exists a function A, not changing sign near infinity, such that

$$\lim_{t \to \infty} \frac{U(tx)/U(t) - x}{A(t)} = x \frac{x - 1}{}; \tag{2.5}$$

where function $|A|$ is regularly varying at infinity with index .

Using notation $A_{n,k} = A(\frac{n+1}{k+1})$, the asymptotic bias of the Hill estimator can be expressed as:

$$\text{ABias} \quad {}_n(k) \sim A_{n,k} \frac{1}{k} \sum_{i=1} \frac{i}{k+1} \sim \frac{A_{n,k}}{1 -}; \quad k, n \to \infty; \quad k/n \to 0.$$

The asymptotic variance of the Hill estimator is

$$\text{AVar} \quad {}_n(k) \sim \frac{{}^2}{k}; \quad k, n \to \infty; \quad k/n \to 0.$$

Notice that the bias will be small only if $A_{n,k}$ is small, which in turn is k to be small, and the variance will be small if k is large. The asymptotic normality of the Hill estimator can be expected when $k, n \to \infty$, $k/n \to 0$ and if $\sqrt{k} A_{n,k} \to 0$,

$$\sqrt{k} \quad {}_n(k)/ \;-1 \xrightarrow{d} N(0, 1).$$

An important step in Hill estimation is to choose the value of k. An ideal method is to select k by minimizing the asymptotic mean squared error of ${}_n(k)$, which is defined as

$$k_0(n) := \arg\min_k \text{AMSE}(n, k) = \arg\min_k \text{Asy}E \quad {}_n(k) - \quad {}^2. \tag{2.6}$$

The main problem is that there is an unknown parameter in (2.6). Due to this consideration, Danielsson et al. [2001] propose to select k by minimizing $\text{Asy}E\,(M_n(k) - 2(\;{}_n(k))^2)^2$, where $M_n(k) = \frac{1}{k} \sum_{i=1}^{k} (\log X_{n,n-i+1} - \log X_{n,n-k})^2$. They show that under some conditions (see Theorem 1 and 2 in Danielsson et al. [2001]), the k-value that minimizes $\text{AMSE}(n, k)$ and the k-value that minimizes $\text{Asy}\; E\,(M_n(k) - 2(\;{}_n(k))^2)^2$ are of the same general order (with respect to n). Moreover, in order to yield an AMSE estimator which is asymptotic to $\text{AMSE}(n, k)$, they use a two-step subsample bootstrap algorithm, which is shown in Algorithm 1 below.

Suppose drawing $\mathcal{A}_{n_i} = \{X_1, \ldots, X_{n_i}\} (n_i < n, i = 1, 2)$ from $\mathcal{A}_n = \{X_1, \ldots, X_n\}$ with replacement and $X_{n_i,1} \le \ldots \le X_{n_i,n_i}$ denote the order statistics of $\mathcal{A}_{n_i}$. Define

$$\gamma_{n_i}(k_i) = \frac{1}{k_i} \sum_{j=1}^{k_i} \left(\log X_{n_i,n_i-j+1} - \log X_{n_i,n_i-k_i} \right),$$

$$M_{n_i}(k_i) = \frac{1}{k_i} \sum_{j=1}^{k_i} \left(\log X_{n_i,n_i-j+1} - \log X_{n_i,n_i-k_i} \right)^2,$$

$$L(n_i, k_i) = E\left[\left(M_{n_i}(k_i) - 2\left(\gamma_{n_i}(k_i)\right)^2 \right)^2 \middle| \mathcal{A}_n \right].$$

Suppose $k_{i,0}(n_i)$ minimizes $L(n_i, k_i)$. Then the optimal choice of k in Danielsson et al. [2001] is defined as

$$\hat{k}_0(n) = \frac{\left(k_{1,0}(n_1)\right)^2}{k_{2,0}(n_2)} \left(\frac{(\log k_{1,0}(n_1))^2}{\left(2\log n_1 - \log k_{1,0}(n_1)\right)^2} \right)^{\frac{\log n_1 - \log k_{1,0}(n_1)}{\log n_1}}. \tag{2.7}$$

Algorithm 1 Two-step Subsample Bootstrap Method

1: Input step size h and the number of bootstrap samples B; set $N = \lceil (n - \sqrt{n})/h \rceil$;
2: **for** each integer $j \in [1, N]$ **do**
3: set $n_{1,j} = \sqrt{n} + (j-1)h$, and draw B bootstrap samples of size $n_{1,j}$ from $\mathcal{A}_n$;
4: calculate $L(n_{1,j}, k_1)$ at each integer $k_1 \in [1, n_{1,j}]$, and find the $k_{1,0}(n_{1,j})$ that minimizes $L(n_{1,j}, k_1)$;
5: set $n_{2,j} = (n_{1,j})^2/n$, and draw B bootstrap samples of size $n_{2,j}$ from $\mathcal{A}_n$;
6: calculate $L(n_{2,j}, k_2)$ at each integer $k_2 \in [1, n_{2,j}]$, and find the $k_{2,0}(n_{2,j})$ that minimizes $L(n_{2,j}, k_2)$;
7: Calculate $R(n_{1,j}) = L(n_{1,j}, k_{1,0})^2 / L(n_{2,j}, k_{2,0})$.
8: **end for**
9: set $n_1 = \arg\min_{n_{1,j}} R(n_{1,j})$, and repeat step 4, 5, 6 to get $k_{1,0}(n_1)$ and $k_{2,0}(n_2)$;
10: Calculate $\hat{k}_0(n)$ with equation (2.7).

From Danielsson et al. [2001], the tail estimator $\gamma_n(\hat{k}_0)$ based on $\hat{k}_0(n)$ above will have the same asymptotic efficiency as $\gamma_n(k_0)$ based on $k_0(n)$ defined in (2.6). All the details are shown in Theorems 4, 6 and Corollaries 5, 7 in Danielsson et al. [2001],

2.1.3 Extreme quantile estimation

Recall equation (1.1) in Section 1, the VaR at a confidence level $1-p$ is defined as the $(1-p)$-quantile, so estimating VaR is actually estimating a quantile. Following Beirlant et al. [2006], by assuming that the ultimate linearity of the Pareto quantile plot persists from the largest k observations till infinity, we can summarize the quantile plot $\left(-\log\frac{i-1}{n}, \log X_{n,n-i+1}\right)$, $i = 1, \ldots, k+1$ with line:

$$y = \log X_{n,n-k} + \hat{\gamma}_n(k)\left(x + \log\frac{k}{n}\right),$$

where $\hat{\gamma}_n(k)$ is the Hill estimator of γ (see equation (2.4)).

To get an estimator of $Q(1-p)$, take $x = -\log p$, and the estimator (first proposed by Weissman [1978]) is given by

$$\begin{aligned} \hat{Q}_{k,1-p} &= \exp\left(\log X_{n,n-k} + \hat{\gamma}_n(k)\left(-\log p + \log\frac{k}{n}\right)\right) \\ &= X_{n,n-k}\left(\frac{k}{np}\right)^{\hat{\gamma}_n(k)}. \end{aligned} \tag{2.8}$$

Denote the asymptotic expectation operator by E_1. When $p = p_n \to 0$ and $np_n \to c > 0$ Beirlant et al. [2006] conclude that,

$$E_1\left(\log\frac{\hat{Q}_{k,1-p}}{Q(1-p)}\right) \sim \frac{A_{n,k}}{1-\rho}\log\frac{k}{np} + A_{n,k}\frac{1-\left(\frac{np}{k}\right)^{-\rho}}{\rho}$$

and

$$\text{AVar}(\log\hat{Q}_{k,1-p}) \sim \frac{\gamma^2}{k}\left(1+\log^2\frac{k}{np}\right), \quad k, n\to\infty, \quad k/n\to 0.$$

Furthermore, when $k, n\to\infty$ and $k/n\to 0$ such that $\sqrt{k}E_1\left(\log\frac{\hat{Q}_{k,1-p}}{Q(1-p)}\right)\to 0$,

$$\sqrt{k}\left(1+\log^2\frac{k}{np}\right)^{-1/2}\left(\frac{\hat{Q}_{k,1-p}}{Q(1-p)}-1\right) \xrightarrow{d} N(0, \gamma^2).$$

2.1.4 Multivariate extreme value distributions

The problem of CoVaR estimation is inherently bivariate, so it is necessary for us to extend univariate EVT in Section 2.1.1 to the multivariate setting. For the univariate case, the ordering principle is clear and unambiguous, but ordering is not unique in multivariate setting. Barnet [1976] discusses several different categories of order relations for multivariate data, and the most popular one is called marginal ordering: for d-dimensional vectors $\mathbf{x} = (x_1, \ldots, x_d)$ and $\mathbf{y} = (y_1, \ldots, y_d)$, the relation $\mathbf{x} \le \mathbf{y}$ is defined as $x_j \le y_j$ for all $j = 1, \ldots, d$. With this ordering principle, the component-wise maximum of random vectors $\mathbf{x}$ and $\mathbf{y}$ is defined as $\mathbf{x} \vee \mathbf{y} := (x_1 \vee y_1, \ldots, x_d \vee y_d)$.

Definition 5. *Consider a sample of i.i.d d-dimensional observations* $\mathbf{Y}_i \in \mathbb{R}^d (i = 1, \ldots, n)$ *with df* F *and margins* $F_j, j = 1, \ldots, d$*. The sample maximum* $\mathbf{M}_n = (M_{n,1}, \ldots M_{n,d})$ *is defined as a vector of component-wise maxima (i.e.* $\mathbf{M}_n = \bigvee_{i=1}^{n} \mathbf{Y}_i$*). The multivariate df* F *is said to belong to the (maximum) domain of attraction of a multivariate df* G *(written as* $F \in \mathcal{D}_d(G)$*), if there exist* $\mathbb{R}^d$*-sequences* $\mathbf{a}_n > \mathbf{0}$ *and* $\mathbf{b}_n$ *such that for all continuity points* $\mathbf{y}$ *of* G*,*

$$\lim_{n \to \infty} P\left(\frac{\mathbf{M}_n - \mathbf{b}_n}{\mathbf{a}_n} \le \mathbf{y}\right) = \lim_{n \to \infty} F^n(\mathbf{a}_n \mathbf{y} + \mathbf{b}_n) = G(\mathbf{y}), \tag{2.9}$$

where G *is the limit distribution with non-degenerate margins* G_j*,* $\mathbf{a}_n \mathbf{y} = (a_{n,1} y_1, \ldots, a_{n,d} y_d)$ *and* $(\mathbf{M}_n - \mathbf{b}_n)/\mathbf{a}_n = ((M_{n,1} - b_{n,1})/a_{n,1}, \ldots, (M_{n,d} - b_{n,d})/a_{n,d})$*. Then,* G *is called multivariate extreme value df.*

By setting $y_i = \infty, \forall i \neq j$ in (2.9), we have for $j = 1, \ldots, d$,

$$\lim_{n \to \infty} F_j^n(a_{n,j} y_j + b_{n,j}) = G_j(y_j).$$

This indicates that each margins G_j of G is a univariate extreme value df and $F_j \in \mathcal{D}(G_j)$.

Example 2.1.2. A d-dimensional random vector $\mathbf{Y}$ follows a multivariate logistic distribution with dependence parameter $\alpha \in (0, 1]$ if its joint df is

$$G(\mathbf{y}) = \exp\left\{-\left(\sum_{j=1}^{d} z_j^{1/\alpha}\right)^{\alpha}\right\}, \quad z_j \ge 0,$$

where $z_j = \{1 + \xi_j(y_i - \mu_i)/\sigma_i\}^{-1/\xi_i}$ (see Gumbel [1960]).

Example 2.1.3. Let B be the nonempty subsets of $\{1, 2, \ldots, d\}$. Let $B_1 = \{b \in B : |b| = 1\}$, where $|b|$ is the number of elements in the set b, and let $B_{(j)} = \{b \in B : j \in b\}$. The

d-dimensional multivariate asymmetric logistic df (see Tawn [1990]) is given by

$$G(y) = \exp\left\{ -\sum_{b \in B} \left[\sum_{j \in b} (\theta_{j,b} z_j)^{1/\alpha_b} \right]^{\alpha_b} \right\},$$

where $z_j = \{1 + \xi_j(y_i - \mu_i)/\sigma_i\}^{1/\xi_i}$, the dependence parameters $\alpha_b \in (0,1]$ for all $b \in B \setminus B_1$, and the asymmetry parameters $\theta_{j,b} \in [0,1]$ for all $b \in B$ and $j \in b$. The constraints $\sum_{j \in B(j)} \theta_{j,b} = 1$ ensure that the marginal distributions are generalized extreme value. The model contains $2^d - d - 1$ dependence parameters and $d(2^{d-1} - 1)$ asymmetry parameters.

2.2 Tail Dependence Function

In addition to characterizing the marginal distributions of multivariate extremes, another important aspect of studying multivariate extremes is their dependence structure. There exists a great variety of equivalent descriptions of extreme value dependence structures (for more details, refer to Beirlant et al. [2006]), and the tail dependence (TD) function is one of the popular ways. As the CoVaR estimation problem involves bivariate random vectors, we restrict our attention to the TD function in the bivariate setting.

2.2.1 Definition

Consider a random vector (X, Y) with df F and margins F_1, F_2. Let $\mathbf{U} = (U_1, U_2) = (F_1(X), F_2(Y))$. Obviously, the random vector $\mathbf{U}$ has standard uniform margins.

Definition 6. *The df F is said to have the upper TD function $R(x, y)$ if for all $x, y > 0$, the following limit exists:*

$$\begin{aligned} &\lim_{u \to 0} \frac{\mathrm{P}\{U_1 \geq 1 - ux, U_2 \geq 1 - uy\}}{u} \\ =&\lim_{u \to 0} \frac{\mathrm{P}\{1 - F_1(X) \leq ux, 1 - F_2(Y) \leq uy\}}{u} \\ =&R(x, y). \end{aligned} \tag{2.10}$$

In addition to the upper TD function we use in this study, another popular TD function is the stable TD function introduced by Huang [1992].

Definition 7. *The df F is said to have a stable TD function $l(x;y)$ if for all $x;y > 0$, the following limits exists:*

$$\begin{aligned} &\lim_{u \to 0} \frac{\mathrm{P}\{U_1 \geq 1 - ux \text{ or } U_2 \geq 1 - uy\}}{u} \\ = &\lim_{u \to 0} \frac{\mathrm{P}\{1 - F_1(X) \leq ux \text{ or } 1 - F_2(Y) \leq uy\}}{u} \\ = &l(x;y). \end{aligned} \tag{2.11}$$

It is easy to find the stable TD function for EVDs. Suppose the random vector $(X;Y)$ has a joint df G. Then the stable TD function can be expressed as:

$$l(x;y) = -\log G\left(G_1^{-1}(e^{-x}); G_2^{-1}(e^{-y})\right); \tag{2.12}$$

where G_1 and G_2 are the margins of G. Therefore, we can obtain the upper TD function for the bivariate extreme value distribution function with formula:

$$R(x;y) = x + y - l(x;y). \tag{2.13}$$

The upper TD function is differentiable almost surely and homogeneous of degree one[3] (see e.g., Nikoloulopoulos et al. [2009]). With Euler's theorem[4] (Wilson [1912]) on homogeneous functions, the upper TD functions can be written as:

$$R(x;y) = x\frac{\partial R(x;y)}{\partial x} + y\frac{\partial R(x;y)}{\partial y};$$

where partial derivatives $\partial R/\partial x$, $\partial R/\partial y$ are homogeneous of order 0 and bounded. With the sufficient condition of continuous second-order partial derivatives, the order of limits and differentiation can be exchanged. Then we have (see (2.3), Nikoloulopoulos et al. [2009])

$$\frac{\partial R(x;y)}{\partial x} = \lim_{u \to 0} \mathrm{P}\{U_2 > 1 - uy | U_1 = 1 - ux\};$$

and

$$\frac{\partial R(x;y)}{\partial y} = \lim_{u \to 0} \mathrm{P}\{U_1 > 1 - ux | U_2 = 1 - uy\}.$$

[3]If $f : V \to W$ is a function between two vector spaces on a field F, and k is an integer, then f is said to be homogeneous of degree k if $f(\alpha \mathbf{v}) = \alpha^k f(\mathbf{v})$ for all $\alpha \in F$ and $\mathbf{v} \in V$.

[4]Let $f : \mathbb{R}_+^n \to \mathbb{R}$ be continuous and also differentiable on $\mathbb{R}_+^n$. Then f is homogeneous of degree k if and only if for all $\mathbf{x} \in \mathbb{R}_+^n$, $kf(\mathbf{x}) = \sum_{i=1}^n x_i \frac{\partial f}{\partial x_i}$.

Therefore, the upper TD function can be rewritten as

$$
\begin{aligned}
R(x, y) &= x \lim_{u \to 0} \mathrm{P}\{U_2 > 1 - uy | U_1 = 1 - ux\} + y \lim_{u \to 0} \mathrm{P}\{U_1 > 1 - ux | U_2 = 1 - uy\} \\
&= x \lim_{u \to 0} \mathrm{P}\{Y > Q_1(1 - uy) | X = Q_2(1 - ux)\} \\
&\quad + y \lim_{u \to 0} \mathrm{P}\{X > Q_1(1 - ux) | Y = Q_2(1 - uy)\},
\end{aligned} \tag{2.14}
$$

where Q_1 and Q_2 are the lower quantile functions of the marginal distributions for X and Y, respectively.

2.2.2 Examples

In this section, we give examples of the TD function for five popular distributions. These five example models will be used later in the simulation and empirical studies.

Example 2.2.1. The bivariate logistic df with standard Fréchet margins is given by

$$
G(x, y; \alpha) = \exp\left\{-(x^{-1/\alpha} + y^{-1/\alpha})^{\alpha}\right\}, \tag{2.15}
$$

where $x, y > 0$ and $\alpha \in (0, 1]$ is the dependence parameter. The dependence increases as α decreases and $\alpha = 1$ stands for independence while $\alpha = 0$ represents comonotonicity (perfect dependence). The corresponding margins are $G_1(x) = G_2(x) = \exp(-1/x)$, so that $G_1^{-1}(e^{-x}) = G_2^{-1}(e^{-x}) = 1/x$. Using equation (2.12), the stable TD function is given by (Gumbel [1960])

$$
l(x, y; \alpha) = -\log G(1/x, 1/y; \alpha) = (x^{1/\alpha} + y^{1/\alpha})^{\alpha}.
$$

And then the upper TD function of the bivariate logistic distribution has the form:

$$
R(x, y; \alpha) = x + y - (x^{1/\alpha} + y^{1/\alpha})^{\alpha}. \tag{2.16}
$$

Example 2.2.2. The bivariate Hüsler-Reiss df (Hüsler and Reiss [1989]) with standard Fréchet margins is

$$
G(x, y; \lambda) = \exp\left\{-x^{-1}\Phi\left(\lambda^{-1} + \frac{\lambda}{2}\log(y/x)\right) - y^{-1}\Phi\left(\lambda^{-1} + \frac{\lambda}{2}\log(x/y)\right)\right\}, \tag{2.17}
$$

where $x, y > 0$, $\lambda > 0$ and $\Phi(\cdot)$ is the standard normal df. As λ increases, the dependence will increase, that is when $\lambda = 0$, X and Y are independent and when $\lambda \to \infty$, X and Y are perfectly dependent. Following the same procedure as above, the corresponding stable TD function is expressed as

$$l(x, y; \lambda) = x\Phi\left(\lambda^{-1} + \frac{\lambda}{2}\log(x/y)\right) + y\Phi\left(\lambda^{-1} + \frac{\lambda}{2}\log(y/x)\right),$$

and the tail dependence function is given by

$$R(x, y; \lambda) = x + y - x\Phi\left(\lambda^{-1} + \frac{\lambda}{2}\log(x/y)\right) - y\Phi\left(\lambda^{-1} + \frac{\lambda}{2}\log(y/x)\right). \tag{2.18}$$

Example 2.2.3. The bilogistic df (Smith [1990]) with standard Fréchet margins is written as

$$G(x, y; \alpha, \beta) = \exp\left(-x^{-1}q^{1-\alpha} - y^{-1}(1-q)^{1-\beta}\right), \quad x, y > 0, \tag{2.19}$$

where q is the root of the equation $(1-\alpha)x^{-1}(1-q)^{\beta} - (1-\beta)y^{-1}q^{\alpha} = 0$, and $0 < \alpha, \beta < 1$. The dependence becomes stronger as each of α, β decreases. As $\alpha = \beta = 0$, X and Y are independent and $\alpha = \beta \to 1$ represents comonotonicity. The stable TD function of the bilogistic model is

$$l(x, y; \alpha, \beta) = \int_0^1 \max\left\{(1-\alpha)t^{-\alpha}x, (1-\beta)(1-t)^{-\beta}y\right\}dt,$$

and the upper TD function is expressed as

$$R(x, y; \alpha, \beta) = x + y - \int_0^1 \max\left\{(1-\alpha)t^{-\alpha}x, (1-\beta)(1-t)^{-\beta}y\right\}dt. \tag{2.20}$$

Example 2.2.4. The bivariate asymmetric logistic distribution with standard Fréchet margins has df of the form

$$G(x, y; \theta_1, \theta_2, \alpha) = \exp\left\{-(1-\theta_1)/x - (1-\theta_2)/y - \left((\theta_1/x)^{1/\alpha} + (\theta_2/y)^{1/\alpha}\right)^{\alpha}\right\}, \tag{2.21}$$

where $x, y > 0$, α is the dependence parameter and $\theta_1, \theta_2 \in [0, 1]$ are asymmetry parameters. We should note that when the asymmetry parameters are zero, the distribution becomes the bivariate logistic distribution (see Example 2.2.1). The stable TD function of the bivariate asysmmetric logistic distribution is introduced as an extension of that of the logistic distribution (Tawn [1988]) and has the form:

$$l(x, y; \theta_1, \theta_2, \alpha) = (1-\theta_1)x + (1-\theta_2)y + \left((x\theta_1)^{1/\alpha} + (y\theta_2)^{1/\alpha}\right)^{\alpha}.$$

Therefore, we could get the upper TD function as

$$R(x;y;\ _1;\ _2;\ ) =\ _1x +\ _2y - (x\ _1)^{1=} + (y\ _2)^{1=}\ : \qquad (2.22)$$

Example 2.2.5. Suppose $W = (X;Y)$ has a bivariate t distribution with location parameter $= (0;0)^T$, scale parameter $= \begin{matrix} 1 & \\ & 1 \end{matrix}$, degree of freedom and equal margins $F_X(\cdot) = F_Y(\cdot) = F_T(\cdot;\ )$, where $F_T(\cdot;\ )$ is the df of a univariate Stundet's t random variable with degrees of freedom. Then the joint density function of W is written as

$$f_T(w;\ ;\ ) = \frac{(\ + 2)=2}{p\ 1 - \ ^2 \quad (\ =2)} \ 1 + \frac{1}{\ } w^T \quad ^1 w \quad ^{(\ +2)=2}; w \in R^2: \qquad (2.23)$$

The dependence strength increases as decrease or increases. And X and Y are independent as $\to\infty$ and $= 0$. They are perfectly dependent as $\to 0$ and $= 1$.

Proposition 2.2.1. *(Demarta and McNeil [2005]) Suppose a random vector* $(X;Y)^T$ *has the joint density function in* (2.23)*. Then its upper TD function is given by (see A.2 for proof.)*

$$\begin{aligned} R(x;y;\ ;\ ) = xF_T & \quad r \frac{\ + 1}{1 - \ ^2} \quad - (y=x)^{\ 1=} \ ;\ \ + 1 \\ + yF_T & \quad r \frac{\ + 1}{1 - \ ^2} \quad - (x=y)^{\ 1=} \ ;\ \ + 1 \ : \end{aligned} \qquad (2.24)$$

Example 2.2.6. Suppose $W = (X;Y)^T$ has a bivariate skew-t distribution with location parameter $= (0;0)^T$, shape parameter $= (\ _1;\ _2)^T$, scale parameter $= \begin{matrix} 1 & \\ & 1 \end{matrix}$, and degree of freedom, denoted as $ST_2(0;\ ;\ ;\ )$. Then its joint density is given by (Azzalini and Capitanio [2003])

$$f_{ST}(w) = 2f_T(w;\ ;\ )F_T \quad ^T w \quad r \frac{\ + 2}{\ + w^T \quad ^1 w};\ \ + 2 \ : \qquad (2.25)$$

When $_1 =\ _2 = 0$, the bivariate skew-t distribution reduces to a bivariate t distribution. The dependence strength is controlled by the parameters and , which is the same as that in the bivariate t distribution. Before we introduce the TD function of the bivariate skew-t distribution, we need to give a definition of the extended skew-t distribution, which will be used to specify the upper TD function.

Definition 8. *A random variable* $Y \in R$ *follows a univariate extended skew-t distribution with location parameter , scale parameter* $!$*, shape parameter , extended parameter and* *degree of freedom* $\nu > 0$, *denoted by* $Y \sim \mathrm{EST}_1(\xi, \omega, \alpha, \tau, \nu)$, *if its density function is*

$$f_{\mathrm{EST}}(y) = \frac{f_{\mathrm{T}}(y;\xi,\omega,\nu)}{F_{\mathrm{T}}(\tau/\sqrt{1+\alpha^2};\nu)} F_{\mathrm{T}}\left(\sqrt{\frac{\nu+1}{\nu+z^2}}(\alpha z+\tau);\ \nu+1\right),$$

where $z = (y-\xi)/\omega$. *When* $\tau = 0$, *the distribution becomes the univariate skew-t distribution.*

Proposition 2.2.2. *Suppose a random vector* $(X, Y)^T$ *has joint density function in* (2.25). *Then its upper TD function is given by (see A.3 for proof)*

$$\begin{aligned} R(x,y;\nu,\alpha_1,\alpha_2,\rho) = {} & x \cdot \bar{F}_{EST}\left(\frac{\sqrt{\nu+1}}{\sqrt{1-\rho^2}}\left((x^*/y^*)^{1/\nu}-\rho\right); 0, 1, \sqrt{1-\rho^2}\,\alpha_2, \tau_1, \nu+1\right) \\ & + y \cdot \bar{F}_{EST}\left(\frac{\sqrt{\nu+1}}{\sqrt{1-\rho^2}}\left((y^*/x^*)^{1/\nu}-\rho\right); 0, 1, \sqrt{1-\rho^2}\,\alpha_1, \tau_2, \nu+1\right), \end{aligned} \tag{2.26}$$

where $x^* = xF_T(\bar{\alpha}_2\sqrt{\nu+1};\nu)$, $y^* = xF_T(\bar{\alpha}_1\sqrt{\nu+1};\nu)$, $\bar{\alpha}_1 = \frac{\alpha_1+\rho\alpha_2}{\sqrt{1+(1-\rho^2)\alpha_2^2}}$, $\bar{\alpha}_2 = \frac{\alpha_2+\rho\alpha_1}{\sqrt{1+(1-\rho^2)\alpha_1^2}}$, $\tau_1 = \sqrt{\nu+1}(\alpha_1+\rho\alpha_2)$, $\tau_2 = \sqrt{\nu+1}(\alpha_2+\rho\alpha_1)$ *and* $\bar{F}_{\mathrm{EST}}$ *is the survival function of the extended skew-t distribution in Definition 8.*

2.2.3 Estimation of the tail dependence function

There exists a number of methods that can be used to estimate the TD function, however, due to data sparsity, efficiency can be gained by assuming a parametric model for the TD function. Coles and Tawn [1991] and Joe et al. [1992] apply maximum likelihood estimation to estimate, while Ledford and Tawn [1996] and Smith [1994] use a censored likelihood approach to implement the estimation. Einmahl et al. [2008] point out that these likelihood-based estimation methods require the smoothness (or even existence) of the partial derivatives of the TD function. Therefore, they propose an estimator based on the method-of-moments for dimension two, which requires smaller set of conditions. In the simulation studies and subsequent data analysis, we adopt the method-of-moments (M-estimator) proposed in Einmahl et al. [2008] and later extended in Einmahl et al. [2012]. This extended M-estimator can be used in arbitrary dimension d and its consistency and asymptotic normality hold under weak conditions.

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be independent random vectors in $\mathbb{R}^2$ with a common continuous df F and margins F_1 and F_2. Let R_i^X and R_i^Y denote the rank of X_i among $X_1, \ldots, X_n$ and the rank of Y_i among $Y_1, \ldots, Y_n$, respectively, where $i \in \{1, \ldots, n\}$. Then for $1 \leq m \leq n$, a nonparametric estimator for the bivariate upper TD function R is defined as:

$$\hat{R}_n(x, y) := \frac{1}{m} \sum_{i=1}^{n} 1\left\{R_i^X \geq n + \frac{1}{2} - mx, R_i^Y \geq n + \frac{1}{2} - my\right\}; \tag{2.27}$$

where $m = m_n \in \{1, \ldots, n\}$ is an intermediate sequence.

We suppose that the function R belongs to some parametric family $\{R(\cdot, \cdot; \theta) : \theta \in \Theta\}$, where $\Theta \subset \mathbb{R}^p (p \geq 1)$ is the parameter space. Let $g = (g_1, \ldots, g_p)^T : [0, 1]^2 \to \mathbb{R}^p$ be a vector of integrable functions. Define function $\varphi : \Theta \to \mathbb{R}^p$ as:

$$\varphi(\theta) := \int\int_{[0,1]^2} g(x, y) R(x, y; \theta) dxdy; \tag{2.28}$$

where φ is a homeomorphism[5] between Θ and its image. For example, in the logistic and Hüsler-Reiss distribution, $\varphi(\theta)$ is a 1-1 mapping of the dependence parameter. In the asymmetric distribution, one component of $\varphi(\theta)$ is the mapping of the dependence parameter and the other two components are mappings of the asymmetry parameters.

Let θ_0 denote the true value of parameter θ. The M-estimator $\hat{\theta}_n$ of θ_0 is defined as a minimizer of the function

$$S_{m,n}(\theta) = \left\| \varphi(\theta) - \int\int_{[0,1]^2} g(x, y) \hat{R}_n(x, y) dxdy \right\|^2; \tag{2.29}$$

where $||\cdot||$ is the Euclidean norm. However, we should note that the choice of function φ is not unique, which means that the M-estimation is not unique. How to choose a proper φ will be discussed in Section 3.2.1.

Theorem 2.2.3. (Existence, uniqueness and consistency of $\hat{\theta}_n$) *Define $\hat{\Theta}_n$ as the set of minimizers of $S_{m,n}$ in* (2.29). *Let $g : [0, 1]^2 \to R^p$ be integrable.*

(i) If φ is a homeomorphism from $\Theta \to \varphi(\Theta)$ and if there exists $\epsilon_0 > 0$ such that the set $\{\theta \in \Theta : ||\theta - \theta_0|| \leq \epsilon_0\}$ is closed, then for every ϵ such that $\epsilon_0 > \epsilon > 0$, as $n \to \infty$,

$$P\left(\hat{\Theta}_n \neq \emptyset \text{ and } \hat{\Theta}_n \subset \{\theta \in \Theta : ||\theta - \theta_0|| \leq \epsilon\}\right) \to 1.$$

[5] A function $f : X \to Y$ between two topological space (X, τ_X) and (Y, τ_Y) is homeomorphism if: (1) f is bijection (one-to-one and onto); (2) f is continuous; (3) the inverse function f^{-1} is continuous (f is open mapping).

(ii) *If in addition to the assumptions of (i), θ_0 is in the interior of the parameter space, φ is twice continuously differentiable and its derivative matrix $D\varphi(\theta_0)$ is of full rank, then, with probability tending to one, $S_{m,n}$ in equation* (2.29) *has a unique minimizer $\hat{\theta}_n$. Hence,*

$$\hat{\theta}_n \xrightarrow{P} \theta_0 \qquad \text{as } n \to \infty.$$

Denote W as a mean-zero Wiener process on $[0,\infty]^2 \setminus \{(\infty,\infty)\}$ with covariance function

$$E\left[W(x_1,y_1)W(x_2,y_2)\right] = R(x_1 \wedge x_2, y_1 \wedge y_2)$$

and for $x, y \in [0,\infty)$, let $W_1(x) := W(x,\infty)$, $W_2(x) := W(\infty, y)$. Further, for $(x,y) \in [0,\infty)^2$, let $R_1(x,y)$ and $R_2(x,y)$ be the right-hand partial derivatives of R at the point (x,y) with respect to the first and second coordinate, respectively. Write

$$B(x,y) = W(x,y) - R_1(x,y)W_1(x) - R_2(x,y)W_2(y),$$

$$\tilde{B} = \int\int_{[0,1]^2} g(x,y)B(x,y)dxdy.$$

Theorem 2.2.4. (Asymptotic normality of $\hat{\theta}_n$) *In addition to the assumptions of Theorem 2.2.3(ii), if the following two conditions also hold:*

(i) $t^{-1}\mathrm{P}\{1 - F_X(X) \le ux, 1 - F_Y(Y) \le uy\} - R(x,y) = O(t^{\alpha})$, *uniformly on the set* $\{(x,y) : x + y = 1, x \ge 0, y \ge 0\}$ *as* $u \to 0$, *for some* $\alpha > 0$;

(ii) $m = m_n \to \infty$ *and* $m = o(n^{2\alpha/(1+2\alpha)})$ *as* $n \to \infty$,

then as $n \to \infty$,

$$\sqrt{m}(\hat{\theta}_n - \theta_0) \xrightarrow{d} D\varphi(\theta_0)^{-1}\tilde{B}. \tag{2.30}$$

Theorems 2.2.3 and 2.2.4 establish the existence and asymptotic normality of M-estimator. For proofs of these two theorems, please refer to Einmahl et al. [2012].

Chapter 3

Methodology

In this chapter, we focus on the so-called conditional Value-at-Risk or CoVaR as a way to capture systemic risk contributions, and explore how the proposed EVT-based semi-parametric methodology can be utilized in the stationary setting to produce estimates of CoVaR. We also further examine the performance with several simulation studies.

3.1 CoVaR Estimation in a Stationary Setting

In this study, we adopt a modified definition of CoVaR proposed by Girardi and Ergün [2013], where the financial distress is specified by a loss in excess of the VaR, rather than being at the VaR level. Mainik and Schaanning [2014] show that, under this definition, CoVaR is dependence consistent in the sense that an increase in the strength of dependence does lead to an increase in systemic risk as measured by CoVaR.

Let X and and Y denote, respectively, losses for a financial institution and a system proxy. The VaR at a confidence level $p \in (0; 1)$ for X is defined as the $1-p$-quantile of the underlying distribution (assuming the quantile is single-valued):

$$\mathrm{P}\{X \geq \mathrm{VaR}_X(1-p)\} = p;$$

and $\mathrm{CoVaR}_{Y|X}(1-p)$ is defined by the $(1-p)$-quantile of the conditional distribution:

$$\mathrm{P}\{Y \geq \mathrm{CoVaR}_{Y|X}(1-p) | X \geq \mathrm{VaR}_X(1-p)\} = p. \tag{3.1}$$

The idea of our methodology is to link the definition of CoVaR to the tail dependence function. As CoVaR is actually a high-quantile of the conditional distribution of Y, it seems sensible to make the following assumption: there exists a constant   such that

$$\mathrm{CoVaR}_{Y|X}(1-p) = \mathrm{VaR}_Y(1-p\ ). \tag{3.2}$$

Under this assumption, we can re-write equation (3.1) as

$$\frac{\mathrm{P}\{Y > \mathrm{VaR}_Y(1-p\ ), X > \mathrm{VaR}_X(1-p)\}}{p} = p. \tag{3.3}$$

Suppose the two-dimensional random vector (X, Y) has joint distribution function F and continuous margins F_1 and F_2. Assume that

(i) $F \in \mathcal{D}_2(G)$, where G is a bivariate extreme value distribution, and

(ii) $F_2 \in \mathcal{D}(\ _{1/}\ )$ for some $\ > 0$.

From (ii), we have (see Chapter 6 in de Haan and Ferreira [2006])

$$\lim_{u \to 0} \frac{\mathrm{P}\{F_1(X) > 1 - ux, F_2(Y) > 1 - uy\}}{u} = R(x, y), \quad x, y > 0, \tag{3.4}$$

where R is known as the upper TD function discussed in Section 2.2.

Combining equation (3.3) and (3.4), we have for p close to zero:

$$\begin{aligned} &\frac{\mathrm{P}\{Y > \mathrm{VaR}_Y(1-p\ ), X > \mathrm{VaR}_X(1-p)\}}{p} \\ =&\frac{\mathrm{P}\{F_2(Y) > 1-p\ , F_1(X) > 1-p\}}{p} \approx R(1,\ ). \end{aligned} \tag{3.5}$$

Due to the fact that the distribution function of Y is in the domain of attraction with positive index $1/$, we have $1 - F_2 \in RV_{\ 1/}$ (see Theorem 2.1.2), which means $\mathrm{VaR}_Y(1-p\ ) \approx$ $\ \mathrm{VaR}_Y(1-p)$ (see Proposition 0.8 (v) in Resnick [1987]). Hence, if we can find an   such that $R(1; \ ) = p$, we obtain the following approximation for CoVaR valid for p close to zero:

$$\text{CoVaR}_{Y|X}(1-p) = \text{VaR}_Y(1-p \ \ ) \approx \text{VaR}_Y(1-p) \quad :$$

In order to find , it is necessary for us to estimate the TD function first. There are various ways to estimate the upper TD function. However, due to data sparsity, efficiency can be gained by assuming a flexible parametric model for the TD function $R(x; y) = R(x; y; \ )$, where denotes the parameter vector. Note that the function R is the df of a measure (see Chapter 6.1.5 in de Haan and Ferreira [2006]), so $R(x; y; \ )$ is monotone at x and y. Moreover, we can see that $0 \leq R(x; y) \leq x \wedge y$, so $0 \leq R(1; \ ; \ ) \leq$ when ≤ 1. In this sense, there always exists an $> p$ such that $R(1; \ ; \ ) = p$. If we can find such , then we should have $= g(\ ; p)$, that is, is a function of CoVaR level p and model parameter . We next illustrate plots of $R(1; \ ; \ )$ as a function of for some parametric models and at various parameter values to show how is influenced by the parameter .

Example 3.1.1 (Bivariate logistic distribution)**.** The function $R(1; \ )$ of the bivariate logistic distribution is given by

$$R(1; \ ; \ ) = 1 + \quad - (1 + \quad ^{1=} \ ) \ ; \quad 0 < \quad \leq 1:$$

From Figure 3.1, we can see that there is a negative monotone relationship between $R(1; \ )$ and , which indicates that for a fixed level p, as the strength of dependence increases (i.e., is smaller), becomes samller. Moreover, $R(1; \ ; 1) = 0$ for all . As $\to 0$, $R(1; \ ; \ ) \to$ for < 1 and $R(1; \ ; \ ) \to 1$ for ≥ 1.

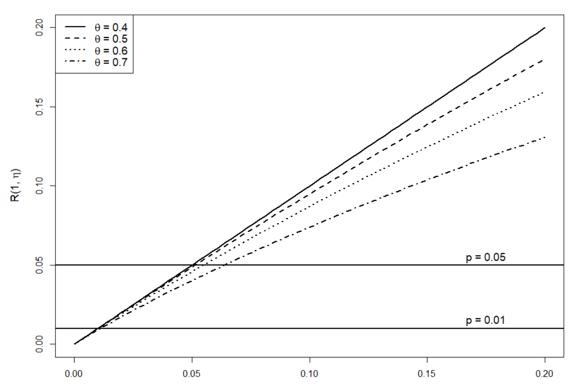


Figure 3.1: Plot of $R(1; \ )$ as a function of for the bivariate logistic distribution for different values of .

Example 3.1.2 (Bivariate Hüsler-Reiss distribution)**.** The function $R(1; \;)$ of the bivariate Hüsler-Reiss distribution is expressed as

$$R(1; \; ; \;) = 1 + \quad - \qquad {}^{1} - \frac{}{2}\log \quad - \qquad {}^{1} + \frac{}{2}\log \quad ; \qquad > 0.$$

From Figure 3.2, we can see that at a given level p, as increases, that is the strength of tail dependence increases, will becomes smaller. Moreover, as $\to 0$, $R(1; \; ; \;) \to 0$ for all . As $\to \infty$, $R(1; \; ; \;) \to$ for < 1 and $R(1; \; ; \;) \to 1$ for > 1.

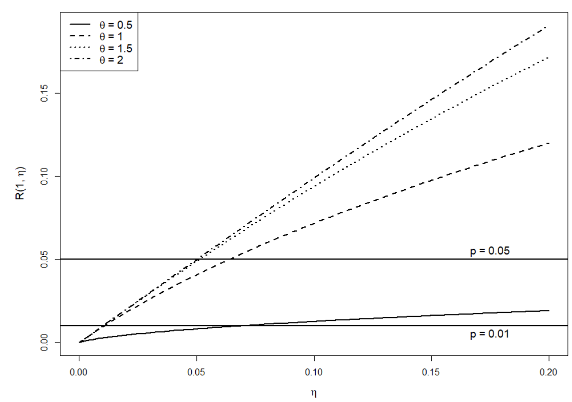


Figure 3.2: Plot of $R(1; \;)$ as a function of for the bivariate Hüsler-Reiss distribution for different values of .

Example 3.1.3 (Bivariate bilogistic distribution)**.** The function $R(1; \;)$ of the bivariate bilogistic distribution is written as

$$R(1; \; ; \; ; \;) = 1 + \quad - \int_0^1 \max \; (1 - \;)t \quad ; (1 - \;)(1 - t) \qquad dt; \quad 0 < \; ; \; < 1.$$

From Figure 3.3, there is a positive monotone relationship between and . That is to say, for a given level p, decreases as the dependence becomes stronger (i.e., each of , decreases).

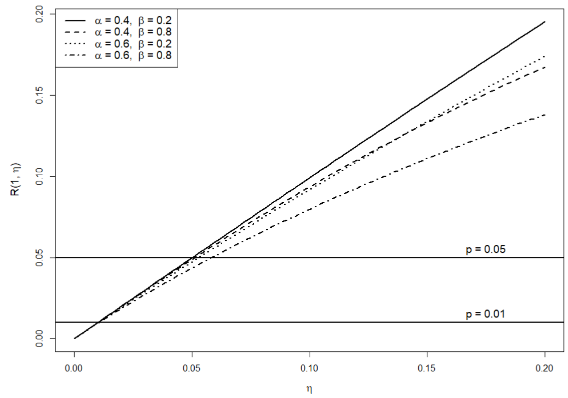


Figure 3.3: Plot of $R(1;\ )$ as a function of for the bivariate bilogistic distribution for different values of and .

Example 3.1.4 (Bivariate asymmetric logistic distribution)**.** The function $R(1;\ )$ of the bivariate asymmetric logistic distribution is given by

$$R(1;\ ;\ _1;\ _2;\ ) = \ _1 + \ _2 - \ _1^{1=} + (\ _2)^{1=} \ ; \quad 0 < \ \leq 1 \text{ and } 0 \leq \ _1;\ _2 \leq 1:$$

Obviously, the effect of dependence strength (the value of) on the value of will be the same as for the bivariate logistic distribution. Furthermore, in Figure 3.4, we can see that will increases as $_1$ or $_2$ decreases. However, when we increase three parameters simultaneously, the change in the value of for a given value of p may increase or decrease.

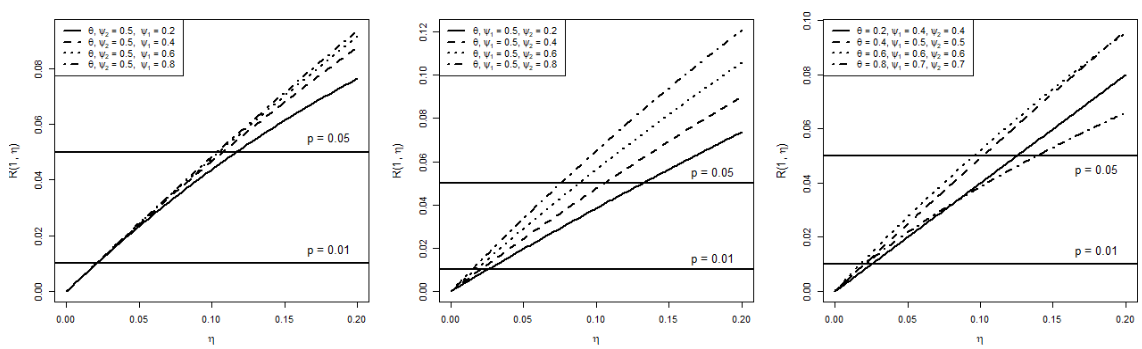


Figure 3.4: Plots of $R(1;\ )$ as a function of for the bivariate asymmetric logistic distribution for different values of , $_1$, $_2$.

Example 3.1.5 (Bivariate t distribution)**.** The function $R(1;\ )$ of the bivariate t distribution with joint density function (2.23) is given by

$$R(1;\ ;\ ;\ ) = F_T\left(\sqrt{\frac{\ +1}{1-\ ^2}}(\ -\ ^{1=});0;1;\ +1\right) + F_T\left(\sqrt{\frac{\ +1}{1-\ ^2}}(\ -\ ^{1=});0;1;\ +1\right);\quad 0\le\ \le 1 \text{ and }\ >0.$$

In Figure 3.5, it is obvious that, increases with stronger dependence (i.e., decreases or increases). Moreover, for every , $R(1;\ ;1;\ )=\ $. As $\to\infty$, the bivariate t distribution becomes the bivariate Gaussian distribution and then $R(1;\ ;0;\ )\to 0$.

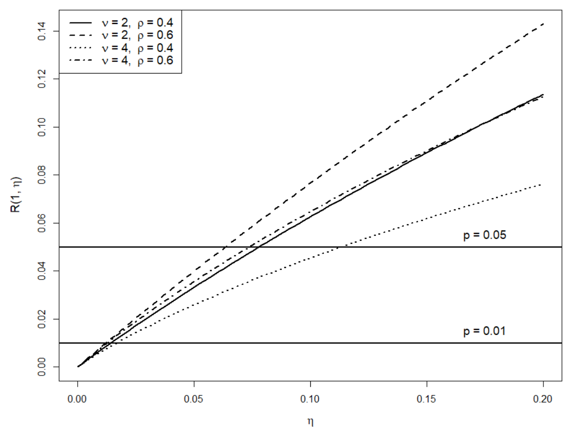


Figure 3.5: Plot of $R(1;\ )$ as a function of for the bivariate t distribution for different values of and .

Let $\{(X_1;Y_1);\dots;(X_n;Y_n)\}$ be a random sample. In an extreme value setting, suppose $p = p_n$ is small relative to sample size n. We estimate the parameter vector in $R(1;\ ;\ )$ with the M-estimator in Section 2.2.3. And then, the estimator of is expressed as

$$\hat{\ } = g(\hat{\ };p_n); \tag{3.6}$$

where $\hat{\ }$ is the M-estimator. Combining the Hill estimator for the tail index in Section 2.1.2, a nonparametric estimator of a high-quantile for $\text{VaR}_Y(1-p)$ in Section 2.1.3 and the estimator for parameter in (3.6), we obtain the following estimator for CoVaR at level p_n:

$$\widehat{\text{CoVaR}}_{Y|X}(1-p_n) = Y_{n;n\ k}\left(\frac{k}{np_n}\right)^{\hat{\ }}\hat{\ }^{\hat{\ }}; \tag{3.7}$$

where $Y_{n;1}\le\dots\le Y_{n;n}$ are the order statistics. Algorithm 2 summarizes the procedure of

CoVaR estimation for the random sample $\{(X_1; Y_1); :::; (X_n; Y_n)\}$.

Algorithm 2 CoVaR Estimation in Stationary Setting

1: Select a parametric model for the tail dependence function, and estimate parameter using M-estimator;
2: Obtain ^ by solving equation $R(1; \; ; \hat{}) = p_n$ for a given p_n;
3: Obtain Hill estimator of with equation (2.4), where k is choosing with the two-step subsample bootstrap method described in Algorithm 1;
4: Estimate $\mathrm{VaR}_Y(1 - p_n)$ by utilizing ^ with equation (2.8);
5: Input all the estimators into equation (3.7) to get the estimator of $\mathrm{CoVaR}_{Y|X}(1 - p_n)$.

3.2 Simulation Studies

In this section, we report results of five simulation studies to show the performance of the proposed method. These studies are based on the following distributions: bivariate logistic distribution, bivariate asymmetric logistic distribution, bivariate Hüsler-Reiss distribution, bivariate bilogistic distribution and bivariate t distribution. The first four models all in the class of bivariate EVDs. Due to the max-stability property of EVDs, they are in the domain of attraction. Moreover, the bivariate t distribution has regularly varying tail, hence is also in the domain of attraction (see e.g., Example 5.21 in Resnick [1987]).

3.2.1 Performance of the M-estimator of the TD Function

The M-estimator discussed in Section 2.2.3 depends on the choices of m in equation (2.27) and the function vector g in equation (2.28). However, what is a proper choice of m and g? To answer this question, we simulated samples with different sizes from several bivariate distributions to explore how the values of m and g influence the M-estimator of model parameters. To illustrate the performance of estimators, bias and root mean squared error (RMSE) are plotted for a range of values of m. For each example, we look at 100 replications of samples and set $m \in \{80; 130; 180; 230; 280; 330\}$ for the first four examples, while for bivariate t distribution, we try $m \in \{50; 100; 150; 200; 250; 300; 350\}$. The sample sizes vary according to the dimension of parameter space .

Example 1. Bivariate logistic distribution

The bivariate logistic distribution has the upper TD function given in equation (2.16). Following the analysis in Einmahl et al. [2012], we take $(@=@\ )R(x; y;\ )$ as the optimal choice of g. Note that the optimal function g depends on the true values of model parameters and hence would not be available in practice. We also consider other choices of g, given by low order polynomials: $g_0(x; y) = 1, g_1(x; y) = x_1$. We simulate random samples of size $n = 2000$ from a bivariate logistic model with $\ = 0{:}6$. Results are shown in Figure 3.6.

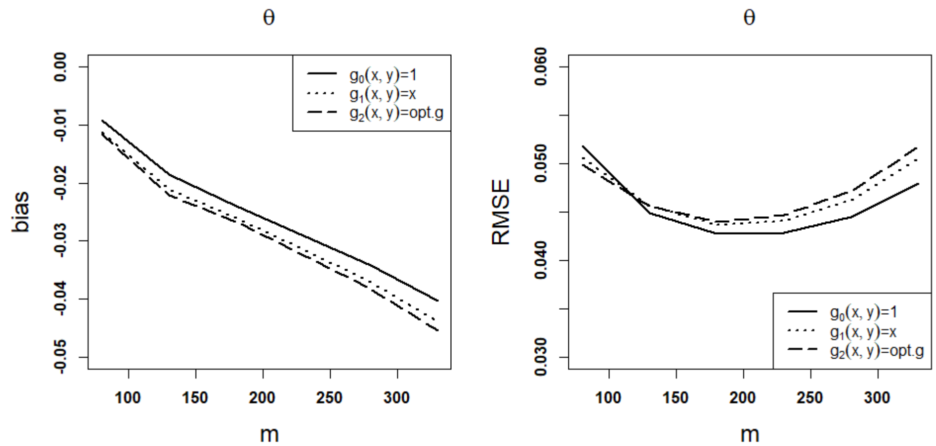


Figure 3.6: The bias and RMSE of M-estimator of based on 100 samples of size 2000 simulated from the bivariate logistic model with parameter $\ = 0{:}6$.

In Figure 3.6, it is observed that, for all choices of function g, the bias increases as m becomes bigger, while RMSE is minimized around $m = 180$. Compared with the optimal choice of g, the simpler options have smaller biases and RMSEs, and the simplest one $g_0(x; y) = 1$ shows the best performance with the smallest bias and RMSE. However, compared with m, the differences among the three choices of g are quite small, which means that in practice, the function g does not affect the estimation much and we can use the simplest form of g directly. This is consistent with findings reported in Einmahl et al. [2012]. However, the value of m appears to matter considerably here. From Figure 3.6, it seems that choosing m between 150 and 250 is reasonable when the sample size is 2000.

Example 2. Bivariate Hüsler-Reiss (HR) distribution

The bivariate HR distribution defined in equation (2.17) has the upper TD function as specified in (2.18). Similarly, we simulate random samples of size $n = 2000$ from a bivariate HR model with $\ = 2{:}5$. Based on the observation above regarding the choice of g, in this case, a simple form is assigned to function g: $g(x; y) = x$.

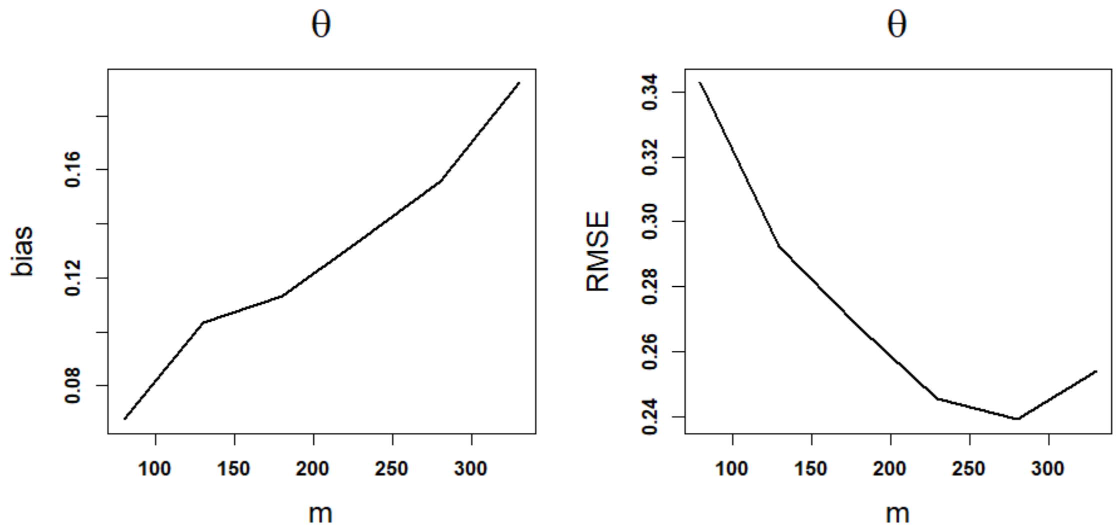


Figure 3.7: The bias and RMSE of M-estimator based on 100 samples of size 2000 simulated from the bivariate HR model with parameter $= 2:5$.

Based on Figure 3.7, the results for the HR distribution show a similar behavior as those for the logistic distribution. To be specific, the bias increases with m, while the RMSE attains a minimal point around $m = 280$. It means that for the HR distribution, when $n = 2000$, a choice of m between 250 and 300 could be considered in practice.

Example 3. Bivariate bilogistic distribution

The upper TD function of the bivariate bilogistic distribution is given in equation (2.18). Following the same steps as in the previous two examples, we simulate from the bivariate bilogistic model with $= 0:4$, $= 0:7$ and $n = 2000$. In this case, two parameters need to be estimated, so we set the function vector as $g(x; y) = (1; x)^T$. The biases and RMSEs are plotted for and separately in Figure 3.8.

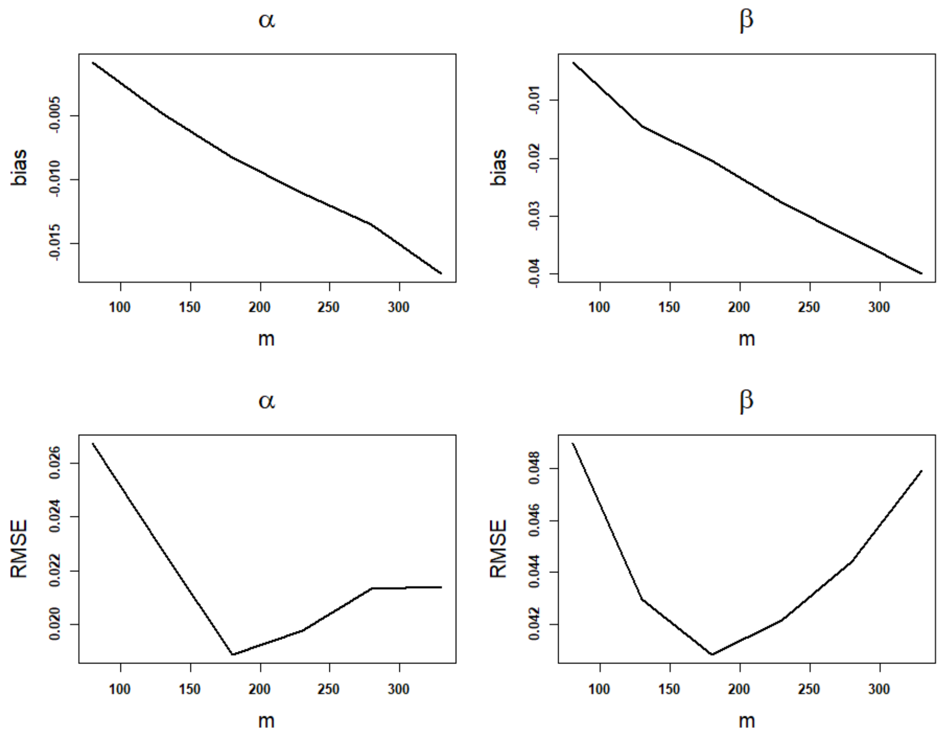


Figure 3.8: The bias and RMSE of M-estimator based on 100 samples of size 2000 simulated from the bivariate bilogistic model with parameter $= 0:4$, $= 0:7$.

In Figure 3.8, we can see that as m increases, the absolute value of bias also increases for both and . Moreover, RMSE's are both minimized at around $m = 180$, which is the same as in Example 1 for the bivariate logistic distribution. These results indicate that for the bivariate bilogistic model, when we have 2000 observations, a value of m between 150 and 200 would be a good choice.

Example 4. Bivariate asymmetric logistic distribution

In Section 2.2, equation (2.22) gives the upper TD function of the bivariate asymmetric logistic distribution, where three parameters need to be estimated. In this simulation, we also only concentrate on the choices of m, and set the function vector g as in Einmahl et al. [2012].

We simulate from a bivariate asymmetric logistical model with $= 0:6$, $_1 = 0:5$ and $_2 = 0:8$ and set $g(x; y) = \; 1; x; 2(x + y) \;^T$. As the dimension of the parameter space here is larger than in the previous examples, we increase the sample size to 2500 to keep the accuracy of estimators.

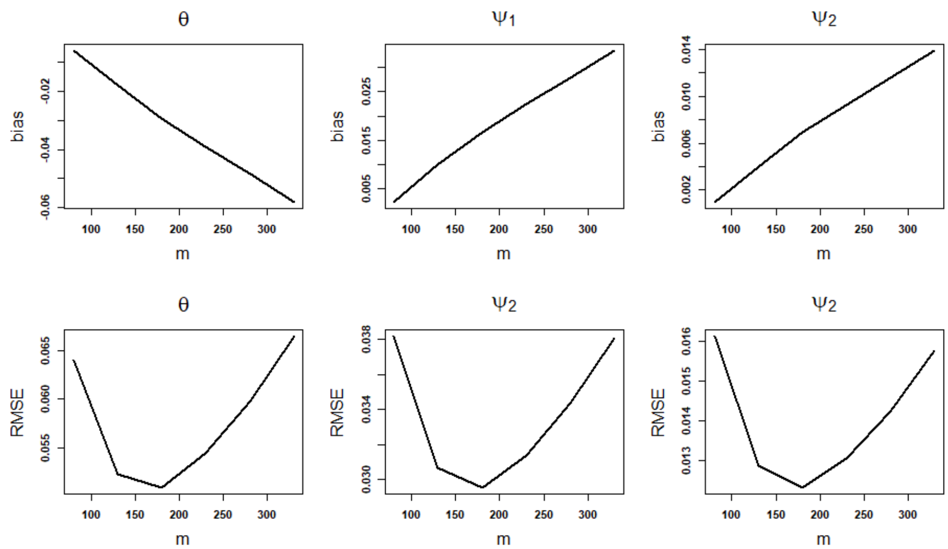


Figure 3.9: The bias and RMSE of M-estimator based on 100 replications of size 2500 simulated for the asymmetric logistic model with parameter $= 0.6$, $_1 = 0.5$, $_2 = 0.8$.

In Figure 3.9, results show an increasing trend in the absolute value of the bias over m and a minimal point of RMSE at $m = 180$ for all three parameters. These observations indicate that a reasonable choice for m is around 150 – 200 when $n = 2500$ for the bivariate asymmetric logistic model.

Example 5. Bivariate t distribution

We give the upper TD function of the standard bivariate t distribution in equation (2.24). There are two parameters and that need to be estimated. We simulate from a bivariate t distribution with $= 0.6$, $= 5$ and set $g(x; y) = (x; x + y)^T$. Due to the difficulty of estimating the tail parameter, we increase the sample size to 3000 to make the estimators more accurate.

Figure 3.10 shows the bias and RMSE of M-estimators of the two parameters. We can see that although the M-estimator of has a good performance and displays a similar behavior with respect to values of m as seen in Examples 1 – 4, most of the bias and RMSE of are bigger than 0.1, accounting for 25% of the true real value and indicating poor performance of the estimator of . Moreover, while the bias of ^ increases as m increases, the RMSE also becomes bigger, indicating the dominance of bias. However, based on the performance of ^, $m = 100$ here is a reasonable choice.

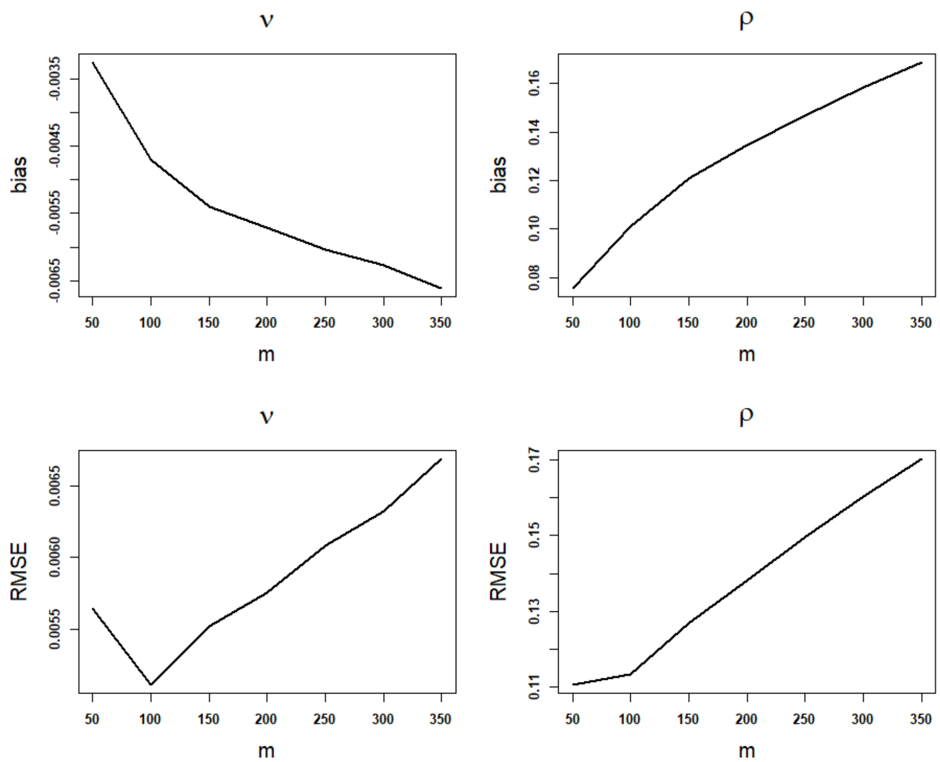
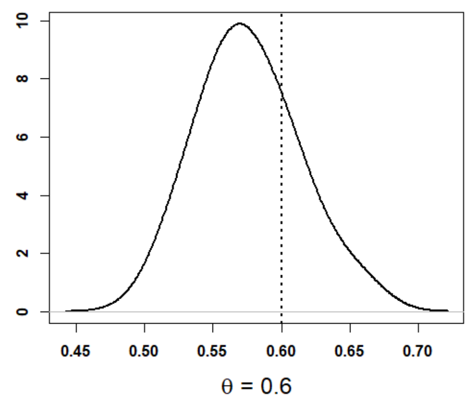


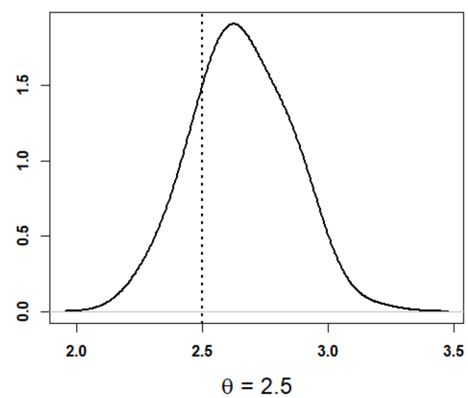
Figure 3.10: The bias and RMSE of simultaneous M-estimator based on 100 samples of size 3000 simulated from the bivariate t model with parameter $= 6$, $= 0{:}6$.

In summary, compared with choosing function g, selection of a suitable value of m is less straightforward. A good choice of m depends on the model, the dimension of the parameter space and also the sample size. It is hard to make a general recommendation. Simulation studies provide some guidance with regard to selection of m for a given model based on bias and RMSE of estimators in finite sample settings. For example, from the results for the bivariate logistic model, we find that a value of $m \in [150; 200]$ for sample size $n = 2000$ may be a good choice, that is about 8% – 10% of the sample size. In our later analysis, including simulation and empirical studies, we would let m equal to 9% of the sample size when estimating parameters of the TD function under the bivariate logistic model.

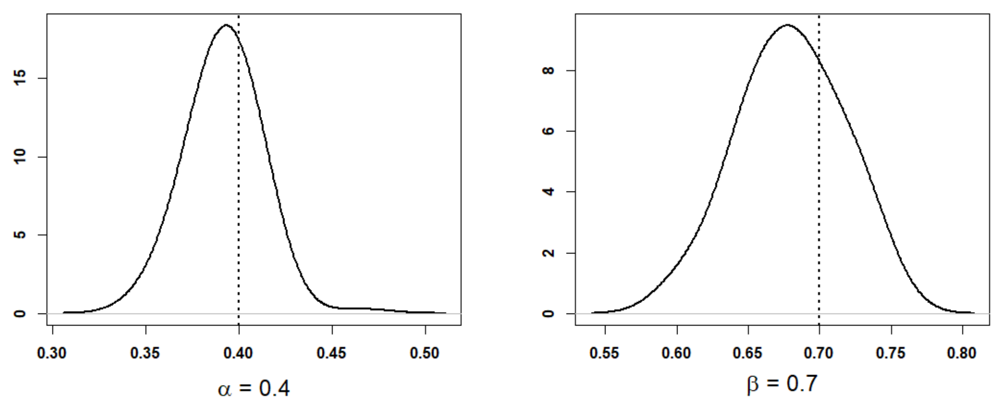
Figure 3.11 displays sampling densities for model parameters using values of m, which minimized RMSE.



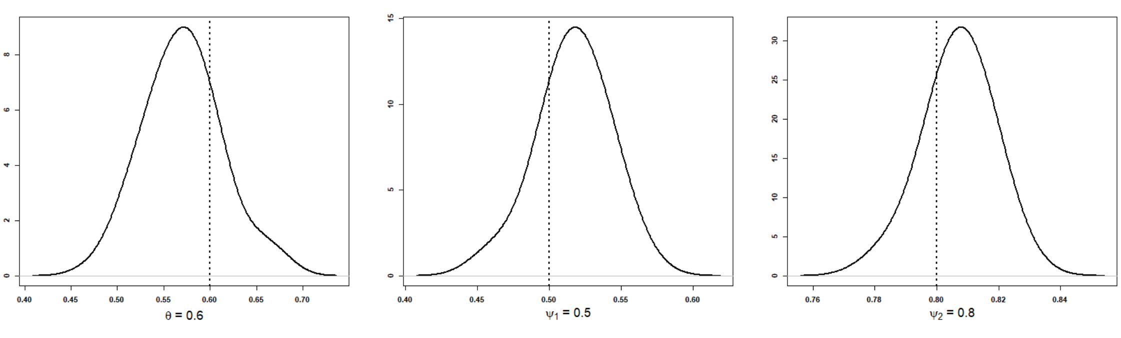
(a) logistic distribution: n = 2000, m = 180



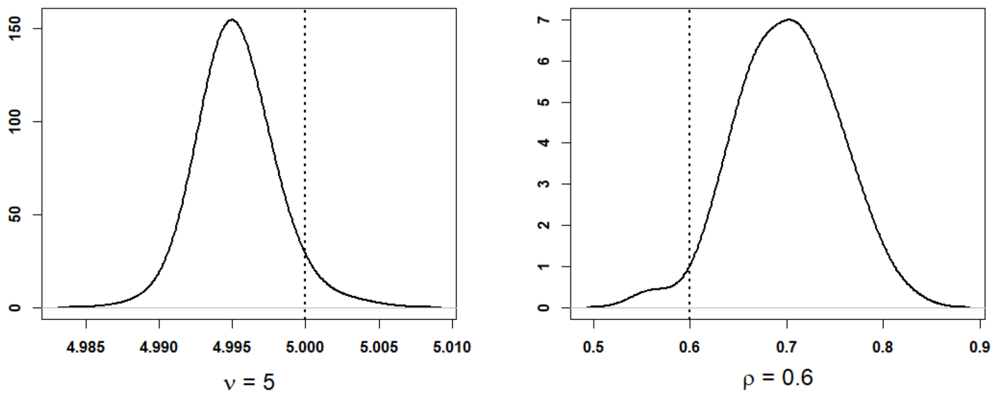
(b) HR distribution: n = 2000, m = 280



(c) Bilogistic distribution: n = 2000, m = 180



(d) Asysmmetric logisitic distribution: n = 2500, m = 180



(e) t distribution: n = 3000, m = 100

Figure 3.11: The sampling densities of estimated parameters based on 100 samples for corresponding parametric models.

From the sampling densities of the model parameter(s), it is clear that the estimated parameters in the first four examples are roughly distributed around the true value. For the bivariate t distribution, the estimated has a bigger bias compared with other parameters, but the estimator of performs well with small variability. Einmahl et al. [2008] and Einmahl et al. [2012] show that under some conditions (for details, see Theorem 2.2.4 in Section 2.2.3), the M-estimators will be asymptotically normal, which is reflected in our plots. Due to the finite sample sizes, estimators exhibit minor biases. Overall, M-estimators of parameters of the TD function perform well in the considered examples, and will hence used in subsequent studies.

3.2.2 Performance of the CoVaR Estimator

Recall equation (3.7) for estimating CoVaR in Section 3.1. For a given p_n, in order to estimate CoVaR, we need to estimate , and $\mathrm{VaR}_Y(1-p_n)$. Tail index is estimated with the Hill estimator; $\mathrm{VaR}_Y(1-p_n)$ is estimated with nonparametric extreme quantile estimator by utilizing ^; is estimated by solving equation $R(1; \ ; \hat{}) = p_n$, where $R(x; y; \ )$ is the upper TD function for an assumed parametric model and ^ is the M-estimator of . All these estimators are then used to produce an estimator of CoVaR. As accuracy of CoVaR estimation is influenced by accuracy of estimators for the three components mentioned above, our initial analysis aims to separate estimation errors attributed to each of the three components.

In this section, we explore the performance of our CoVaR estimator together with ^ , ^ and $\widehat{\mathrm{VaR}}_Y(1-p_n)$. Monte Carlo simulations are carried with the same five example distributions as in Section 3.2.1. At first, the parameters follow the same setting as in Figure 3.11 so that we can also see how the M-estimators affect our CoVaR estimation. To be specific, for logistic distribution, we put $= 0.6$, $n = 2000$ and $m = 180$. For HR distribution, we let $= 2.5$, $n = 2000$ and $m = 280$. For bilogistic distribution, we generate samples with $= 0.4$, $= 0.7$, $n = 2000$ and estimate with $m = 180$. For asymmetric logistic distribution, we have $= 0.6$, $_1 = 0.5$, $_2 = 0.8$, $N = 2500$ and $m = 180$. For t distribution, we let $= 5$, $= 0.6$, $n = 3000$ and $m = 100$. For each model, we let $p_n = 0.05$ and do 100 replications of the CoVaR estimation procedure.

When exploring the performance of $\widehat{\mathrm{CoVaR}}_{Y|X}(1-p_n)$, the true $\mathrm{CoVaR}_{Y|X}(1-p_n)$ is computed by finding the quantile of the conditional distribution, which is given as the root of

$$h(y) = \int_{\{(u,v)\in\mathbb{R}^2: u>c, v>y\}} f(u,v)\,du\,dv = p_n^2, \tag{3.8}$$

where $c = \mathrm{VaR}_X(1-p_n)$ is the true $(1-p_n)$-quantile of the marginal distribution of X, and $f(x,y)$ is the joint density function of (X,Y). Table 3.1 gives the summary statistics of proposed CoVaR estimates with ^ .

Table 3.1: Summary statistics of proposed CoVaR estimates at level $p_n = 0.05$. The margins of the first four distributions are all standard Fréchet distribution and the margins of the bivariate t distribution are all student t distribution.

	logistic	HR	bilogistic	asymmetric logistic	t
True $\mathrm{CoVaR}_{Y\|X}$	367.3064	399.4755	341.5227	281.4862	4.4215
Mean	446.3422	463.4036	460.9443	327.7476	4.5000
Median	425.4983	456.3283	434.5008	314.4263	4.4763
Standard deviation	127.9184	130.7512	149.8606	83.3997	0.5457

To further explore the performance, for each model, we also compute a $\mathrm{CoVaR}_{Y|X}(1-p_n)$ as

$$\mathrm{CoVaR}_{Y|X}(1-p_n) = \mathrm{VaR}_Y(1-p_n), \tag{3.9}$$

where $\mathrm{VaR}_Y(1-p_n)$ is the true $(1-p_n)$-quantile of the marginal distribution of Y and is the value solving equation $R(1, ,) = p_n$ with true . This is also used to examine the performance of ^ . In order to further see how the estimated would affect the estimation of CoVaR, we also make a comparison between $\widehat{\mathrm{CoVaR}}_{Y|X}(1-p_n)$ in (3.7) with estimator

$\widehat{\mathrm{CoVaR}}_{Y|X}(1-p_n) = Y_{n,n-k}(k/(np_n))^{\hat{}}$ $\hat{}$. Moreover, we also compute the real value of in (3.2) as

$$_0 = \frac{P\left(Y > \mathrm{CoVaR}_{Y|X}(1-p_n)\right)}{p_n}, \tag{3.10}$$

where $\mathrm{CoVaR}_{Y|X}(1-p_n)$ is the true value. Apart from exploring the performance of estimators, the distances between $_0$ and , $\mathrm{CoVaR}_{Y|X}$ and $\mathrm{CoVaR}_{Y|X}$ can show us how well the upper TD function approximates the conditional tail probability when p_n is relatively small.

Furthermore, for each replication, we compute the empirical estimate of the conditional exceedance probability as

$$e_n := \frac{\#\{X_i > \widehat{\mathrm{VaR}}_X(1-p_n), Y_i > \widehat{\mathrm{CoVaR}}_{Y|X}(1-p_n)\}}{\#\{X_i > \widehat{\mathrm{VaR}}_X(1-p_n)\}}, \tag{3.11}$$

where $(X_1, Y_1), \ldots, (X_n, Y_n)$ are simulated vectors, and $\widehat{\mathrm{VaR}}_X(1-p_n)$ is obtained from the method given in Section 2.1.3. When estimation of VaR and CoVaR is accurate, the ratio e_n should be close to the probability level p_n. It is analogy to the VaR backtesting (Kuester et al. [2006]) based on the exceedances of VaR, which will be discussed in Section 4.1. This statistic is useful in practice as a check on the accuracy of CoVaR estimator, provided p_n is not too extreme (small) relative to the sample size. All the results are displayed by smoothed histograms of estimators in Figures 3.12, 3.13, 3.14, 3.15 and 3.16.

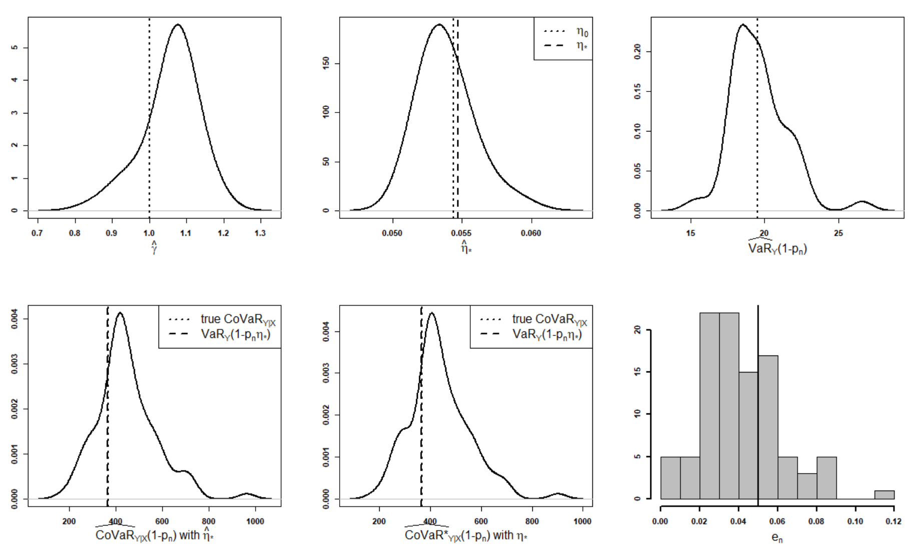


Figure 3.12: The sampling densities of estimates of , , VaR_Y, $\mathrm{CoVaR}_{Y|X}$ and e_n at level $p_n = 0.05$ based on 100 samples of size 2000 for the logistic model with parameter $= 0.6$.

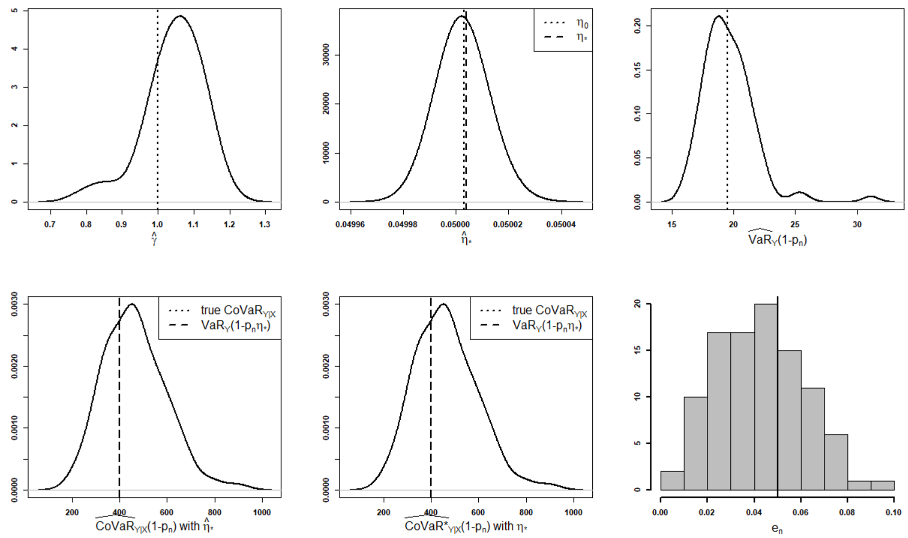


Figure 3.13: The sampling densities of estimates of , , VaR_Y, $\mathrm{CoVaR}_{Y|X}$ and e_n at level $p_n = 0.05$ based on 100 samples of size 2000 for the HR model with parameter $= 2.5$.

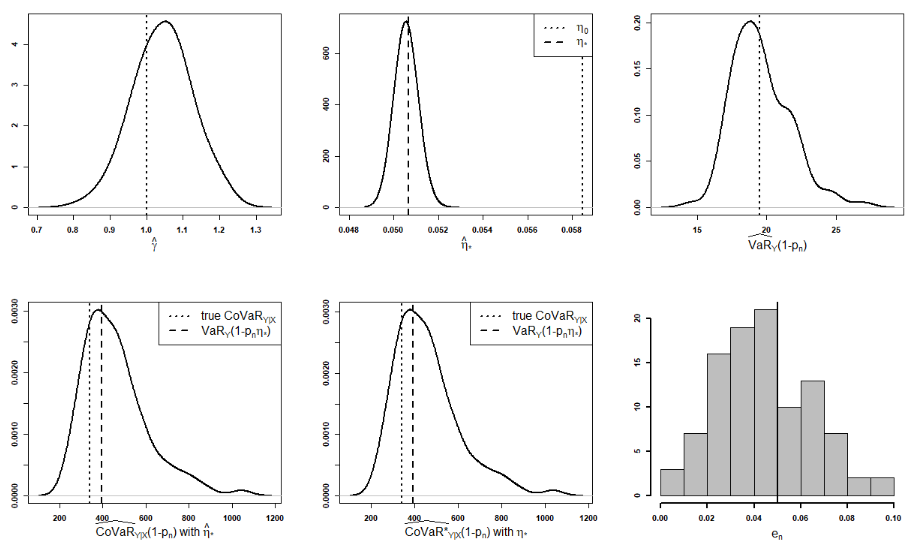


Figure 3.14: The sampling densities of estimates of , , VaR_Y, $\mathrm{CoVaR}_{Y|X}$ and e_n at level $p_n = 0.05$ based on 100 samples of size 2000 for the bilogistic model with parameters $= 0.4$, $= 0.7$.

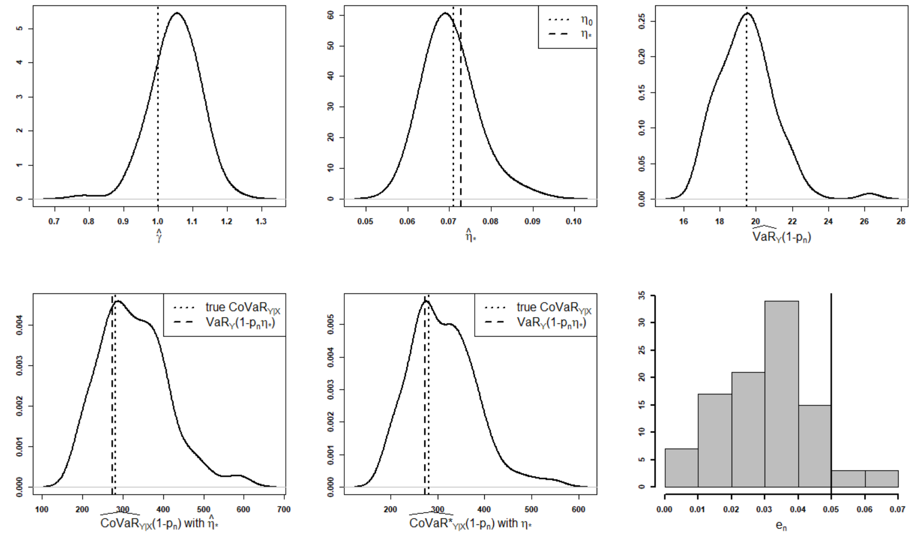


Figure 3.15: The sampling densities of estimates of , , VaR_Y, $\mathrm{CoVaR}_{Y|X}$ and e_n at level $p_n = 0.05$ based on 100 samples of size 2500 for the asymmetric logistic model with parameters $= 0.6$, $_1 = 0.5$, $_2 = 0.8$.

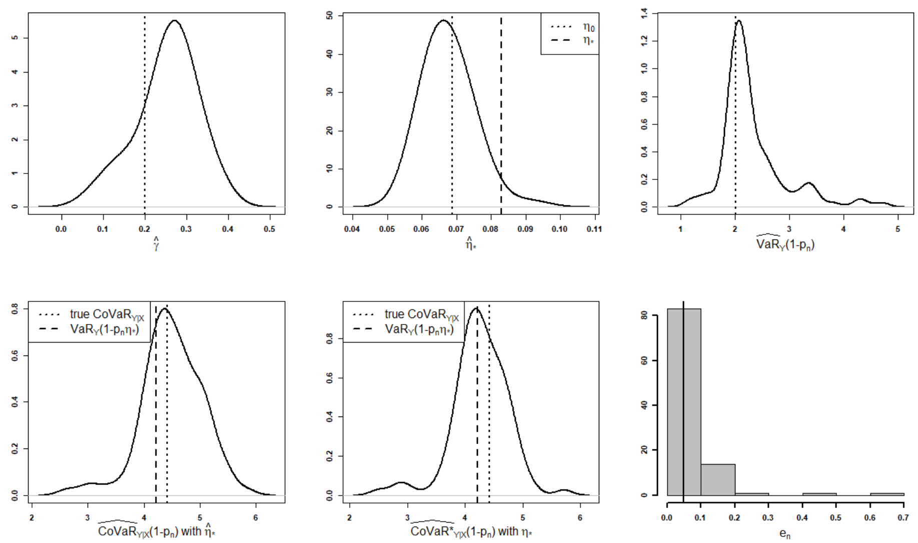


Figure 3.16: The sampling densities of estimates of , , VaR_Y, $\mathrm{CoVaR}_{Y|X}$ and e_n at level $p_n = 0.05$ based on 100 samples of size 3000 for t distribution with parameters $= 5$, $= 0.6$.

Based on the sampling densities, firstly we can observe that the Hill estimator of exhibits a small positive bias in all situations. Secondly, except the case of the bivariate t distribution, the estimators of are nearly unbiased and densities are symmetric over the true , especially in the cases of the HR distribution and the bilogistic distribution. This also indicates that the

M-estimator of   performs well and will not cause a large bias to the estimator of   for the first four examples. The value of   are close to the values of $_0$ in those four cases, while $\hat{}$ in bilogistic distribution underestimate $_0$ considerably from Figure 3.14. This may be due to the fact that the range of estimate values in the bilogistic model is much narrower than that in the logistic distribution and the asymmetric logistic distribution. In the case of the t distribution, although the difference between $_0$ and   is bigger and $\hat{}$ is more biased, an interesting phenomenon is that the density of $\hat{}$ is centered and symmetric over $_0$, which is expected to lead to more accurate CoVaR estimation.

Thirdly, the estimates of $\mathrm{VaR}_Y(1-p_n)$ are distributed around the true value in all examples. However, the standard errors of $\widehat{\mathrm{VaR}}_Y(1-p_n)$ are big, which indicates that this nonparametric estimator is not quite stable and is more likely to fail in practice. Moreover, for the estimation of CoVaR, all the estimates of $\mathrm{CoVaR}_{Y|X}(1-p_n)$ are roughly distributed around $\mathrm{CoVaR}_{Y|X}(1-p_n) = \mathrm{VaR}_Y(1-p_n\ )$, while most of them over-estimate the true value of $\mathrm{CoVaR}_{Y|X}(1-p_n)$. This is sensible, as under the key assumption in our methodology, $\mathrm{Co}\check{\mathrm{V}}\mathrm{aR}(1-p_n)$ is in fact estimating $\mathrm{VaR}_Y(1-p_n\ )$. Furthermore, there is no difference between the densities of $\mathrm{Co}\check{\mathrm{V}}\mathrm{aR}_{Y|X}(1-p_n)$ and those of $\mathrm{Co}\check{\mathrm{V}}\mathrm{aR}_{Y|X}(1-p_n)$, indicating that the estimated   does not have a big influence on the CoVaR estimation. The shapes of densities of $\mathrm{Co}\check{\mathrm{V}}\mathrm{aR}_{Y|X}(1-p_n)$ are similar to those of $\widehat{\mathrm{VaR}}_Y(1-p_n)$, showing that CoVaR estimation is dominated by the estimation of VaR_Y. In logistic model and HR model, the true $\mathrm{CoVaR}_{Y|X}$ and $\mathrm{CoVaR}_{Y|X}$ overlap with each other, which is resulted from the tiny distance between $_0$ and , implying that for these two distribution, $R(1;\ )$ approximates the conditional probability $\mathrm{P}\{Y > \mathrm{VaR}_Y(1-p\ )|X > \mathrm{VaR}_X(1-p)\}$ very well. Finally, from the histograms of the empirical estimates of the conditional exceedance probabilities, we can confirm the earlier observation that there is a tendancy for the proposed CoVaR estimator to over-estimate the true value of CoVaR.

From the results above, it seems that most of the CoVaR estimation error can be attributed to the estimation of $\mathrm{VaR}_Y(1-p_n)$. To further explore the influence of $\widehat{\mathrm{VaR}}_Y(1-p_n)$, for all distributions, we do another round of simulations by increasing the sample sizes to 5000, replication number to 200 and decrease the level p_n to 0.01. These changes allow us to produce more accurate estimates of each component. In addition to applying the nonparametric estimation of $\mathrm{VaR}_Y(1-p_n)$ introduced in Section 2.1.3, we also estimate the $\mathrm{VaR}_Y(1-p_n)$ with the Peaks-Over-Threshold (POT) method (see e.g., Embrechts et al. [1997]), where a generalized Pareto (GP) distribution is used to approximate the distribution function of exceedances above threshold u (Pickands [1975]). The details of this method are given in Appendix A.4.

Paramters of the GP distribution are fitted via maximum likelihood estimation, implemented in function "gdp.fit" of the R package "ismev".

An important aspect of the POT method is the selection of threshold u. Due to the large number of datasets, it is impossible for us to select u manually for each sample with the mean excess plot (Davison and Smith [1990]) or parameter stability plots. In our simulations, we adopt a pragmatic approach and set the threshold consistent with the value of k in the Hill estimator of : $u = Y_{n;n-k}$. For brevity of presentation, we only illustrate the sampling densities of the two estimator of $\mathrm{VaR}_Y(1-p_n)$ together with the corresponding estimators of $\mathrm{CoVaR}_{Y|X}(1-p_n)$. We also plot the histograms of e_n to further see the performance of $\mathrm{CoVaR}_{Y|X}(1-p_n)$ estimation based on the semi-parametric of the quantile. The denominator and nominator of e_n in (3.11) are also computed and plotted in Figure 3.17 – 3.21. They are the number of exceedances of VaR_X:

$$E_n := \# \{ X_i > \widehat{\mathrm{VaR}}_X(1-p_n) \} \tag{3.12}$$

and the number of joint exceedances:

$$E_n^b := \# \{ X_i > \widehat{\mathrm{VaR}}_X(1-p_n); Y_i > \widehat{\mathrm{CoVaR}}_{Y|X}(1-p_n) \} \tag{3.13}$$

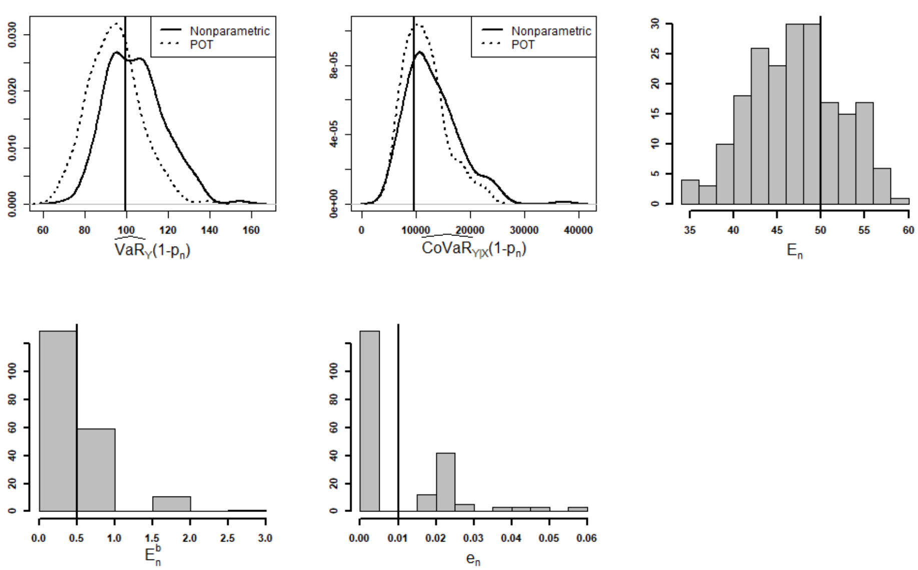


Figure 3.17: The sampling densities of estimates of VaR_Y, $\mathrm{CoVaR}_{Y|X}$, E_n, E_n^b and e_n at level $p_n = 0.01$ based on 200 samples of size 5000 for the logistic model with parameter $= 0.6$. Threshold u is chosen as: $u = Y_{n;n-k}$ with $k = 450$.

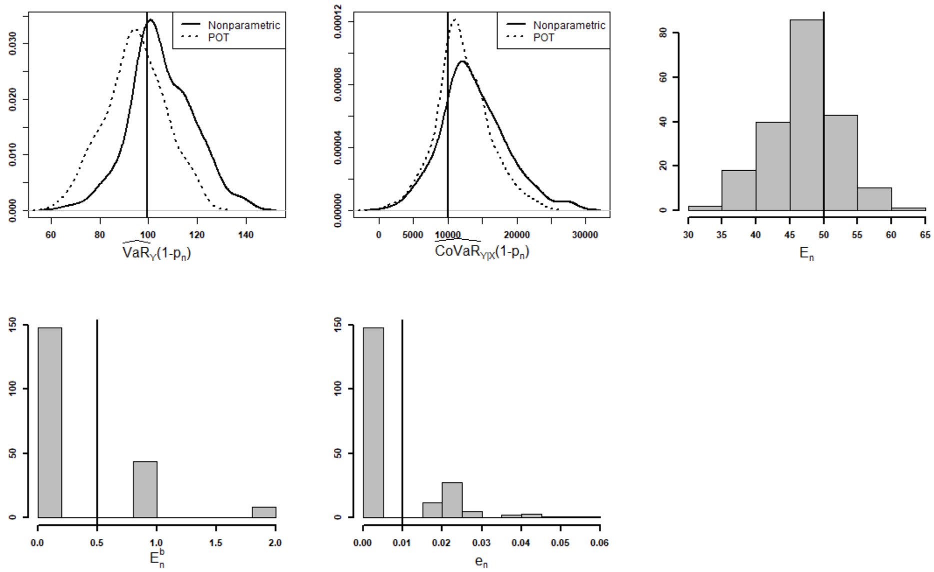


Figure 3.18: The sampling densities of estimates of VaR_Y, $\mathrm{CoVaR}_{Y|X}$, E_n, E_n^b and e_n at level $p_n = 0.01$ based on 200 samples of size 5000 for the HR model with parameter $= 2.5$. Threshold u is chosen as: $u = Y_{n,n-k}$ with $k = 700$.

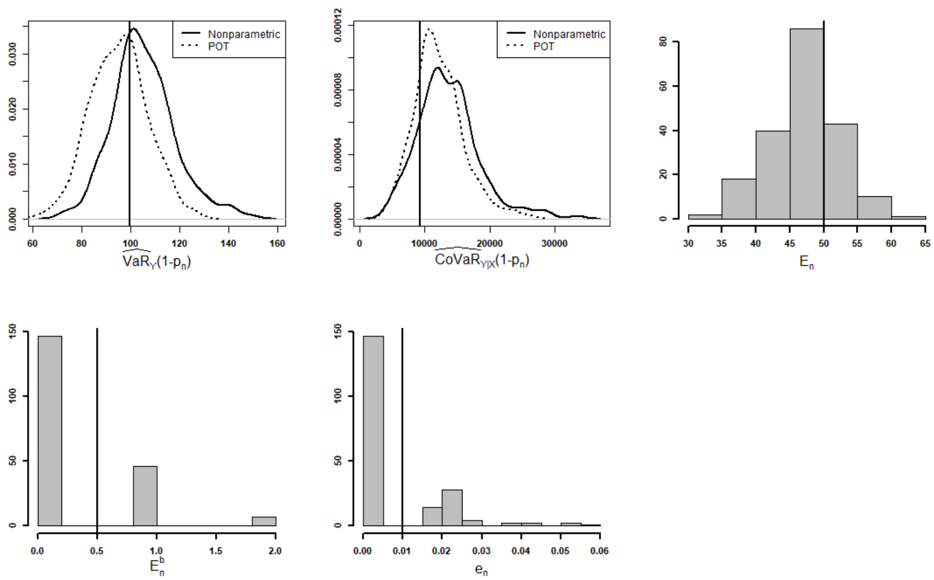


Figure 3.19: The sampling densities of estimates of VaR_Y, $\mathrm{CoVaR}_{Y|X}$, E_n, E_n^b and e_n at level $p_n = 0.01$ based on 200 samples of size 5000 for the bilogistic model with parameter $= 0.4$ and $= 0.7$. Threshold u is chosen as: $u = Y_{n,n-k}$ with $k = 450$.

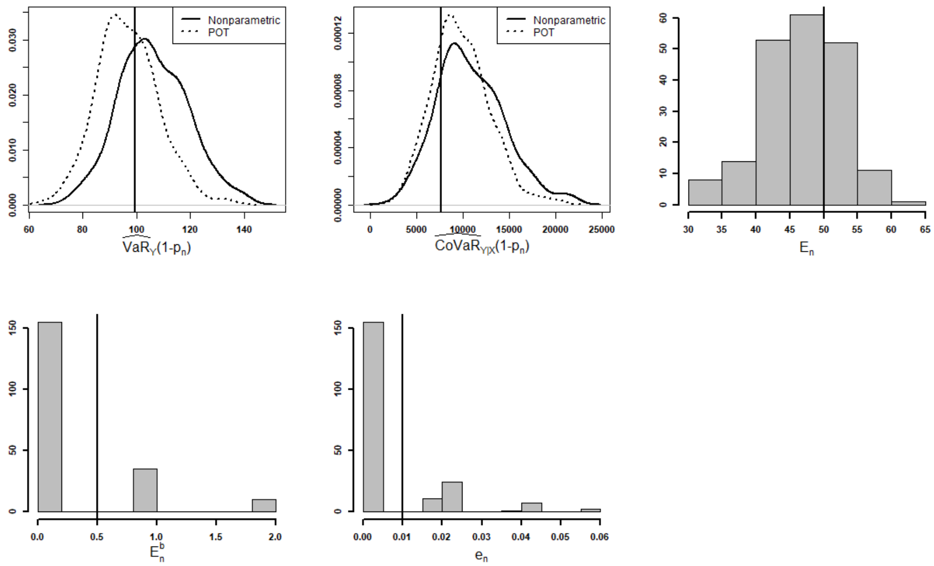


Figure 3.20: The sampling densities of estimates of VaR_Y, $\text{CoVaR}_{Y|X}$, E_n, E_n^b and e_n at level $p_n = 0.01$ based on 200 samples of size 5000 for the asymmetric logistic model with parameter $= 0.6$, $_1 = 0.5$ and $_2 = 0.8$. Threshold u is chosen as: $u = Y_{n,n-k}$ with $k = 400$.

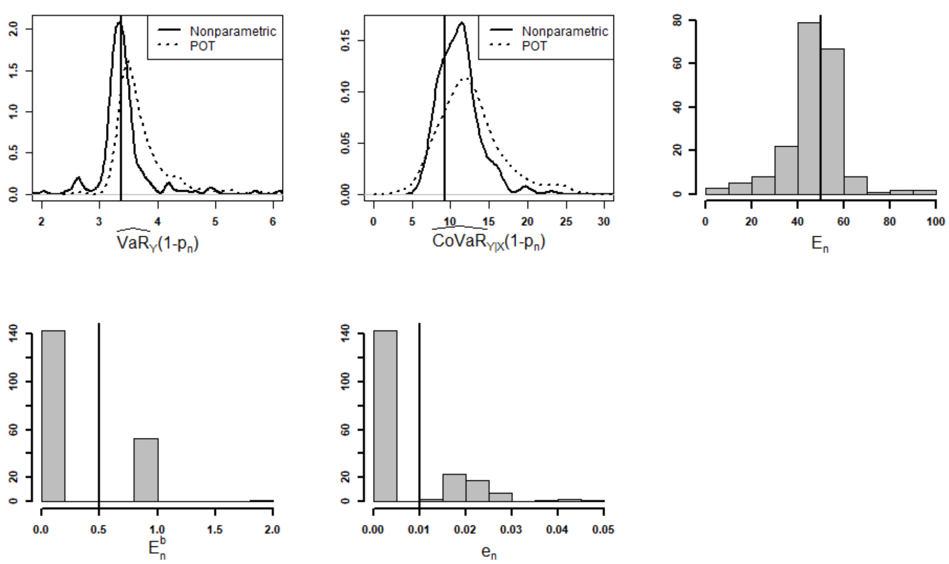


Figure 3.21: The sampling densities of estimates of VaR_Y, $\text{CoVaR}_{Y|X}$, E_n, E_n^b and e_n at level $p_n = 0.01$ based on 200 samples of size 5000 for t distribution with parameter $= 5$ and $= 0.6$. Threshold u is chosen as: $u = Y_{n,n-k}$ with $k = 150$.

According to the sampling densities of estimated $\text{VaR}_Y(1-p_n)$, it seems that the two estimators of $\widehat{\mathrm{VaR}}_Y(1-p_n)$ are quite similar to each other, especially in the cases of the HR model and bilogistic model. Both of them have large variability that may lead to prediction inaccuracies in practice. For all example distributions, the sampling densities of $\widehat{\mathrm{CoVaR}}_{Y|X}(1-p_n)$ share similar shape of those of $\widehat{\mathrm{VaR}}_Y(1-p_n)$, which is consistent with the results given in Figure 3.12 – 3.16, where simulations are generated with smaller sample sizes and at level $p_n = 0.05$. This confirms that the key to improve the performance of CoVaR estimation in our methodology is to improve the accuracy of the estimator of $\mathrm{VaR}_Y(1-p_n)$. From the histograms of E_n, most of the empirical estimators are bigger smaller than the nominal value: $5000 \times 0.01 = 50$, indicating that the estimator of $\mathrm{VaR}_X(1-p_n)$ overestimate the true value slightly. The histograms of conditional probability shows that the CoVaR estimation based on larger sample sizes and smaller p_n is better and more acceptable, because most of the probability values are between 0 and 0.02. This is close our nominal value from the histograms of E_n^b. As by letting $p_n = 0.01$ and $n = 5000$, the nominal nominal value of E_n^b is $n \times p_n^2 = 0.5$. However, the number of exceedances can only be integer, which means that E_n^b is expected to be 0 or 1. When $E_n^b = 0$, $e_n = 0$ for every E_n, and when $E_n = 1$, $e_n = 0.02$ when $E_n = 50$. The summary statistics of proposed estimates are given in Table 3.2.

Table 3.2: Summary statistics of proposed CoVaR estimates at level $p_n = 0.01$. The margins of the first four distributions are all standard Fréchet distribution and the margins of the bivariate t distribution are all student t distribution.

	logistic	HR	bilogistic	asymmetric logistic	t
True $\mathrm{CoVaR}_{Y\|X}$	9719.46	9999.50	7660.29	281.4862	9.2215
Mean	1313.98	463.4036	13772.85	10887.50	11.04
Median	12158.84	456.3283	13166.82	10500.52	10.84
Standard deviation	4986.23	4694.61	3644.54	83.3997	2.64

Chapter 4

Empirical Study

In this chapter, we use the proposed CoVaR estimation method on financial time series in a stationary setting by utilizing the five parametric models discussed in Section 3.2. Due to the volatility of financial time series, we apply the methodology to the realized innovations from a AR(1)-GARCH(1,1) filter. However, Sun and Zhou [2014] demonstrate that the realized innovations from GARCH filter for a sample with normal innovations may follow a heavy-tailed distribution, we also consider to perform the CoVaR estimation on the raw financial time series. A backtesting procedure in Girardi and Ergün [2013] is applied to assess the performance.

Remark 1. Mainik and Schaanning [2014] give a more general definition of CoVaR than presented in equation (3.1). Let $(X; Y)$ be a random vector with joint df F and continuous margins F_1 and F_2. The CoVaR at confidence level $1 - p_2$ is defined as the $(1 - p_2)$-quantile of the conditional distribution:

$$P\left(Y \geq \mathrm{CoVaR}_{Y|X}(p_1; p_2) | X \geq \mathrm{VaR}_X(1 - p_1)\right) = p_2; \tag{4.1}$$

where risk levels p_1 and p_2 can be different. Unlike the definition in Girardi and Ergün [2013], this CoVaR definition allows us to consider different levels of CoVaR when conditioning on the same distress event, and the same level of CoVaR when conditioning on a more severe distress event. It is straightforward to adapt our methodology to this more general definition of CoVaR. Under the same assumptions as stated in Section 3.1, we can re-write equation (4.1) as

$$
\begin{aligned}
p_2 &= \frac{\mathrm{P}\left(Y > \mathrm{VaR}_Y(1-p_2\ \ ), X > \mathrm{VaR}_X(1-p_1)\right)}{p_1} \\
&= \frac{\mathrm{P}\left(F_2(Y) > 1 - \frac{p_2}{p_1}\ \ \times p_1, F_1(X) > 1-p_1\right)}{p_1}.
\end{aligned}
\tag{4.2}
$$

When p_1 is close to 0, (4.2) can be approximated with $R\left(1, \frac{p_2}{p_1}\right)$, where $R(x, y)$ is the upper TD function of (X, Y). Then, if we can find such that $R\left(1, \frac{p_2}{p_1}\right) = p_2$, we can also estimate $\mathrm{CoVaR}_{Y|X}(p_1, p_2)$ with $\mathrm{VaR}_Y(1-p_2)$, where is the tail index for random variable Y.

4.1 Backtesting

In this section, we review the backtesting procedure to assess and compare different CoVaR estimates from various parametric model assumptions on TD function of R. By noting that CoVaR is actually the quantile of a conditional distribution, its evaluation can be achieved with the standard Kupiec [1995] and Christoffersen [1998] tests. The test is first designed for evaluating the accuracy of interval forecasts, and has been applied to assess the predictive performance of VaR (Kuester et al. [2006]) and CoVaR (Girardi and Ergün [2013]).

Let $\{X_t\}_{t\in\mathbb{N}}$ and $\{Y_t\}_{t\in\mathbb{N}}$ denote losses (negative log-returns) for an institution and a system proxy. The VaR at confidence level $1-p_1 \in (0, 1)$ for X_t is written as $\mathrm{VaR}_{X_t}(1-p_1)$ and the CoVaR at confidence level $1-p_2 \in (0, 1)$ for Y_t conditional on $X_t > \mathrm{VaR}_{X_t}(1-p_1)$ is written as $\mathrm{CoVaR}_{Y_t|X_t}(p_1, p_2)$.

Suppose our sample include n observations with $t = 1, \ldots, n$. For each institution, define the "hit sequence" of violation as

$$
I_{X_t} = \begin{cases} 1, & \text{if } X_t > \mathrm{VaR}_{X_t}(1-p_1) \\ 0, & \text{if } X_t \le \mathrm{VaR}_{X_t}(1-p_1) \end{cases}.
$$

For a sub-sample (assumed to have n_1 observations), when the institution is in financial distress $\{X_t > \mathrm{VaR}_{X_t}(1-p_1)\}$, construct the second "hit sequence" of violation as

$$
I_{Y_t|X_t} = \begin{cases} 1, & \text{if } Y_t > \mathrm{CoVaR}_{Y_t|X_t}(p_1, p_2) \\ 0, & \text{if } Y_t \le \mathrm{CoVaR}_{Y_t|X_t}(p_1, p_2) \end{cases}.
$$

We test the correct unconditional coverage of VaR forecasts at level $1 - p_1$ by specifying the following hypotheses:

$$H_0 : E[I_{X_t}] \equiv \pi = p_1 \quad \text{versus} \quad H_A : E[I_{X_t}] \equiv \pi \neq p_1.$$

Under the null hypothesis, the likelihood-ratio test statistic is given by (for more details, see Kupiec [1995]):

$$\mathrm{LR}_{\mathrm{uc}} = 2\left(\mathcal{L}(\hat{\pi}; I_{X_1}, I_{X_2}, \ldots, I_{X_n}) - \mathcal{L}(p_1; I_{X_1}, I_{X_2}, \ldots, I_{X_n})\right) \sim \chi^2_1, \quad \text{for large } n, \tag{4.3}$$

where $\mathcal{L}(x; I_{X_1}, I_{X_2}, \ldots, I_{X_n}) = \log\left[x^{n_1}(1-x)^{n-n_1}\right]$. The maximum-likelihood estimator $\hat{\pi}$ is n_1/T_1, where n_1 is the number of violations, that is $\#\{t : I_{X_t} = 1\}$.

Similarly, the null hypothesis to test the unconditional coverage property of CoVaR forecasts at confidence level $1 - p_2$ is

$$H_0 : E[I_{Y_t|X_t}] \equiv \pi = p_2,$$

and the maximum-likelihood estimation $\hat{\pi}$ under alternative hypothesis is n_2/n_1, where n_2 is the number of violations in the second "hit sequence", that is $\#\{t : I_{Y_t|X_t} = 1\}$.

In addition to the unconditional coverage test based on the binomial distribution, further calibration tests for risk measures are discussed in Nolde and Ziegel [2017]. They propose a two-stage procedures for assessment of risk measure forecasts, which can not only assess calibration of risk measure forecasts, but also allow for a comparison of several methods by supplementing traditional backtests with comparative backtests.

4.2 Data

We consider a partial panel of financial institutions studied in Acharya et al. [2017] with a market capitalization in excess of 5 billion USD as of the end of June 2007. The total number of firms in our sample is 8, including Bank OF AMERICA CORP (BAC), BANK NEW YORK INC (BK), AFLAC INC (AFL), ALLSTATE CORP (ALL), AMERICAN EXPRESS CO (AXP), FRANKLIN RESOURCES INC (BEN), GOLDMAN SACHS GROUP INC (GS) and T ROWE PRICE GROUP INC (TROW). The sample period is from June 26, 2000 to May

9, 2019, consisting of 4747 daily price observations for each institution. 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. Figure B.1 in Appendix B.1 gives the time series plots of daily losses for institutions and DJUSFN.

Before we do the estimation, it is essential for us to validate the tail dependence assumption of our samples. Figure 4.1displays scatter plots of losses for each institution and the market index. For presentation purpose, we standardized the daily losses by subtracting the sample mean and dividing the sample standard deviation. From the plots, for each institution, its dependence with DJUSFN seems to be persistent in both the upper and lower joint tails, especially for the BAC-DJUSFN and BEN-DJUSFN data pairs. This is indicative of the variables being tail dependent.

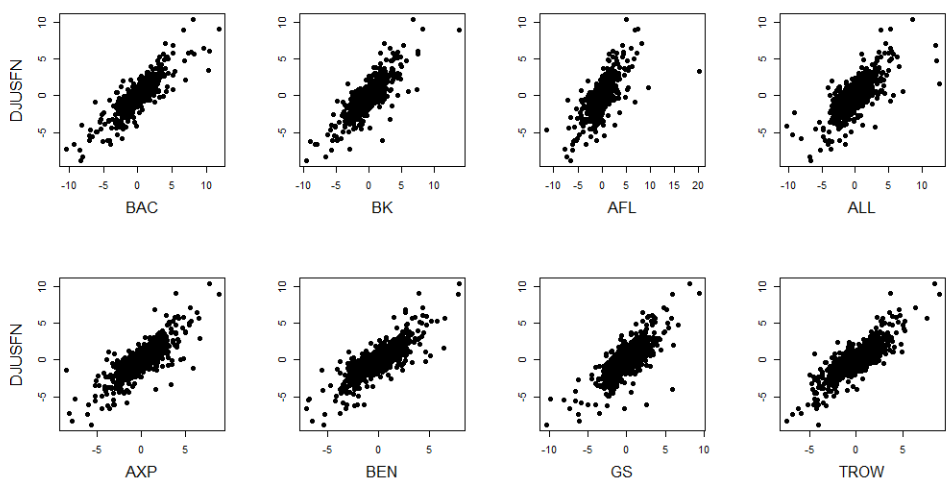


Figure 4.1: Scatter plots of standardized daily losses (%) for time series introduced in Section 4.2 over the period from June 27, 2000 to May 9, 2019.

One way to test for tail dependence is via estimating the extremal dependence measures and (Coles et al. [1999]). Let random variables X and Y denote, respectively, losses of a financial institution and a system proxy with joint df F and margins F_1 and F_2. The extremal dependence measures and are defined as

$$ = \lim_{t \to \infty} \frac{P\left(X > F_1^{-1}(1-1/t), Y > F_2^{-1}(1-1/t)\right)}{P\left(X > F_1^{-1}(1-1/t)\right)};$$

$$= \lim_{t \,!\, 1} \frac{2 \log \mathrm{P} \quad X > F_1^{\ 1}(1 - 1{=}t)}{\log \mathrm{P} \quad X > F_1^{\ 1}(1 - 1{=}t); Y > F_2^{\ 1}(1 - 1{=}t)} - 1;$$

where $0 \leq \quad \leq 1$ and $-1 < \quad \leq 1$ provided the limits exist. If $(X; Y)$ is tail independent, then $\quad = 0$ and $-1 < \quad < 1$; if $(X; Y)$ is tail dependent, then $\quad = 1$ and $0 < \quad < 1$. Define

$$(u) := 2 - \frac{\log C(u; u)}{\log u} \quad \text{and} \quad (u) := \frac{2 \log(1 - u)}{\log C(u; u)} - 1;$$

where $C(u_1; u_2) := F \quad F_1^{\ 1}(u_1); F_1^{\ 1}(u_2)$ for $(u_1; u_2) \in [0; 1]^2$ is the copula of bivariate df F and $C(u_1; u_2) = 1 - u_1 - u_2 + C(u_1; u_2)$ is the survival copula. It is shown (see Coles and Tawn [1991]) that

$$= \lim_{u \,!\, 1} \quad (u) \quad \text{and} \quad = \lim_{u \,!\, 1} \quad (u):$$

By replacing C and C with their empirical estimators, we can obtain simple estimators of $(\; ; \;)$. Figure 4.2 illustrates the (u) and (u) plots for the BAC-DJUSFN data, which are implemented with the function “chi.plot” in R package “evd”. The (u) and (u) plots for other institutions are given in Appendix B.2.

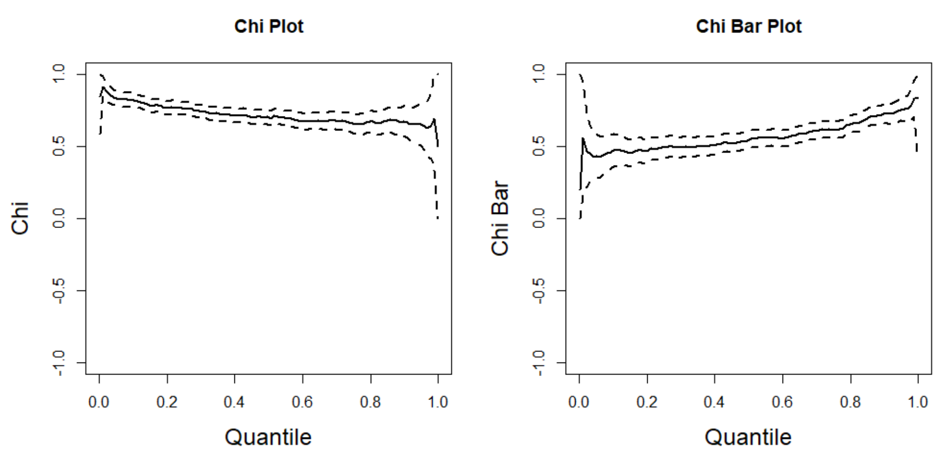


Figure 4.2: (u) and (u) plots for the BAC-DJUSFN data.

From theses plots, when taking sampling variability into account, the (u) estimates appear to be converging to one for $u \to 1$, especially in BEN-DJUSFN and DS-DJUSFN data. This justifies the assumption of tail dependence. However, the estimates of for the ALL-DJUSFN pair have a decreasing trend towards 0 when $u \to 1$, which suggests the possibility of tail independence. In order to further confirm the tail dependence assumption for this dataset, we re-estimate with an alternative method.

Ledford and Tawn [1996] established that under weak conditions

$$\mathrm{P}\quad X > F_1\ {}^{1}(1 - 1{=}t); Y > F_2\ {}^{1}(1 - 1{=}t) \quad \in RV\ _{1=}\ ; \tag{4.4}$$

where is called the residual dependence coefficient. It follows that $= 2 \quad - 1$. Transform the original random variables X and Y to unit Fréchet: $Z_1 = -1{=}\log F_1(X)$ and $Z_2 = -1{=}\log F(Y)$ and define $T = \min\{Z_1; Z_2\}$. We note that (4.4) is equivalent to

$$\mathrm{P}(Z_1 > z; Z_2 > z) = \mathrm{P}(T > z) \in RV\ _{1=}\ :$$

Hence, we can estimate (and) as the shape parameter for random variable T with the maximum likelihood estimator in the POT method. Figure 4.3 gives the plot of estimated versus threshold in the POT method for the ALL-DJUSFN data. From the graph, the 95% confidence band covers the value of one for thresholds exceeding roughly the 70%-quantile. This gives support for the earlier claim that the data could be assumed to be tail dependent.

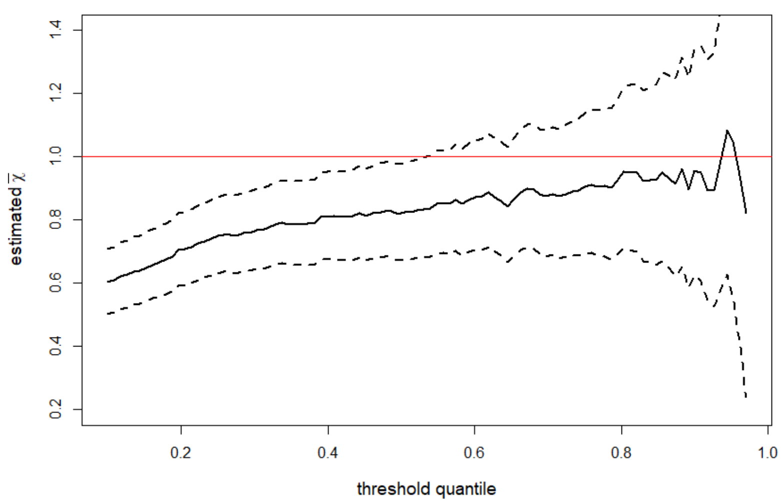


Figure 4.3: Maximum likelihood estimates of with 95% confidence bands based on the profile likelihood for the ALL-DJUSFN data.

4.3 Results

Firstly, we show the results of CoVaR estimation by treating observations in time series of daily losses for firms and a market proxy as i.i.d.. We use the whole dataset to do the estimation and then compare the CoVaR estimates under different parametric models by utilizing the unconditional coverage test in Section 4.1.

Rather than to selecting k in Hill estimator for tail index automatically with the two-step subsample bootstrap algorithm, in these studies, we want to see how the k value will affect the estimator of CoVaR. Due to this consideration, we assess stability of estimates to values of k by plotting k versus CoVaR estimates. All the five parametric models in Section 3.2 are used to do the estimation, and for each model, we consider three choices of m for the M-estimator: $m/n \in \{9\%, 20\%, 30\%\}$ for logistic distribution, $m/n \in \{14\%, 25\%, 35\%\}$ for Hüsler-Reiss distribution, $m/n \in \{9\%, 20\%, 30\%\}$ for bilogistic distribution, $m/n \in \{8\%, 20\%, 30\%\}$ for asymmetric logistic distribution and $m/n \in \{3\%, 15\%, 20\%\}$ for t distribution. For all datasets, $n = 4746$ is the sample size. The first value of m in each considered set corresponds to the best values in terms of RMSE, as it comes from the results of simulation studies in Section 3.2.1. We consider $\mathbf{p} = \mathbf{p}_n = (p_{1,n}, p_{2,n}) = (0.05, 0.05)$ and $(0.02, 0.05)$. For brevity of presentation, here we just show the graphs of BAC company in Figures 4.4 and 4.5. The additional graphs for the other seven institutions are given in Appendix B.3.

From the graphs, we can see that the lines for different m values nearly coincide with each other, especially for the logistic, Hüsler-Reiss and bilogistic distributions, indicating that the choice of m will not have a large effect on CoVaR estimation. Furthermore, for all datasets and parametric models, the CoVaR estimates tend to increase as k increases and have large variability for k values below 100. This is similar to the effect of k in the tail index estimation as we point out in Section 2.1.2: if k is large, the bias will be large, and if k is small, the variance will be large. In this situation, we hope to select a k value that balances bias and variance. From the graphs, it seems that the estimators are much more stable from around $k = 280$ to $k = 380$, thus choosing a k inside $[280, 380]$ seems to be suitable. This corresponds to $k/n \in [6\%, 8\%]$.

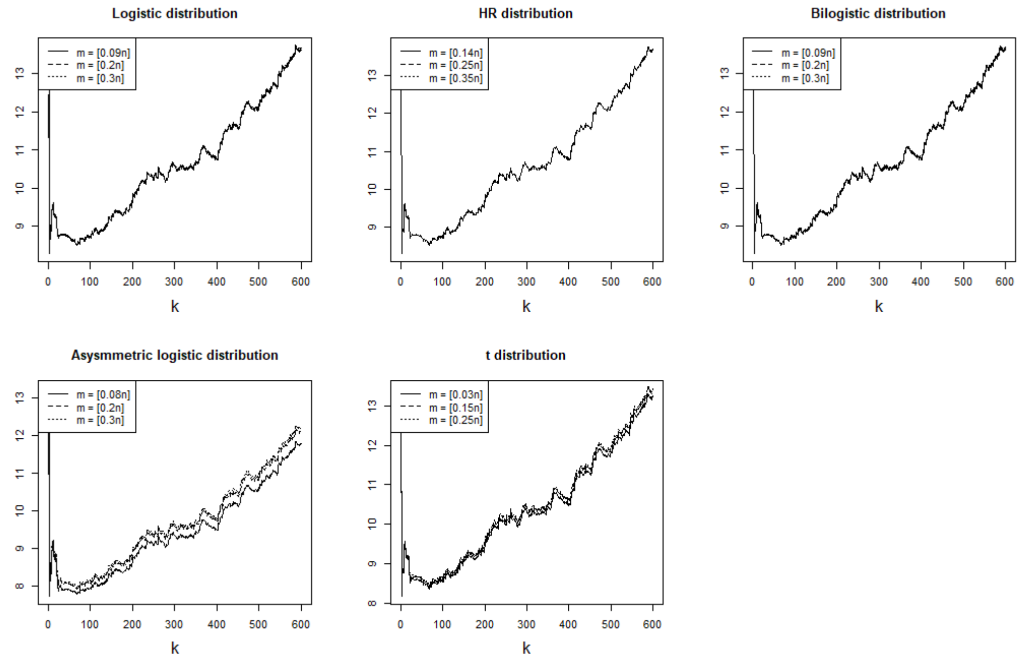


Figure 4.4: Estimates of CoVaR as a function of k at level $p_n = (0{:}05; 0{:}05)$ for different values of m for original BAC-DJUSFN data.

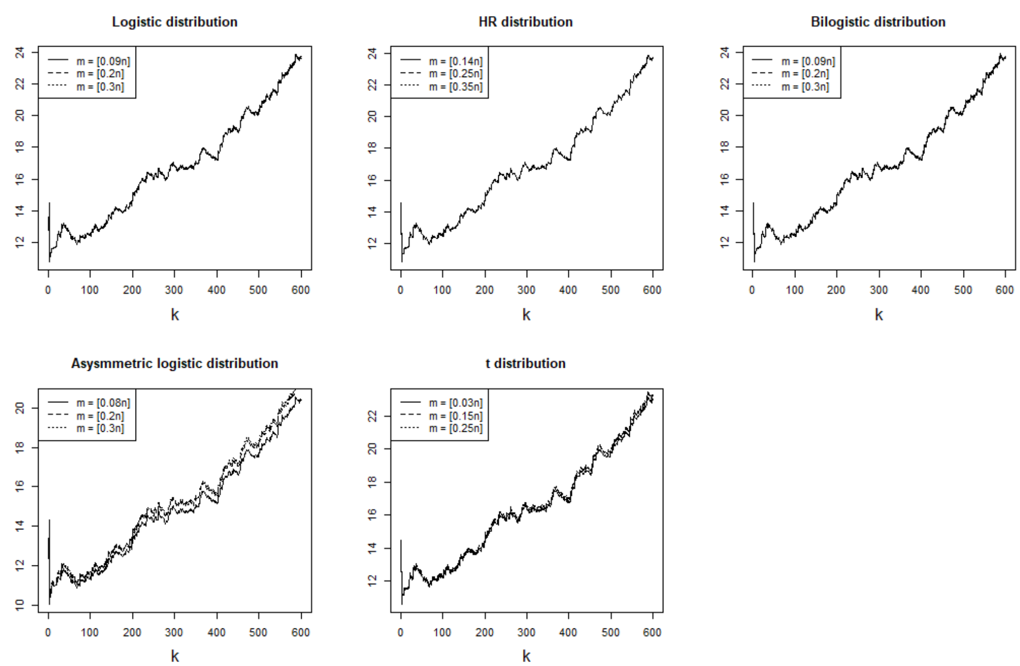


Figure 4.5: Estimates of CoVaR as a function of k at level $p_n = (0{:}02; 0{:}05)$ for different values of m for original BAC-DJUSFN data.

The next step is to evaluate the proposed estimator of CoVaR via the backtesting procedure using the unconditional coverage test. Based on the results above, for each parametric model and institution, we compute the CoVaR estimates with the first value of m in each considered set and $k = 350$. Before we perform the unconditional coverage test, we need to estimate the VaR of institutions, where we also use the nonparametric extreme quantile estimation method and let $k = 300$ (the k values are selected by plotting k versus estimates of VaR). We also backtest these VaR estimates to see how this nonparametric extreme quantile estimation method performs. Results are given in Table 4.1.

From the p-values of VaR backtesting, we can see that for all institutions, the unconditional coverage property for the VaR measure is satisfied at 5% significance level, indicating a good performance of the nonparametric extreme quantile estimation method. For both risk levels, the CoVaR estimates for the logistic, Hüsler-Reiss, bilogistic and t distribution over-estimate CoVaR as the values of E_n^b are smaller than the nominal value: $E_n \times 0.05$. From the p-values at risk level $\mathbf{p}_n = (0.05, 0.05)$, for all institutions, the CoVaR measure under the logistic distribution, the asymmetric logistic distribution and the t distribution satisfies the unconditional coverage property at 5% significance level, while for some institutions, the unconditional coverage property is rejected under the Hüsler-Reiss and the bilogistic distribution. Moreover, the asymmetric logistic distribution seems to produce more accurate estimates, as its E_n^b's are the closest to the nominal value. From the p-values at risk level $\mathbf{p}_n = (0.02, 0.05)$, only the CoVaR estimates under asymmetric logistic distribution satisfy the unconditional coverage property, indicating that for the datasets, asymmetric logistic distribution is the most suitable model among the considered five models, which is in line with the results at level $\mathbf{p}_n = (0.05, 0.05)$.

We should note that the financial time series we plot in Figure B.1 show volatility clustering and are dependent, which violates the i.i.d. assumption. We next design to remove the volatility and time dependence via an AR(1)-GARCH(1,1) filter and see how the proposed CoVaR estimator would perform on the resulting residuals. Let X_t denote the losses for an institution or the market index. By assuming the AR(1)-GARCH(1,1) model, the losses satisfy

$$X_t = \mu_t + \sigma_t Z_t, \qquad \mu_t = \phi_0 + \phi_1 X_{t-1}, \qquad \sigma_t^2 = \omega_0 + \omega_1(\sigma_{t-1} Z_{t-1})^2 + \omega_2 \sigma_{t-1}^2,$$

where the standardized vectors $\{Z_t\}_{t \in \mathbb{N}}$ are i.i.d. with zero mean and variance one. Parameters in the AR(1)-GARCH(1,1) filter are estimated using maximum likelihood assuming a standardized skew-t distribution (Fernández and Steel [1998]) for the innovations $\{Z_t\}$.

Table 4.1: Unconditional coverage tests for VaR of institutions and CoVaR based on raw data. The X and Y variables for VaR and CoVaR are 100 $\times$ log return. "Log, HR, Bilog, Alog, t" stands for logistic, Hüsler-Reiss, bilogistic, asymmertric and t distribution, respectively. E_n is the number of exceedances of the VaR estimate, and E_n^b is the number of joint exceedances of VaR and CoVaR estimates. Moreover, T represents the test statistic value in (4.3).

		p_n = (0:05; 0:05)						p_n = (0:02; 0:05)					
		VaR	CoVaR					VaR	CoVaR				
			Log	HR	Bilog	Alog	t		Log	HR	Bilog	Alog	t
BAC	estimate	3.486	10.659	10.699	10.662	9.414	10.406	5.675	17.003	17.067	17.008	15.019	16.599
	E_n=E_n^b	237	7	6	6	12	7	92	1	1	1	3	1
	T	0.001	2.434	3.684	3.684	0.002	2.434	0.093	4.294	4.294	4.294	0.664	4.294
	p-value	0.984	0.119	**0.055**	**0.055**	0.964	0.119	0.761	**0.038**	**0.038**	**0.038**	0.415	**0.038**
BK	estimate	3.213	10.598	10.696	10.604	9.192	10.192	4.729	16.905	17.063	16.915	14.664	16.258
	E_n=E_n^b	237	7	6	7	14	9	96	1	1	1	3	1
	T	0.001	2.434	3.684	2.434	0.389	0.784	0.013	4.619	4.619	4.619	0.815	4.619
	p-value	0.984	0.119	**0.055**	0.119	0.533	0.376	0.911	**0.032**	**0.032**	**0.032**	0.367	**0.032**
AFL	estimate	2.718	10.503	10.677	10.509	8.841	10.074	4.436	16.754	17.032	16.764	14.103	16.070
	E_n=E_n^b	249	7	6	7	15	10	98	1	1	1	3	1
	T	0.598	2.963	4.315	2.963	0.518	0.543	0.751	4.783	4.783	4.783	0.895	4.783
	p-value	0.439	**0.085**	**0.038**	**0.085**	0.472	0.461	0.751	**0.029**	**0.029**	**0.029**	0.344	**0.029**
ALL	estimate	2.279	10.383	10.656	10.395	8.639	9.974	3.733	16.563	16.999	16.582	13.780	15.911
	E_n=E_n^b	239	7	7	7	15	10	102	1	1	1	3	1
	T	0.013	2.520	2.520	2.520	0.761	0.354	0.526	5.113	5.1136	5.113	1.061	5.113
	p-value	0.910	0.112	0.112	0.112	0.383	0.552	0.468	**0.024**	**0.024**	**0.024**	0.303	**0.024**
AXP	estimate	3.196	10.624	10.698	10.629	9.282	10.222	4.810	16.947	17.066	16.955	14.806	16.307
	E_n=E_n^b	246	7	6	7	13	8	99	1	1	1	3	1
	T	0.332	2.828	4.154	2.828	0.041	1.796	0.177	4.865	4.865	4.865	0.936	4.865
	p-value	0.565	**0.093**	**0.042**	**0.093**	0.839	0.180	0.674	**0.027**	**0.027**	**0.027**	0.333	**0.027**
BEN	estimate	3.157	10.644	10.699	10.649	9.254	10.296	4.589	16.979	17.067	16.987	14.763	16.424
	E_n=E_n^b	248	7	6	7	12	7	103	1	1	1	3	1
	T	0.501	2.918	4.261	2.918	0.014	2.918	0.683	5.196	5.196	5.196	1.105	5.196
	p-value	0.479	**0.087**	**0.039**	**0.088**	0.907	**0.088**	0.409	**0.023**	**0.023**	**0.023**	0.293	**0.023**
GS	estimate	3.228	10.567	10.693	10.572	9.039	9.956	4.709	16.856	17.057	16.864	14.419	15.881
	E_n=E_n^b	257	7	6	6	12	7	97	1	1	1	3	1
	T	1.678	3.335	4.751	4.751	0.061	3.335	0.046	4.701	4.701	4.701	0.855	4.701
	p-value	0.195	**0.068**	**0.029**	**0.029**	0.806	**0.068**	0.830	**0.030**	**0.030**	**0.030**	0.355	**0.030**
TROW	estimate	3.227	10.658	10.702	10.661	9.408	10.220	4.781	17.002	17.071	17.006	15.008	16.302
	E_n=E_n^b	246	7	6	6	12	8	103	1	1	1	3	1
	T	0.332	2.828	4.154	4.154	0.008	1.796	0.683	5.196	5.196	5.196	1.105	5.196
	p-value	0.565	**0.093**	**0.042**	**0.042**	0.930	0.180	0.409	**0.023**	**0.023**	**0.023**	0.293	**0.023**

With the estimates of μ_t and σ_t, we can obtain the sequence of residuals as: $\hat{Z}_t = (X_t - \hat{\mu}_t)/\hat{\sigma}_t$. The following estimation is based on these realized residuals $\{\hat{Z}_t\}$. We give the time series plots of realized residuals for institutions and market index in Figure B.2 of Appendix B.1.

Before we perform the estimation, we produce the scatter plots of institution's residuals versus market's residuals to validate the tail dependence assumption. Graphs are shown in Figure 4.6. We can see that the upper tail dependence is obvious for each pair.

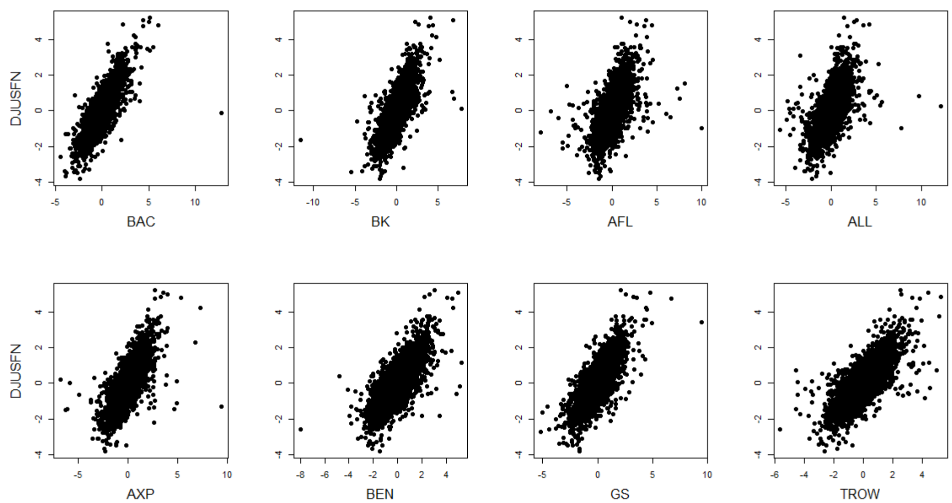


Figure 4.6: Scatter plots of realized residuals from time series introduced in Section 4.2 over the period from June 27, 2000 to May 9, 2019.

Remark 2. Throughout the nonparametric extreme quantile estimation we discussed in Section 2.1.3 and the analysis we constructed before, we estimate the VaR at a confidence level $1-p$ for the i.i.d. random variables $Y_1, Y_2, \ldots, Y_n$ via formula:

$$\widehat{VaR}_Y(1-p) = Y_{n,n-k}\left(\frac{k}{np}\right)^{\hat{\gamma}_n(k)},$$

where $Y_{n,1} \leq Y_{n,2} \leq \ldots \leq Y_{n,n}$ are the order statistics and $\hat{\gamma}_n(k)$ is the Hill estimator of the tail index with sample fraction k. However, a more general way to estimate the VaR is

$$\widehat{VaR}_Y(1-p) = Y_{n,n-k_2}\left(\frac{k_2}{np}\right)^{\hat{\gamma}_n(k_1)}, \tag{4.5}$$

where the intermediate sequences $k_1 = k_{1,n}, k_2 = k_{2,n}$ can be different. From this formula, to estimate the VaR, we first select a proper k_1 for the tail index using Hill plot or the bootstrap-based method. Then, we put the estimates $\hat{}_n(k_1)$ into (4.5) and do sensitivity analysis between k_2 and VaR estimates to select a proper value of k_2 and get an accurate estimate of VaR. In the following estimation, we will apply this procedure to estimate the VaR of market index and institutions. We denote k_1^i and k_2^i as the sample fractions of the tail index estimate and VaR estimate for institution i ($i \in \{$AFL, ALL, AXP, BAC, BEN, BK, GS, TROW$\}$), respectively. The parameter k_1^s and k_2^s are denoted as the sample fractions of the tail index estimate and VaR estimate for the market index, respectively.

For each institution, we consider $p_n = (0:02; 0:05)$ and the first value of m in each considered set. We select the k_1^i's and k_1^s through the stability of Hill plot and the results are: $k_1^s = 230$, $k_1^{\text{AFL}} = 230$, $k_1^{\text{ALL}} = 230$, $k_1^{\text{AXP}} = 240$, $k_1^{\text{BAC}} = 240$, $k_1^{\text{BEN}} = 240$, $k_1^{\text{BK}} = 245$, $k_1^{\text{GS}} = 250$ and $k_1^{\text{TROW}} = 240$. To select k_2^s, we plot VaR estimates as well as the plots of CoVaR estimates under different parametric model assumptions as a function of k_2^s. For the sake of brevity, we also just show the graphs of BAC company here in Figure 4.7. The others are shown in Figure B.6 in Appendix B.3.

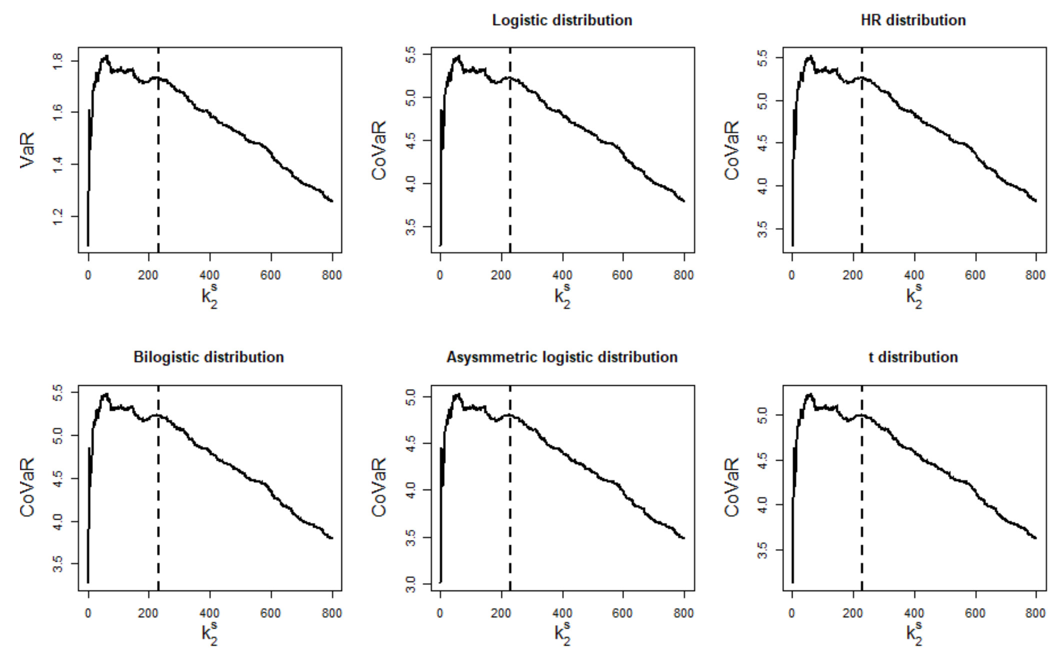


Figure 4.7: Estimates of CoVaR as a function of k_2^s at level $p_n = (0:02; 0:05)$ for estimated residuals from BAC-DJUSFN data. The dotted vertical line represent the $k_2^s = k_1^s = 230$.

From the plots, firstly we can see that the curves of CoVaR estimates under different model assumptions have the same shape as the curve of VaR estimates. This is sensible, as the quotient between CoVaR estimator and VaR estimator is ^ ^, which is uncorrelated with the value of k_2^s. Moreover, the curves seem to be stable from around 100 to 250, and then decrease linearly. Thus, it seems that selecting $k_2^s = k_1^s = 230$ is reasonable for all pairs. Compared with the graphs in Figures 4.4 and 4.5, these plots are more informative to select a proper value of sample fraction, as the stable region seems to be more obvious. This may indicate that constructing sensitivity analysis following procedures in Remark 2 is more suitable for the proposed CoVaR estimation methodology.

To select the value of k_2^i, we plot the VaR estimates against sample fraction k_2^i and the graphs suggest selecting $k_2^i = 200$ for all i. The next step is to assess the CoVaR estimates with the unconditional coverage test. Based on the results above, we put the CoVaR estimates with $k_2^s = 230$ into backtesting. The results are given in Table 4.2.

From the test results of estimated VaR of institutions, all VaR estimates perform well and satisfy the unconditional coverage property, although the VaR estimates for company GS appear to be slightly overestimate. This indicates that $k_2^i = 200$ for all i is a good choice here. From the estimates of CoVaR, it seems that the logistic, Hüsler-Reiss and bilogistic distributions have similar performance, as their estimates for each institution are close to each other. However, from the results of E_n^b and p-value of the unconditional coverage test, we can see that the estimates under these three parametric models over-estimate CoVaR and the unconditional coverage test is rejected at 5% significance level, especially under the Hüsler-Reiss distribution. However, the estimates under asymmetric logistic and t distributions have E_b^s that are much closer to the nominal value: $E_n \times 0.05$, and all pass the unconditional coverage test at 5% significance level, indicating more accurate estimation of CoVaR. These are similar to the backtesting results of the raw data in Table 4.1. However, for the asymmetric logistic and t distributions, although they seem to have the same E_n^b value for some institutions, their estimates are quite different from each other. This means that although this traditional calibration backtesting procedure allows us to assess a particular risk measure estimation method, it may fail to provide an adequate comparison of the results among several methods.

To make further comparisons across these five parametric models, we then calculate an average quantile score for each institution under each model. Following the notations in Section 4.1, define $\mathcal{I} = \{t : I_{X_t} = 1\}$ and $n_1 = |\mathcal{I}|$. The average quantile score S is expressed as

$$S = \frac{1}{n_1} \sum_{t \in I} S\left(\widehat{\text{CoVaR}}_{Y_t|X_t}(p_1, p_2), Y_t\right),$$

Table 4.2: Unconditional coverage tests for VaR of institutions and CoVaR based on realized residuals at level $p_n = (0.02, 0.05)$. "Log, HR, Bilog, Alog, t" stands for logistic, Hüsler-Reiss, bilogistic, asymmertric and t distribution, respectively. E_n is the number of exceedances of the VaR estimate, and E_n^b is the number of joint exceedances of VaR and CoVaR estimates. Moreover, T represents the test statistic value in (4.3).

		VaR	CoVaR				
			Log	HR	Bilog	Alog	t
BAC	estimate	2.239	5.220	5.257	5.223	4.791	4.989
	E_n=E_n^b	90	1	1	1	6	3
	T	0.265	4.133	4.133	4.133	0.479	0.593
	p-value	0.607	**0.042**	**0.042**	**0.042**	0.489	0.441
BK	estimate	2.178	5.168	5.251	5.173	4.638	4.879
	E_n=E_n^b	98	1	1	1	6	4
	T	0.101	4.782	4.782	4.782	0.243	0.185
	p-value	0.751	**0.029**	**0.029**	**0.029**	0.622	0.667
AFL	estimate	2.147	5.028	5.206	5.252	4.374	4.579
	E_n=E_n^b	97	1	**0**	**0**	4	4
	T	0.046	4.701	1.373	1.373	0.166	0.166
	p-value	0.830	**0.030**	**NaN**	**NaN**	0.683	0.683
ALL	estimate	2.088	4.960	5.197	4.974	4.271	4.403
	E_n=E_n^b	94	2	**0**	2	4	4
	T	0.009	2.063	1.370	2.063	0.115	0.115
	p-value	0.924	0.151	**NaN**	0.151	0.734	0.734
AXP	estimate	2.200	5.162	5.249	5.166	4.606	4.885
	E_n=E_n^b	96	1	1	1	6	4
	T	0.012	4.619	4.619	4.619	0.294	0.148
	p-value	0.911	**0.032**	**0.032**	**0.032**	0.588	0.700
BEN	estimate	2.221	5.192	5.253	5.195	4.693	4.902
	E_n=E_n^b	93	1	1	1	6	3
	T	0.040	4.375	4.375	4.375	0.379	0.701
	p-value	0.842	**0.037**	**0.037**	**0.037**	0.538	0.402
GS	estimate	2.181	5.197	5.252	5.200	4.7078	4.908
	E_n=E_n^b	85	**0**	**0**	**0**	5	2
	T	1.096	1.361	1.361	1.361	0.132	1.547
	p-value	0.295	**NaN**	**NaN**	**NaN**	0.716	0.214
TROW	estimate	2.146	5.181	5.253	5.183	4.686	4.907
	E_n=E_n^b	97	1	1	1	6	3
	T	0.046	4.701	4.701	4.701	0.268	0.855
	p-value	0.830	**0.030**	**0.030**	**0.030**	0.605	0.355

where $S(r, x)$ is a consistent scoring function for VaR at level $1 - p_2$, which is expressed as (for more details, see Nolde and Ziegel [2017]):

$$S(r, x) = (p_2 - 1\{x > r\})\, r + 1\{x > r\} x.$$

A lower score indicates better performance. The results are reported in Table 4.3. From the values of S, the asymmetric logistic distribution and the t distributions has superior performance on CoVaR estimation, compared with the other three distributions. In addition, among the first four extreme value distributions, it seems that the multi-parameter model (asymmetric logistic distribution) has the better performance than the one-parameter models (logistic and HR distributions). This is making sense, as from the M-estimates of model parameters, for each dataset, $\hat{}_1$ is around 0.1 while $\hat{}_2$ is around 0.6, showing big difference and thus obvious asymmetry. Furthermore, the HR distribution seems to have the worst performance among these five distributions for the most of institutions. These results are consistent with the results of unconditional coverage tests in Table 4.2. Finally, the model with the best performance seems to vary across different institutions, but can only be either the asymmetric logistic distribution or the t distribution.

Table 4.3: The average quantile scores S of CoVaR estimates based on realized residuals at level $\mathbf{p}_n = (0.02, 0.05)$.

	BAC	BK	AFL	ALL	AXP	BEN	GS	TROW
Log	0.262	0.256	0.253	0.251	0.259	0.260	0.259	0.260
HR	0.263	0.263	0.260	0.260	0.263	0.263	0.263	0.263
Bilog	0.262	0.260	0.263	0.251	0.259	0.261	0.260	0.260
Alog	**0.253**	0.254	0.244	**0.227**	0.255	0.254	**0.249**	0.254
t	0.255	**0.252**	**0.243**	0.244	**0.252**	**0.253**	0.249	**0.253**

Chapter 5

Conclusion

This thesis develops a semi-parametric method for estimating a high quantile of a conditional distribution given that one of the components of a bivariate random vector is extreme, and then applies this method to systemic risk estimation with CoVaR as a risk measure. The methodology rests on asymptotic results from multivariate EVT, combining parametric modelling of the tail dependence function to address the issue of data sparsity in the joint tail regions and nonparametric univariate tail estimation techniques.

The main advantage of EVT-based estimators is that they capture useful information in the tail of the data without restricting the behaviour of the central part. We note that there is another EVT-based estimation method for CoVaR, which makes use of the limit expression for the conditional probability given that one of the components of a bivariate random vector is extreme; see Nolde and Zhang [2018]. Although this method is shown to provide a competitive alternative to a flexible fully-parametric method, its requirement of multivariate regular variation imposes the restriction of the same tail index for both the institutional and system losses, and a somewhat restrictive parametric assumption on the extremal dependence structure. Thus, another advantage of the framework in our methodology is that it allows to balance model uncertainty under a milder condition of multivariate domain of attraction.

In the simulation studies in Section 3.2, it was shown that the variance of the proposed CoVaR estimator is dominated by the variability in the estimation of the high quantile. We also observed a minor positive bias under all models used in data simulation. The methodology has been applied to real data including log-returns on eight US financial firms with returns on the

Dow Jones US Financials Index used as market proxy. A test for unconditional coverage was carried out on CoVaR estimates under several parametric model assumptions on the tail dependence function. It was found that there were minor differences among the CoVaR estimates under symmetric models, and the asymmetric model showed the best performance, since the backtest for unconditional coverage was passed under the asymmetric logistic distribution for all financial firms. A similar analysis was conducted after applying an AR(1)-GARCH(1,1) filter to the individual time series. Through the backtesting results, the asymmetric logistic distribution and t distribution were found to have superior performance on the resulting CoVaR estimates and passed the unconditional coverage test for financial institutions. The other three models also resulted in similar CoVaR estimates and the unconditional coverage property were rejected under these three models.

As the next step, we would like to conduct a more rigorous backtesting of the proposed CoVaR estimation procedure in order to be able to identify the best performing model for the tail dependence function, as well as to compare our methodology with other available approaches. Through the simulation studies, we were able to assess finite sample properties of the proposed estimator. In the follow-up work, we aim to study asymptotic properties of the estimator, including consistency and normality. Finally, we are interested in extending our methodology to the case where the losses of an institution and of the system exhibit tail independence as the current framework only applies to situations in which the tail dependence function is not identically zero, i.e., only in the presence of tail dependence.

Bibliography

B. Abdous, A.L. Fougères, and K. Ghoudi. Extreme behaviour for bivariate elliptical distributions. *Canadian Journal of Statistics*, 33:317–334, 2005. → page 3

V.V. Acharya, R. Engle, and M. Richardson. Capital shortfall: A new approach to ranking and regulating systemic risks. *American Economic Review*, 102:59–64, 2012. → page 1

V.V. Acharya, L.H. Pedersen, T. Philippon, and M. Richardson. Measuring systemic risk. *The Review of Financial Studies*, 30:2–47, 2017. → pages 1, 48

T. Adrian and M.K. Brunnermeier. CoVaR. Technical report, National Bureau of Economic Research, 2011. → page 2

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. → page 19

V. Barnet. The ordering of multivariate data (with discussion). *Journal of the Royal Statistics Society, Series A*, 139:318–354, 1976. → page 14

J. Beirlant, Y. Goegebeur, J. Segers, and J.L. Teugels. *Statistics of Extremes: Theory and Applications*. John Wiley & Sons, 2006. → pages 9, 10, 13, 15

S. Benoit, J.E. Colliard, C. Hurlin, and C. Pérignon. Where the risks lie: A survey on systemic risk. *Review of Finance*, 21:109–152, 2017. → page 1

P.F. Christoffersen. Evaluating interval forecasts. *International Economic Review*, pages 841–862, 1998. → page 47

S.G. Coles and J.A. Tawn. Modelling extreme multivariate events. *Journal of the Royal Statistical Society: Series B*, 53:377–392, 1991. → pages 20, 50

S.G. Coles, J. Heffernan, and J.A. Tawn. Dependence measures for multivariate extremes. *Extremes*, 2:339–365, 1999. → page 49

J. Danielsson, L. de Haan, L. Peng, and C.G. de Vries. Using a bootstrap method to choose the sample fraction in tail index estimation. *Journal of Multivariate Analysis*, 76:226–248, 2001. → pages 4, 11, 12

B. Das and S.I. Resnick. Conditioning on an extreme component: Model consistency with regular variation on cones. *Bernoulli*, 17:226–252, 2011. → page 3

Anthony C Davison and Richard L Smith. Models for exceedances over high thresholds. *Journal of the Royal Statistical Society: Series B (Methodological)*, 52(3):393–425, 1990. → page 42

O. De Bandt and P. Hartmann. Systemic risk: a survey. 2000. → page 1

L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer Science & Business Media, 2006. → pages 3, 6, 24, 25

L. de Haan and S.I. Resnick. Second-order regular variation and rates of convergence in extreme-value theory. *The Annals of Probability*, 24:97–124, 1996. → page 10

Stefano Demarta and Alexander J McNeil. The t copula and related copulas. *International statistical review*, 73(1):111–129, 2005. → page 19

F.X. Diebold, T. Schuermann, and J.D. Stroughair. Pitfalls and opportunities in the use of extreme value theory in risk management. *The Journal of Risk Finance*, 1:30–35, 2000. → page 5

J.H. Einmahl, A. Krajina, and J. Segers. A method of moments estimator of tail dependence. *Bernoulli*, 14:1003–1026, 2008. → pages 4, 20, 36

J.H. Einmahl, A. Krajina, and J. Segers. An m-estimator for tail dependence in arbitrary dimensions. *The Annals of Statistics*, 40:1764–1793, 2012. → pages 4, 20, 22, 30, 32, 36

P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events: for Insurance and Finance*. Springer Berlin Heidelberg, 1997. → pages 2, 5, 8, 41

C. Fernández and M. Steel. On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93:359–371, 1998. → page 54

R.A. Fisher and L.H.C. Tippett. 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 University Press, 1928. → page 6

G. Girardi and A.T. Ergün. Systemic risk measurement: Multivariate GARCH estimation of CoVaR. *Journal of Banking & Finance*, 37:3169–3180, 2013. → pages 2, 23, 46, 47

B. Gnedenko. Sur la distribution limite du terme maximum d'une serie aléatoire. *Annals of Mathematics*, pages 423–453, 1943. → pages 6, 8

E.J. Gumbel. Distributions des valeurs extremês en plusiers dimensions. *Publ. Inst. Statist. Univ. Paris*, 9:171–173, 1960. → pages 14, 17

J.E. Heffernan and S.I. Resnick. Limit laws for random vectors with an extreme component. *The Annals of Applied Probability*, 17:537–571, 2007. → page 3

J.E. Heffernan and J.A. Tawn. A conditional approach for multivariate extreme values (with discussion). *Journal of the Royal Statistical Society: Series B*, 66:497–546, 2004. → page 3

B.M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, pages 1163–1174, 1975. → pages 4, 9

X. Huang. Statistics of bivariate extremes. *Tinbergen Institute Research Series*, 22, 1992. → page 15

J. Hüsler and R.D. Reiss. Maxima of normal random vectors: between independence and complete dependence. *Statistics & Probability Letters*, 7:283–286, 1989. → page 17

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. → page 7

H. Joe, R.L. Smith, and I. Weissman. Bivariate threshold methods for extremes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 54:171–183, 1992. → page 20

H. Joe, H. Li, and A.K. Nikoloulopoulos. Tail dependence functions and vine copulas. *Journal of Multivariate Analysis*, 101:252–270, 2010. → page 3

G.G. Kaufman and K.E. Scott. What is systemic risk, and do bank regulators retard or contribute to it? *The Independent Review*, 7:371–391, 2003. → page 1

S. Kotz and S. Nadarajah. *Multivariate t-distributions and Their Applications*. Cambridge University Press, 2004. → page 68

K. Kuester, S. Mittnik, and M.S. Paolella. Value-at-risk prediction: A comparison of alternative strategies. *Journal of Financial Econometrics*, 4:53–89, 2006. → pages 38, 47

P. Kupiec. Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3, 1995. → pages 47, 48

A.W. Ledford and J.A. Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83:169–187, 1996. → pages 20, 50

G. Mainik and E. Schaanning. On dependence consistency of CoVaR and some other systemic risk measures. *Statistics & Risk Modeling*, 31:49–77, 2014. → pages 2, 23, 46

A.K. Nikoloulopoulos, H. Joe, and H. Li. Extreme value properties of multivariate t copulas. *Extremes*, 12:129–148, 2009. → pages 3, 16, 68

N. Nolde and J. Zhang. Conditional extremes in asymmetric financial markets. *Journal of Business & Economic Statistics*, pages 1–29, 2018. → pages 3, 61

N. Nolde and J. Ziegel. Elicitability and backtesting: Perspectives for banking regulation. *The annals of applied statistics*, 11:1833–1874, 2017. → pages 48, 60

S.A. Padoan. Multivariate extreme models based on underlying skew-t and skew-normal distributions. *Journal of Multivariate Analysis*, 102:977–991, 2011. → pages 68, 69

I.J. Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3: 119–131, 1975. → page 41

S.I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer, 1987. → pages 6, 24, 29

S.I. Resnick. *Heavy-tail Phenomena: Probabilistic and Statistical Modeling*. Springer Science & Business Media, 2007. → page 7

S.I. Resnick. *Extreme Values, Regular Variation and Point Processes*. Springer, 2013. → page 67

R.L. Smith. Extreme value theory. *Handbook of Applicable Mathematics*, 7:437–471, 1990. → page 18

R.L. Smith. Multivariate threshold methods. In *Extreme Value Theory and Applications*, pages 225–248. Springer, 1994. → page 20

P. Sun and C. Zhou. Diagnosing the distribution of GARCH innovations. *Journal of Empirical Finance*, 29:287–303, 2014. → page 46

J.A. Tawn. Bivariate extreme value theory: models and estimation. *Biometrika*, 75:397–415, 1988. → page 18

J.A. Tawn. Modelling multivariate extreme value distributions. *Biometrika*, 77:245–253, 1990. → page 15

R. Von Mises. La distribution de la plus grande de n valuers. *Rev. Math. Union Interbalcanique*, 1:141–160, 1936. → page 7

I. Weissman. Estimation of parameters and large quantiles based on the k largest observations. *Journal of the American Statistical Association*, 73:812–815, 1978. → page 13

E.B. Wilson. *Advanced Calculus*. Ginn, 1912. → page 16

Appendix A

Proofs

In this appendix, we collect several supplementaty theoretical results and derivations referred to in the main body of the thesis.

A.1 Regular Variation of Skew-t Distribution

Let $Y \sim \mathrm{ST}_1(0; 1; \;\; ; \;\;)$ be a random variable with density function

$$f_S T(y) = 2 f_T(y; \;\;) F_T \quad y \sqrt{\frac{\; + 1}{\; + y^2}}; \quad + 1 \quad ;$$

where $f_T(\cdot; \;\;)$ and $F_T(\cdot; \;\;)$ are the density function and df of a univariate Student's t random variable with degree of freedom.

It can be verified that the tail of the standard skew-t distribution, for $\; > 1$ and $\; \in \mathrm{R}$, is a regularly varying function with index $-$ (see p 13-17, Resnick [2013]), and $1 - F_{\mathrm{ST}}(y) \sim \; ^{1} y \;\; s(y; \;\; ; \;\;)$ as $y \to \infty$, where

$$s(y; \;\; ; \;\;) = \frac{2 \;\; \{(\;\; + 1)=2\}}{(\;\; =2)} \quad \frac{1}{y^2} + \frac{1}{\;} \quad ^{(\;\; +1)=2} F_{\mathrm{T}}(\;\; \sqrt{\;\; + 1}; 0; 1; \;\;)$$

is a slowly varying function.

Define $Q_{\ ;\ }(\cdot)$ as the lower quantile function of distribution $\mathrm{ST}_1(0; 1;\ ;\ )$. Then we can get as $u \to 0$ (see (4), Padoan [2011]),

$$Q_{\ ;\ }(1-u) \sim (s(\ ;\ )/u)^{1/\ }\ ; \tag{A.1}$$

where

$$s(\ ;\ ) = \frac{\{(\ +1)/2\}^{\ /2} F_T(\ \sqrt{\ +1}; 0; 1;\ )}{(\ /2)\sqrt{\ }}.$$

A.2 Proof of Proposition 2.2.1

The proof is following Nikoloulopoulos et al. [2009], where the TD function is derived in multivariate version.

Lemma A.2.1 (Kotz and Nadarajah [2004]). *Let* $X = (X_1^T; X_2^T)^T \in \mathrm{R}^d$ *follows the multivariate t distribution with parameter* $\ = 0$*, scale parameter and degree of freedom, where*

$$\ = \begin{pmatrix} _{11} & _{12} \\ _{21} & _{22} \end{pmatrix}; \quad _{22;1} = \ _{22} - \ _{21}\ _{11}^{\ 1}\ _{12};$$

If X_1 *and* X_2 *are k and* $d-k$ *dimensional sub-vectors, respectively, then*

$$\mathrm{P}\{X_2 \le x_2 | X_1 = x_1\} = F_T\left(\sqrt{\frac{\ +k}{\ + x_1^T\ _{11}^{\ 1} x_1}}\, x_{2;1}; 0;\ _{22;1};\ +k\right);$$

where $x_{2;1} = x_2 - \ _{21}\ _{11}^{\ 1} x_1$.

We know that the student's t distribution is a special case of the standard skew-t distribution with $\ = 0$. So from equation (A.1), the $Q_1(1-ux)$ and $Q_2(1-uy)$ in (2.14) satisfy:

$$Q_1(1-ux) \sim (s(0;\ )/ux)^{1/\ } \text{ and } Q_2(1-uy) \sim (s(0;\ )/uy)^{1/\ }\ ; \quad u \to 0.$$

Then the upper tail dependence function of the bivariate t distribution can be written as

$$\begin{aligned} R(x; y;\ ;\ ) = {} & x \lim_{u \to 0} \mathrm{P}\{Y > (s(0;\ )/(uy))^{1/\ } | X = (s(0;\ )/(ux))^{1/\ }\} \\ & + y \lim_{u \to 0} \mathrm{P}\{X > (s(0;\ )/(ux))^{1/\ } | Y = (s(0;\ )/(uy))^{1/\ }\}. \end{aligned} \tag{A.2}$$

From Lemma A.2.1, we obtain

$$\begin{aligned} &\mathrm{P}\{Y \le \{s(0;\nu)/(uy)\}^{1/\nu} \mid X = (s(0;\nu)/(ux))^{1/\nu}\} \\ &= F_T\left(\sqrt{\frac{(\nu+1)(s(0;\nu)/(ux))^{2/\nu}}{(1-\rho^2)(\nu+(s(0;\nu)/(ux))^{2/\nu})}}\left(\left(\frac{y}{x}\right)^{1/\nu}-\rho\right);0;1;\nu+1\right). \end{aligned} \tag{A.3}$$

Observing that

$$\lim_{u\to 0}\frac{(s(0;\nu)/ux)^{2/\nu}}{\nu+(s(0;\nu)/ux)^{2/\nu}} = 1,$$

we obtain the final expression for the upper tail function of the bivariate t distribution:

$$\begin{aligned} R(x,y;\rho,\nu) = {} & xF_T\left(\sqrt{\frac{\nu+1}{1-\rho^2}}\left(\rho-\left(\frac{y}{x}\right)^{1/\nu}\right);\nu+1\right) \\ & + yF_T\left(\sqrt{\frac{\nu+1}{1-\rho^2}}\left(\rho-\left(\frac{x}{y}\right)^{1/\nu}\right);\nu+1\right). \end{aligned} \tag{A.4}$$

A.3 Proof of Proposition 2.2.2

The proof is following Padoan [2011], where the TD function is derived in multivariate version.

Lemma A.3.1 (Padoan [2011]). *Let* $\mathbf{X} = (X_1, X_2)^T \sim \mathrm{ST}_2(0, \bar{\Omega}, \alpha, \nu)$, *where* $\bar{\Omega} = \begin{pmatrix} 1 & \omega \\ \omega & 1 \end{pmatrix}$. *Then for* $i \neq j \in \{1, 2\}$,

- $X_j \sim \mathrm{ST}_1(0, 1, \alpha_j^*, \nu)$, *where* $\alpha_j^* = \frac{\alpha_j + \omega\alpha_i}{\sqrt{1+\alpha_i^2(1-\omega^2)}}$.
- *The conditional distribution is univariate extended skew-t distribution:*

$$\mathrm{P}(X_i \le x_i \mid X_j = x_j) = F_{\mathrm{EST}}\left((x_i - \omega x_j)/\sqrt{Q_j(1-\omega^2)}; 0, 1, \sqrt{1-\omega^2}\,\alpha_i, \tau_{ij}, \nu+1\right),$$

where $Q_j = (\nu + x_j^2)/(\nu+1)$, $\tau_{ij} = \sqrt{(\nu+1)/(\nu+x_j^2)}\,(\alpha_j + \omega\alpha_i)x_j$.

Then from Lemma A.3.1, for the bivariate skew-t distribution in Example 2.2.6, we can rewrite the conditional probability in equation (2.14) as

$$
\begin{aligned}
&\lim_{u\to 0} \mathrm{P}\{U_2 > 1-uy \mid U_1 = 1-ux\} \\
=&\lim_{u\to 0} \mathrm{P}\{Y > Q_2(1-uy) \mid X = Q_1(1-ux)\} \\
=&\lim_{u\to 0} \bar{F}_{\mathrm{EST}}\left(\frac{\frac{Q_2(1-uy)}{Q_1(1-ux)} - \rho}{\sqrt{1-\rho^2}} \sqrt{\frac{\nu+1}{\frac{\nu}{Q_1(1-ux)^2}+1}}; 0, 1; \sqrt{1-\rho^2}\,\alpha_2, \tau_1, \nu+1 \right);
\end{aligned}
$$

where $\tau_1 = \sqrt{\frac{\nu+1}{\nu/q_{\xi_1}(1-uy)^2+1}}(\alpha_2 + \rho\alpha_1)$ and $\bar{F}_{\mathrm{EST}}$ is the survival function of F_{EST}. Furthermore, Q_1, Q_2 are the lower quantile functions for $\mathrm{ST}_1(0,1,\alpha_1,\nu)$ and $\mathrm{ST}_1(0,1,\alpha_2,\nu)$, respectively. The parameters α_1 and α_2 are given in Lemma A.3.1.

Applying regular variation of the tail of the skew-t distribution, we have

$$
\lim_{u\to 0} \frac{Q_2(1-uy)}{Q_1(1-ux)} = \left(\frac{y}{x}\right)^{1/\nu}
$$

and

$$
\lim_{u\to 0} \nu/Q_1(1-ux)^2 = 0
$$

where $x = xF_{\mathrm{T}}(\alpha_2\sqrt{\nu+1}; 0, 1, \nu)$ and $y = yF_{\mathrm{T}}(\alpha_1\sqrt{\nu+1}; 0, 1, \nu)$. Similarly, we can get the form of $\lim_{u\to 0} \mathrm{P}\{U_2 > 1-uy \mid U_1 = 1-ux\}$. Combining all the results, we have the final expression for the the upper tail dependence function of the bivariate skew-t distribution function as

$$
\begin{aligned}
R(x, y; \rho, \alpha_1, \alpha_2, \nu) = {} & x \cdot \bar{F}_{EST}\left(\frac{\sqrt{\nu+1}}{\sqrt{1-\rho^2}} \left(\left(\frac{x}{y}\right)^{1/\nu} - \rho \right); 0, 1; \sqrt{1-\rho^2}\,\alpha_2, \tau_1, \nu+1 \right) \\
& + y \cdot \bar{F}_{EST}\left(\frac{\sqrt{\nu+1}}{\sqrt{1-\rho^2}} \left(\left(\frac{y}{x}\right)^{1/\nu} - \rho \right); 0, 1; \sqrt{1-\rho^2}\,\alpha_1, \tau_2, \nu+1 \right);
\end{aligned}
\tag{A.5}
$$

where $x = xT(\bar{\alpha}_2\sqrt{\nu+1})$, $y = yT(\bar{\alpha}_1\sqrt{\nu+1})$, $\bar{\alpha}_1 = \frac{\alpha_1+\rho\alpha_2}{\sqrt{1+(1-\rho^2)\alpha_2^2}}$, $\bar{\alpha}_2 = \frac{\alpha_2+\rho\alpha_1}{\sqrt{1+(1-\rho^2)\alpha_1^2}}$, $\tau_1 = \sqrt{\nu+1}(\alpha_1 + \rho\alpha_2)$, $\tau_2 = \sqrt{\nu+1}(\alpha_2 + \rho\alpha_1)$.

A.4 Peaks-Over-Threshold Method

One way to define an extreme event is when the process exceeds a high threshold; i.e., events of the form $\{X > u\}$ for large threshold u. The question is can we find a limiting distribution for excesses over u, $X - u$, conditional on the event that X has exceeded u, i.e., $X > u$.

Definition 9. *The distribution with df of the form*

$$H_{\xi;\sigma}(y) = \begin{cases} 1 - \left(1 + \xi\frac{y}{\sigma}\right)_+^{-1/\xi}; & \xi \neq 0 \\ 1 - e^{-y/\sigma}; & \xi = 0 \end{cases}$$

is called a generalized Pareto distribution with scale parameter $\sigma > 0$ *andd shape parameter* $\xi \in \mathbb{R}$*, written as* $GP(\sigma; \xi)$*.*

Suppose $X_1; \ldots; X_n$ are i.i.d random variables with df F, where $F \in \mathcal{D}(G_\xi)$. That is to say, when n is large, we have

$$F^n(a_n x + b_n) \approx G_\xi(x) = \exp\left\{-(1 + \xi x)_+^{-1/\xi}\right\}: \tag{A.6}$$

Define $x_F = \sup\{x : F(x) < 1\}$. In particular, we can assume that there exist u, near x_F, such that (A.6) holds for all x such that $a_n x + b_n > u$. Let $a_n x + b_n = y$, $b_n = \mu$ and $a_n = \sigma$. Then we have

$$F(y) \approx \exp\left\{-\left(1 + \xi\frac{y-u}{\sigma}\right)_+^{-1/\xi}\right\} \approx 1 - \left(1 + \xi\frac{y-u}{\sigma}\right)_+^{-1/\xi}; \quad y > u:$$

Now consider the conditional df of excess over threshold u:

$$\begin{aligned} P(X - u \geq y | X > u) &= \frac{P(X > u + y)}{P(X > u)} = \frac{1 - F(u + y)}{1 - F(u)} \\ &\approx \left[\frac{\left(1 + \xi\frac{u+y}{\sigma}\right)_+}{\left(1 + \xi\frac{u}{\sigma}\right)_+}\right]^{-1/\xi} \\ &= \left(1 + \xi\frac{y}{\sigma_u}\right)_+^{-1/\xi}; \quad y > 0; \end{aligned}$$

where $\sigma_u = \sigma + \xi(u - \mu)$. From Definition 9, we can see that $X - u | X > u \sim GP(\sigma_u; \xi)$ for u large.

Theorem A.4.1 (Pickands-Balkema-de Haan)**.** *For* $\xi \in \mathbb{R}$*, the following are equivalent:*

1. *$F \in \mathcal{D}(G_{\xi})$.*
2. *There exists a positive function $\beta(\cdot)$ such that*

$$\lim_{u \uparrow x_F} \sup_{y \in (0, x_F - u)} |F_u(y) - H_{\xi, \beta(u)}(y)| = 0,$$

where $F_u(y) := \mathrm{P}(X - u \leq y | X > u)$, $y \leq 0$, $u < x_F \leq \infty$.

The above theorem reveals duality between the limiting distribution of maxima and excesses over a high threshold; if the normalized maxima have a $GEV(\mu, \sigma, \xi)$ distribution as the limit, then conditional excesses over a limiting threshold have a generalized Pareto distribution with the same shape parameter ξ.

Once we obtain the maximum likelihood estimates of ξ and σ_u, the $(1 - p)$-th quantile can be estimated with:

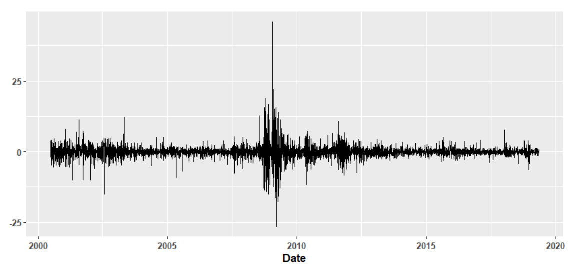
$$\hat{Q}_{1-p} = u + \frac{\hat{\sigma}_u}{\hat{\xi}} \left((p/\hat{\zeta}_u)^{-\hat{\xi}} - 1 \right), \quad \hat{\xi} \neq 0,$$

where $\hat{\zeta}_u = \#\{X_i > u\}/n$ is the empirical estimate of the probability of exceedance over threshold u.

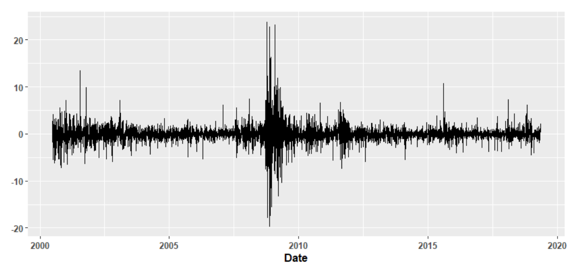
Appendix B

Tables and Figures

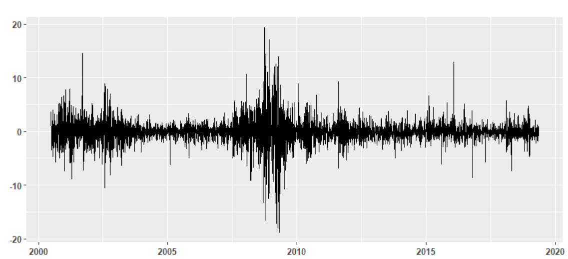
B.1 Time series plots



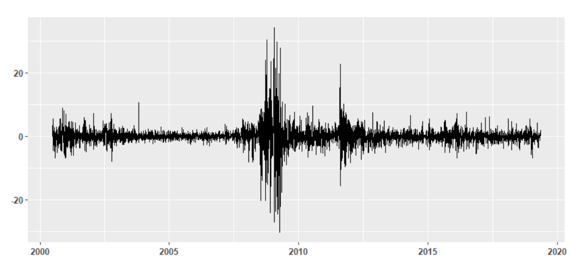
(a) AFL



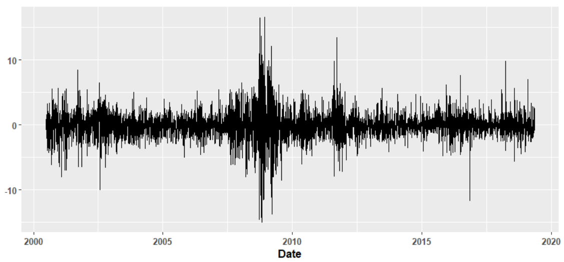
(b) ALL



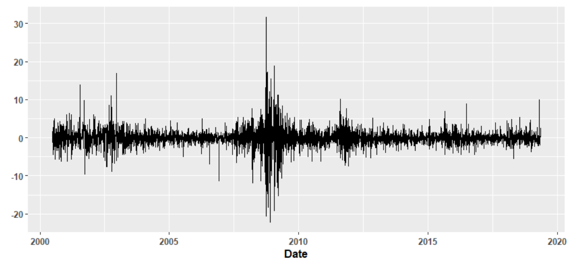
(c) AXP



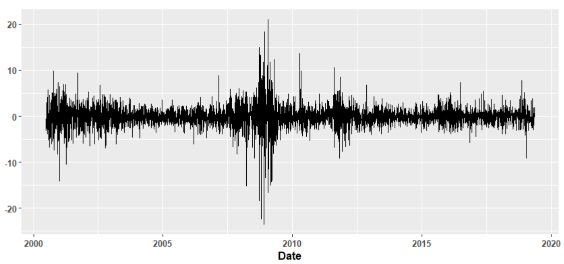
(d) BAC



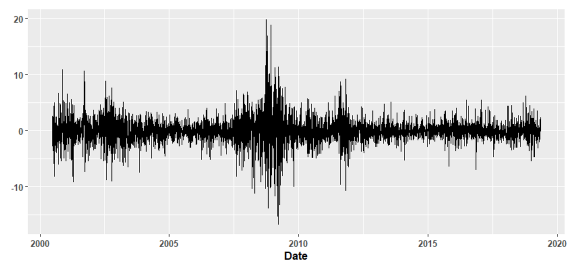
(e) BEN



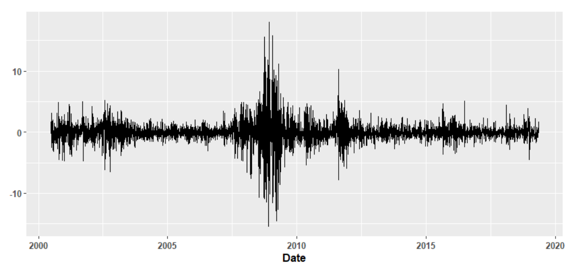
(f) BK



(g) GS

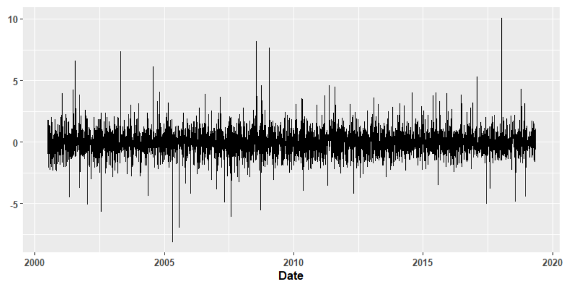


(h) TROW

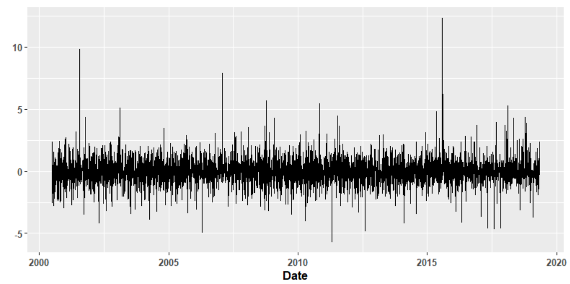


(i) DJUSFN

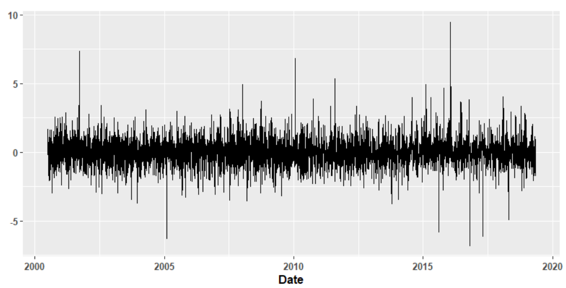
Figure B.1: Time series plots of daily losses for institutions and financial system.



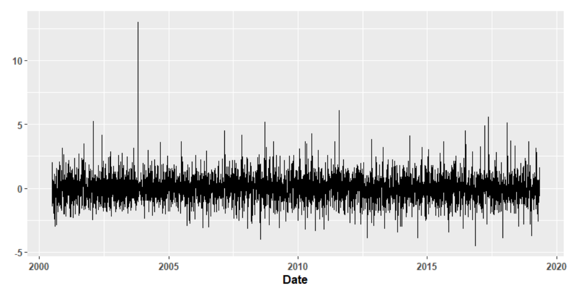
(a) AFL



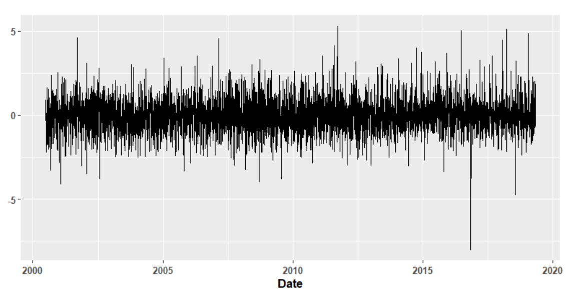
(b) ALL



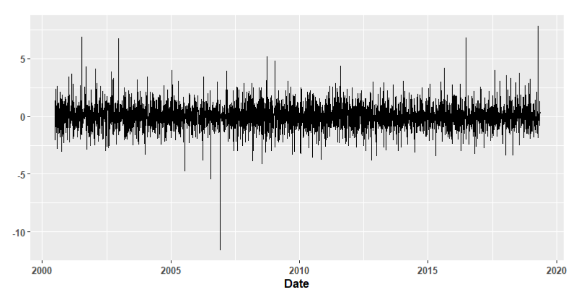
(c) AXP



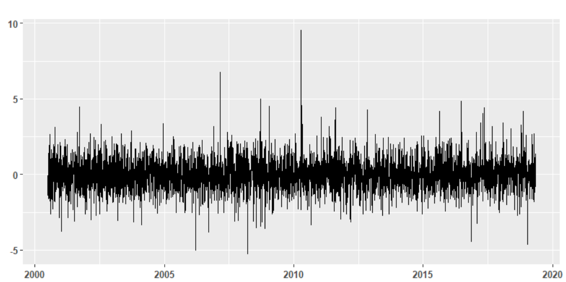
(d) BAC



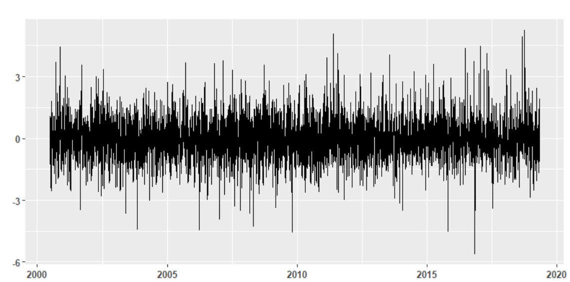
(e) BEN



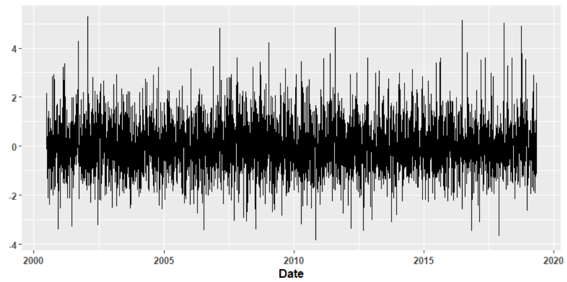
(f) BK



(g) GS



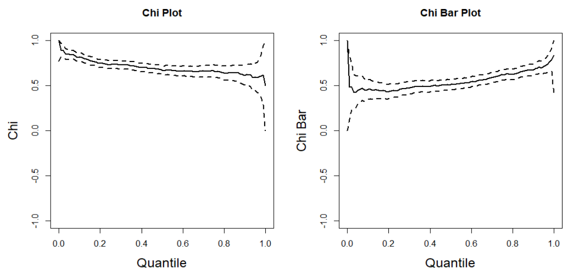
(h) TROW



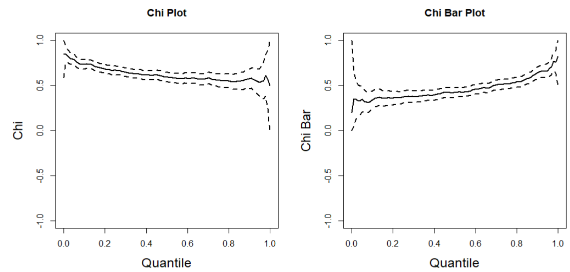
(i) DJUSFN

Figure B.2: Time series plots of realized residuals for institutions and financial system.

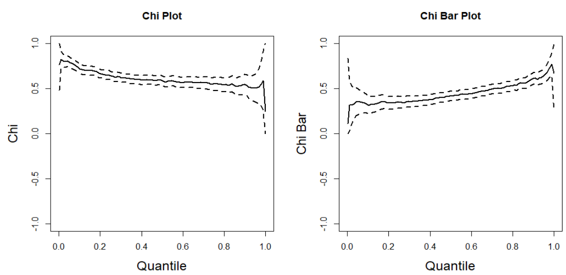
B.2 (u) and $^{-}(u)$ plots



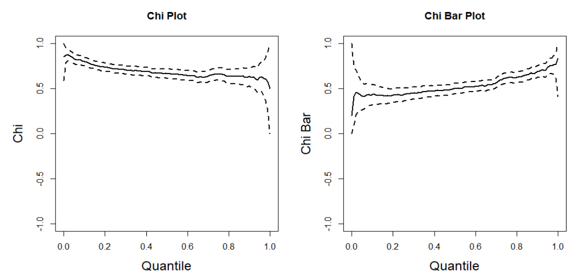
(a) BK-DJUSFN



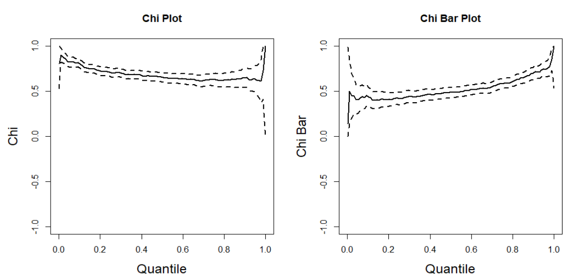
(b) AFL-DJUSFN



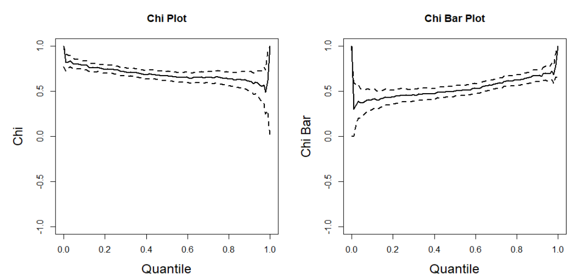
(c) ALL-DJUSFN



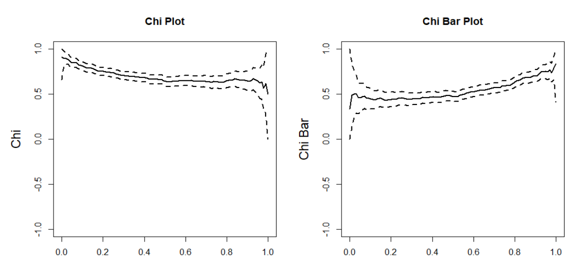
(d) AXP-DJUSFN



(e) BEN-DJUSFN



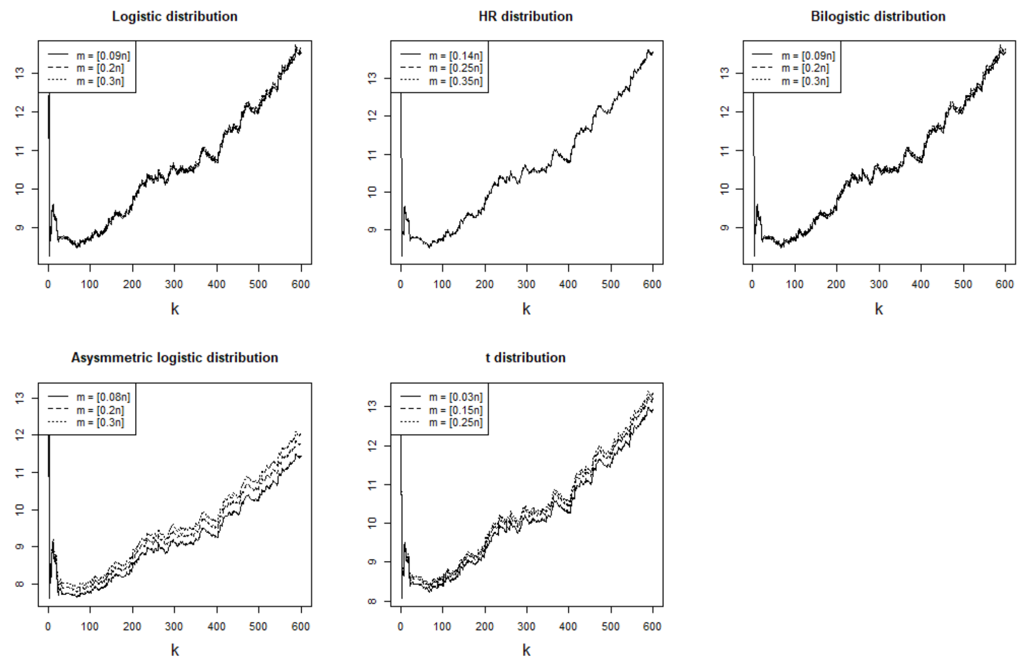
(f) GS-DJUSFN



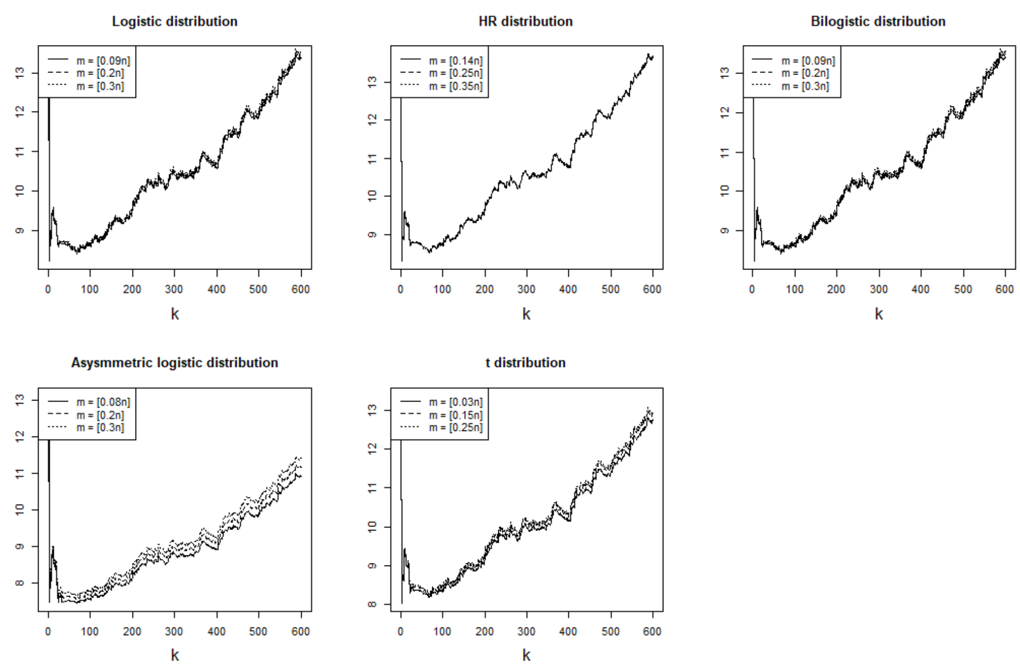
(g) TROW-DJUSFN

Figure B.3: (u) and (u) plots for other seven institutions

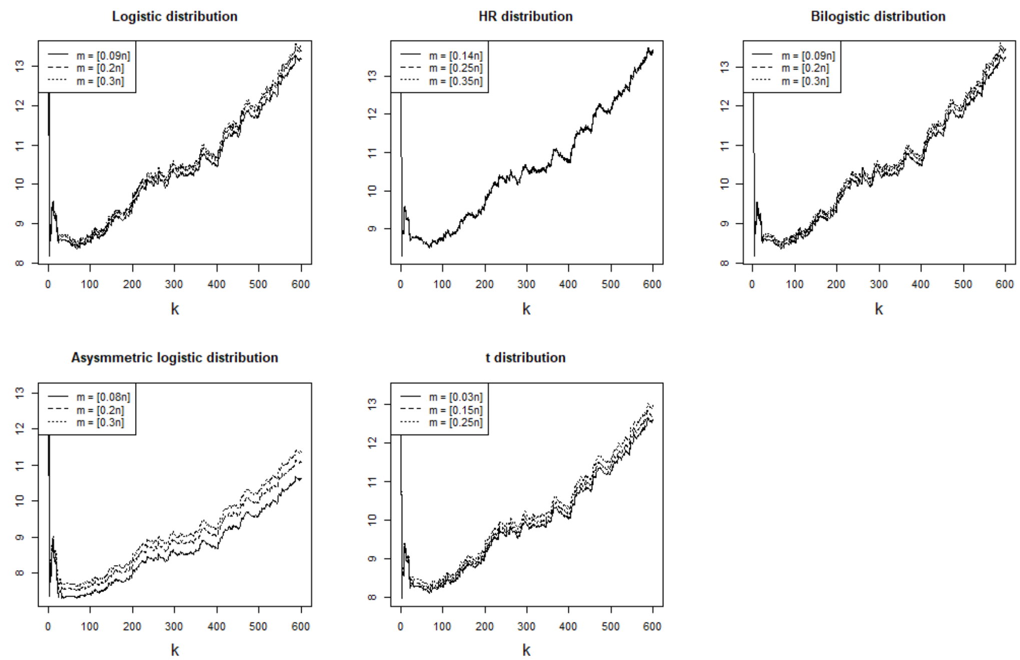
B.3 Plots of CoVaR estimates against the sample fraction for other seven institutions



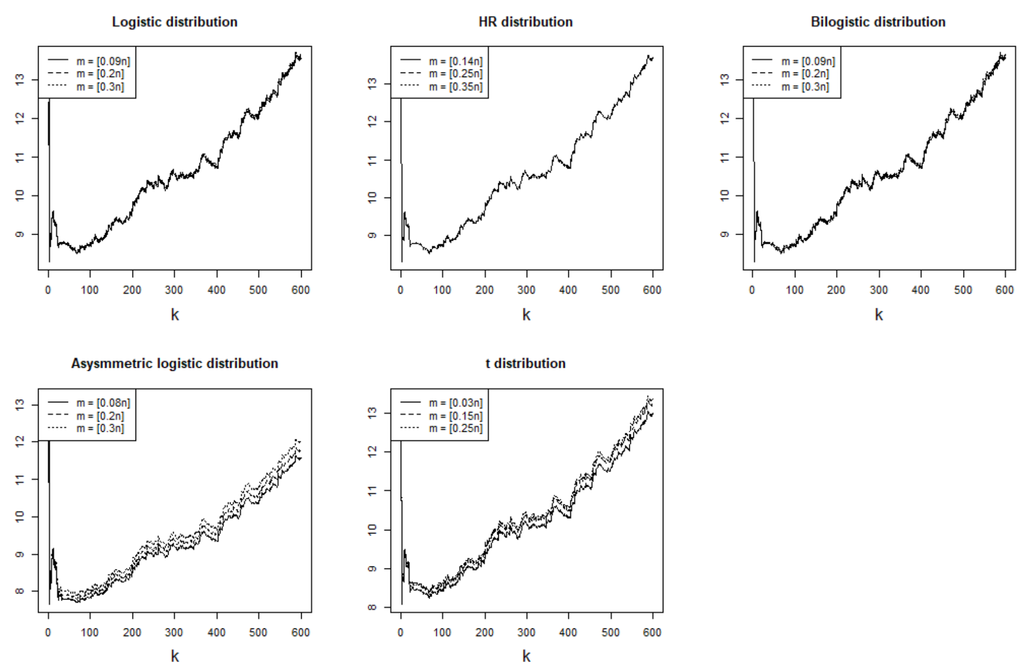
(a) BK-DJUSFN



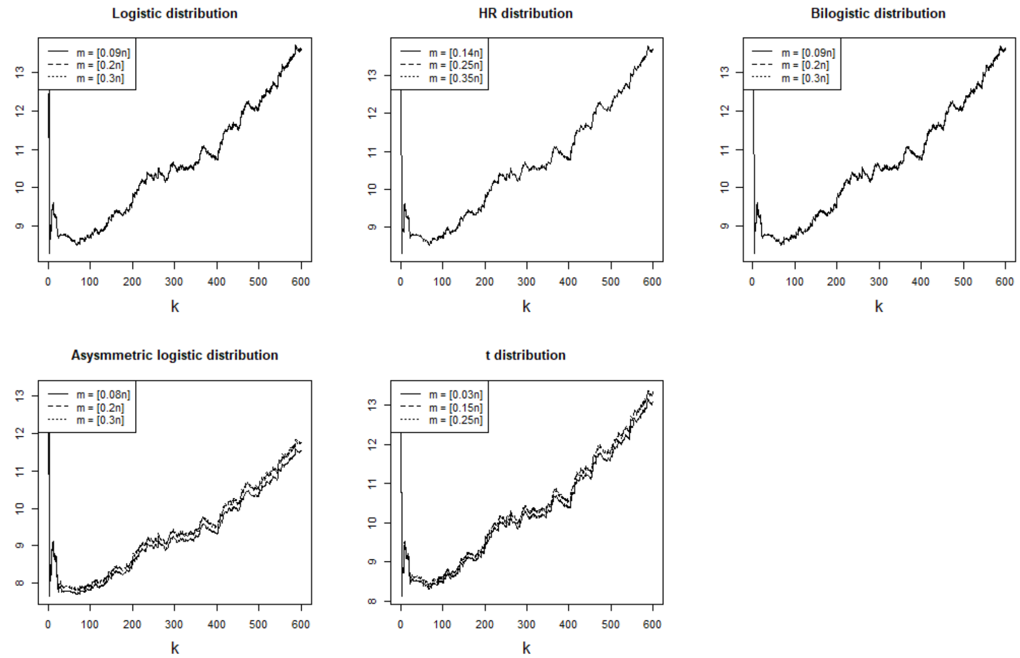
(b) AFL-DJUSFN



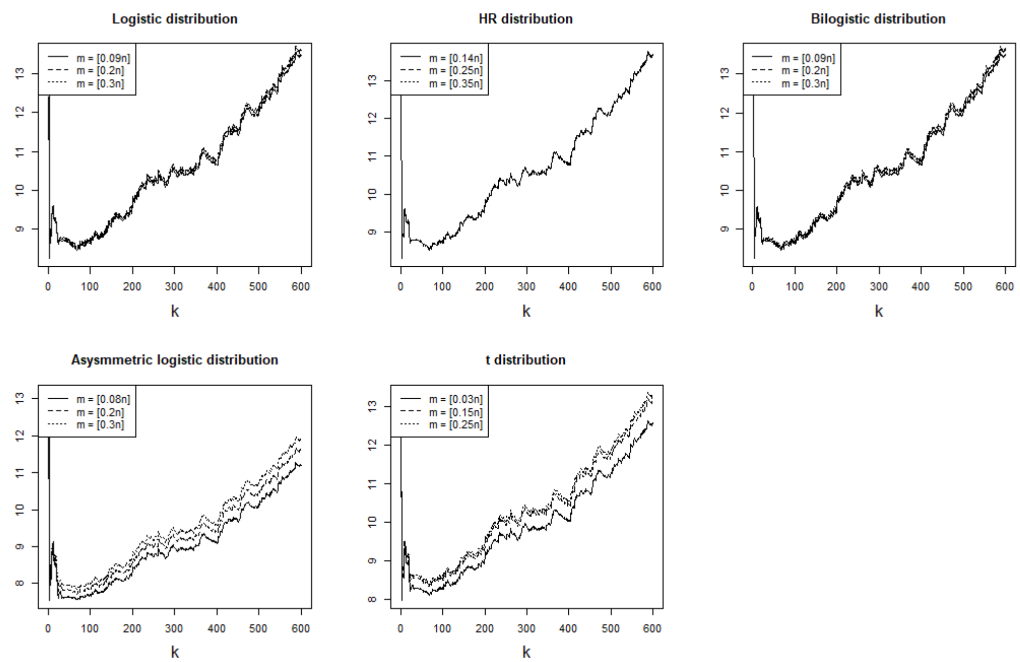
(c) ALL-DJUSFN



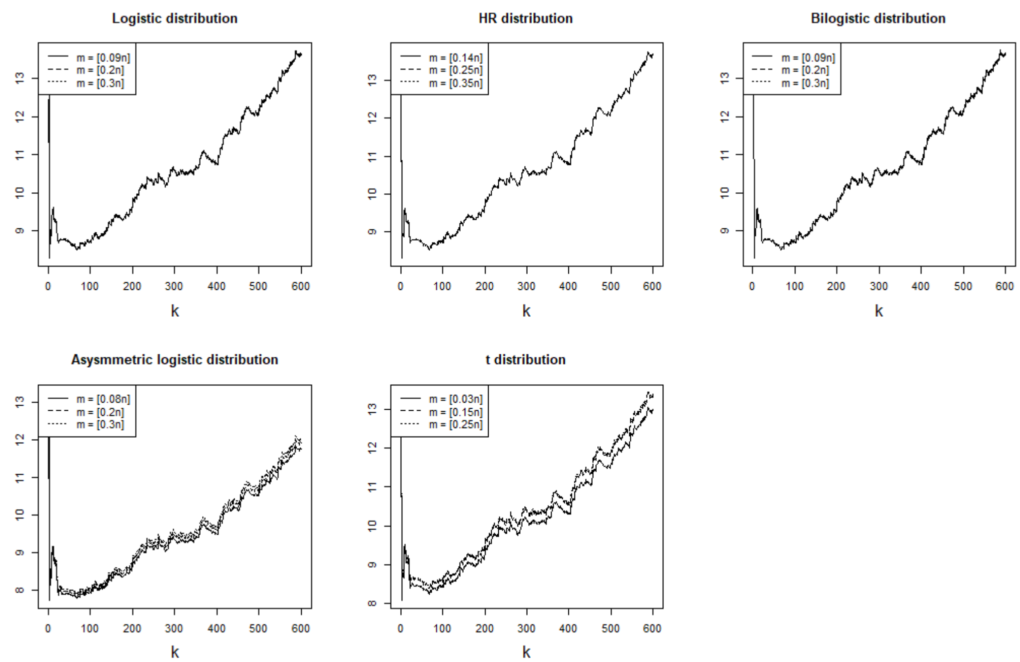
(d) AXP-DJUSFN



(e) BEN-DJUSFN

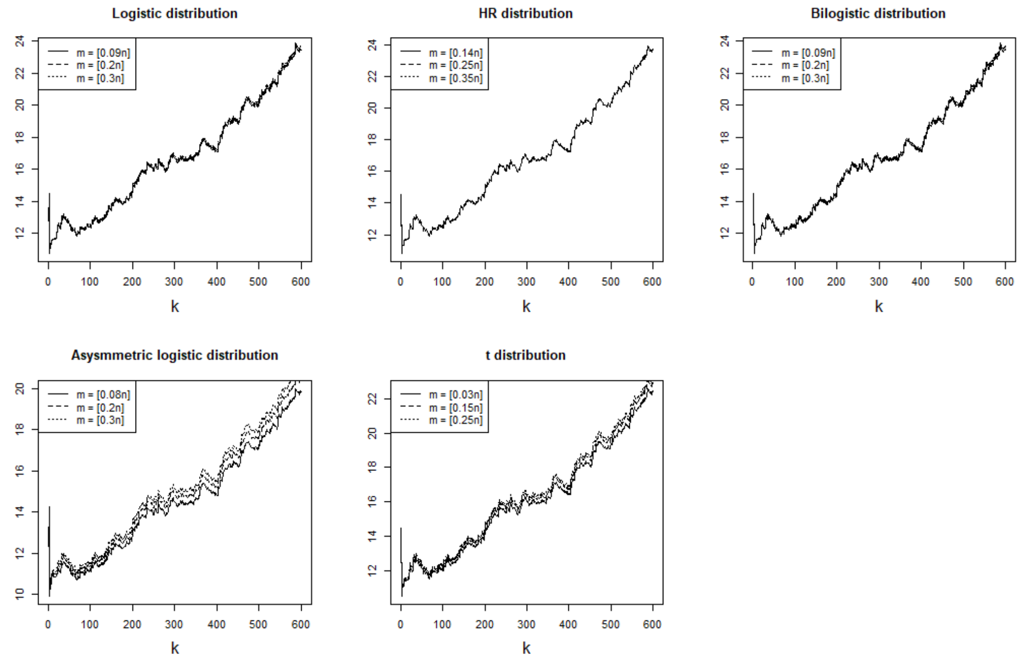


(f) GS-DJUSFN

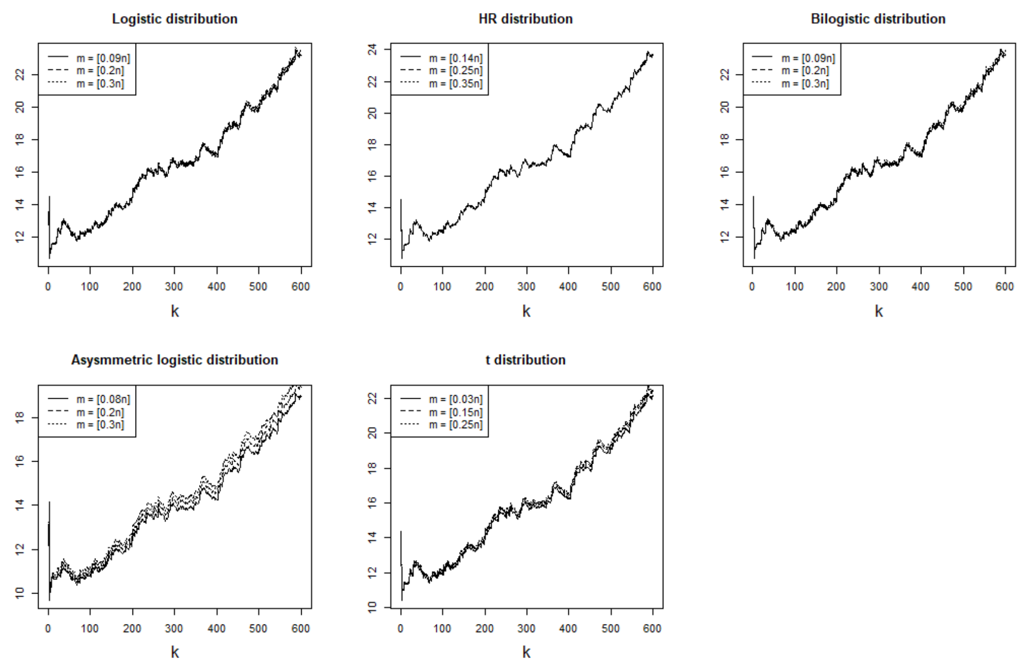


(g) TROW-DJUSFN

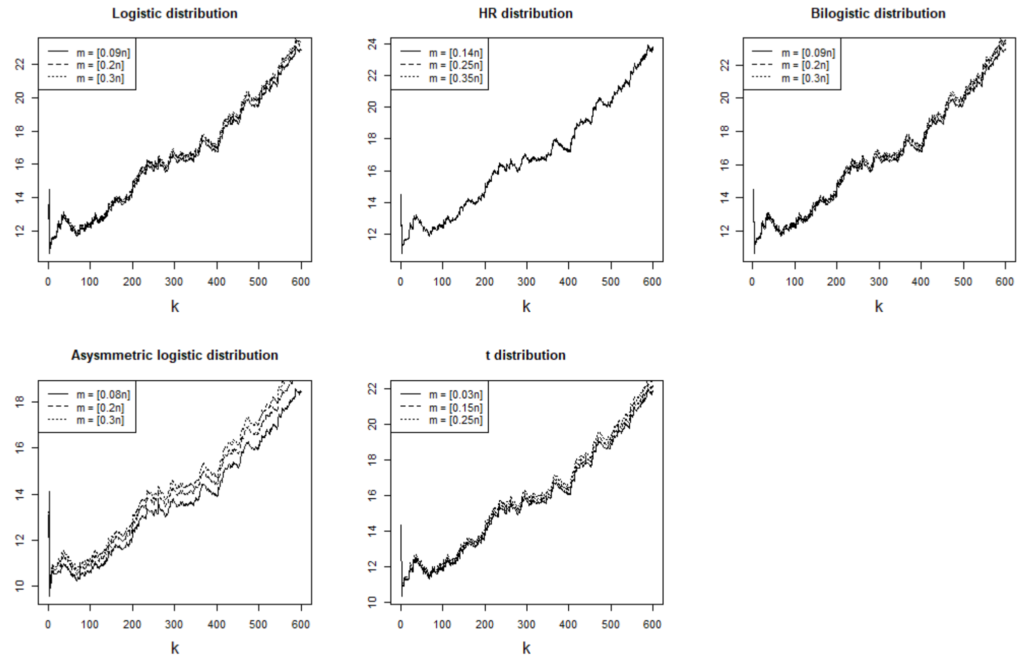
Figure B.4: Estimates of CoVaR as a function of k with raw data at level $\mathrm{p}_n = (0.05, 0.05)$ for different values of m.



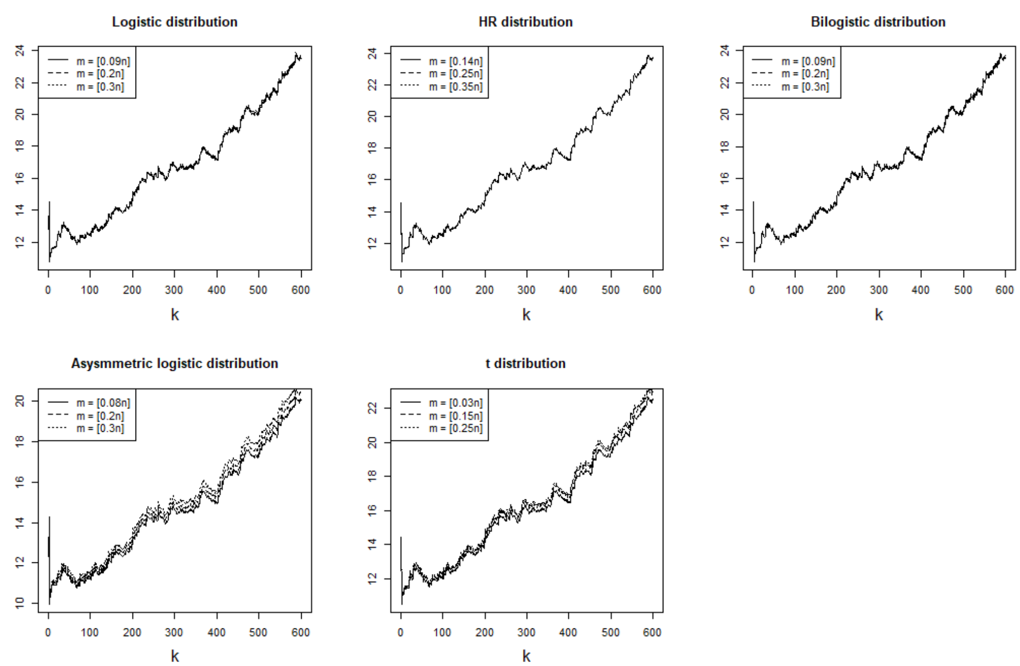
(a) BK-DJUSFN



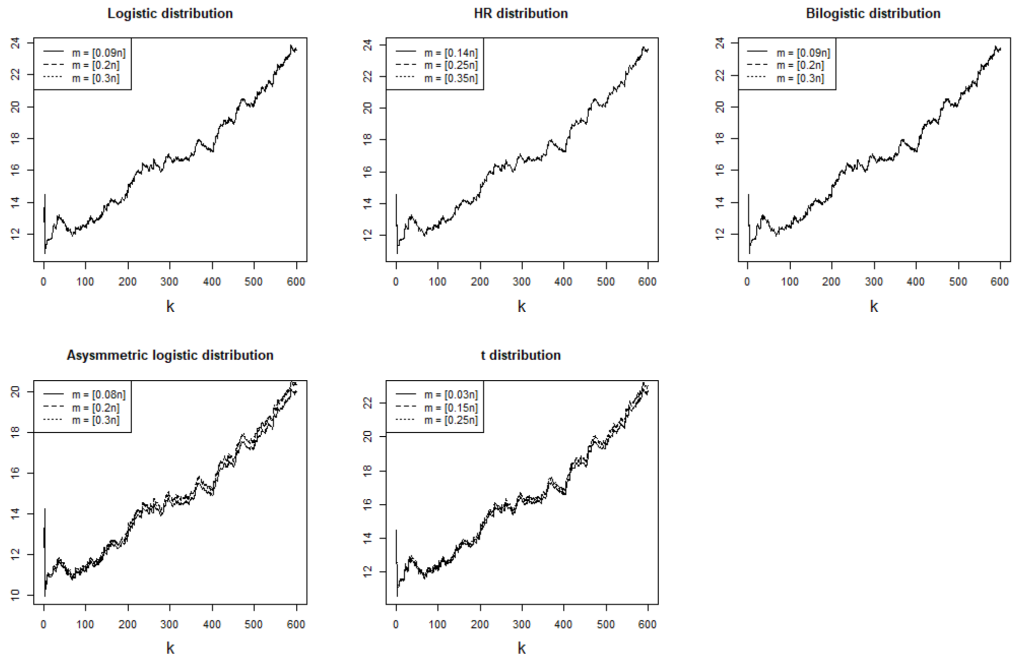
(b) AFL-DJUSFN



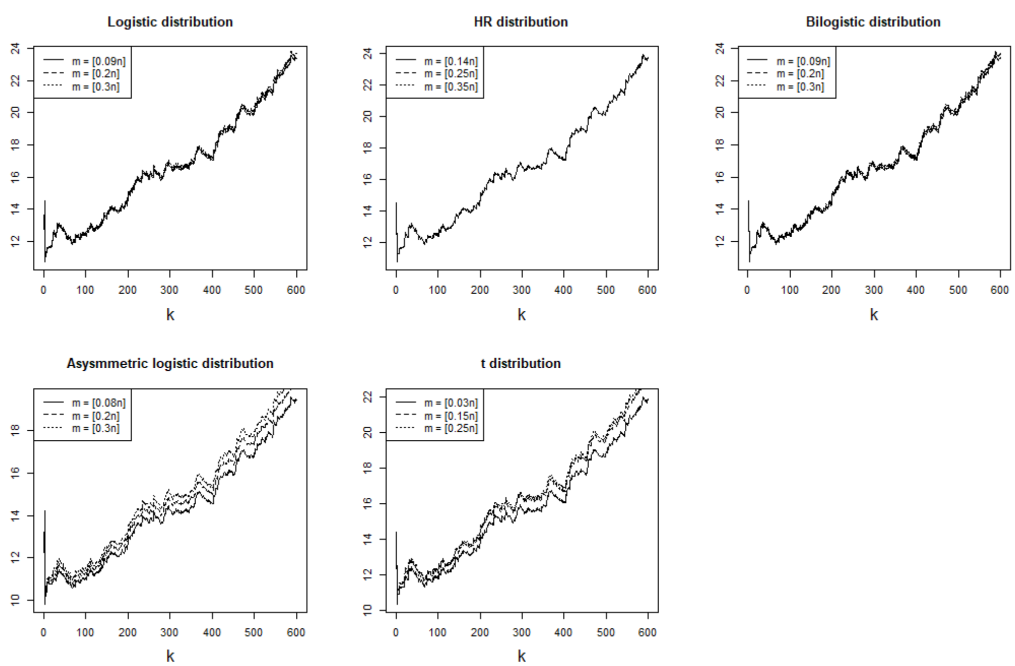
(c) ALL-DJUSFN



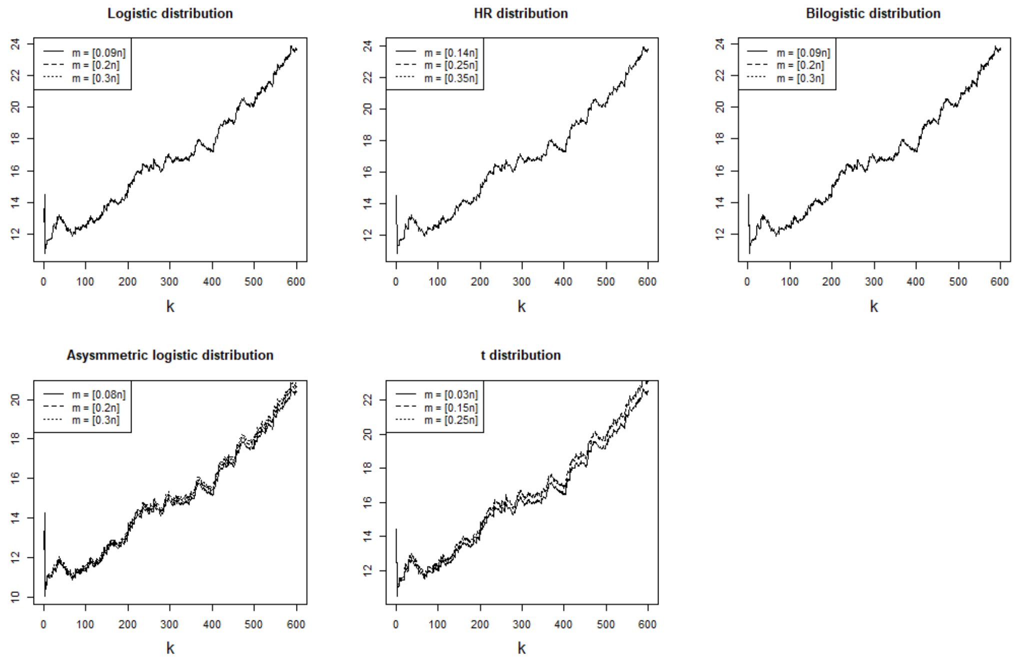
(d) AXP-DJUSFN



(e) BEN-DJUSFN

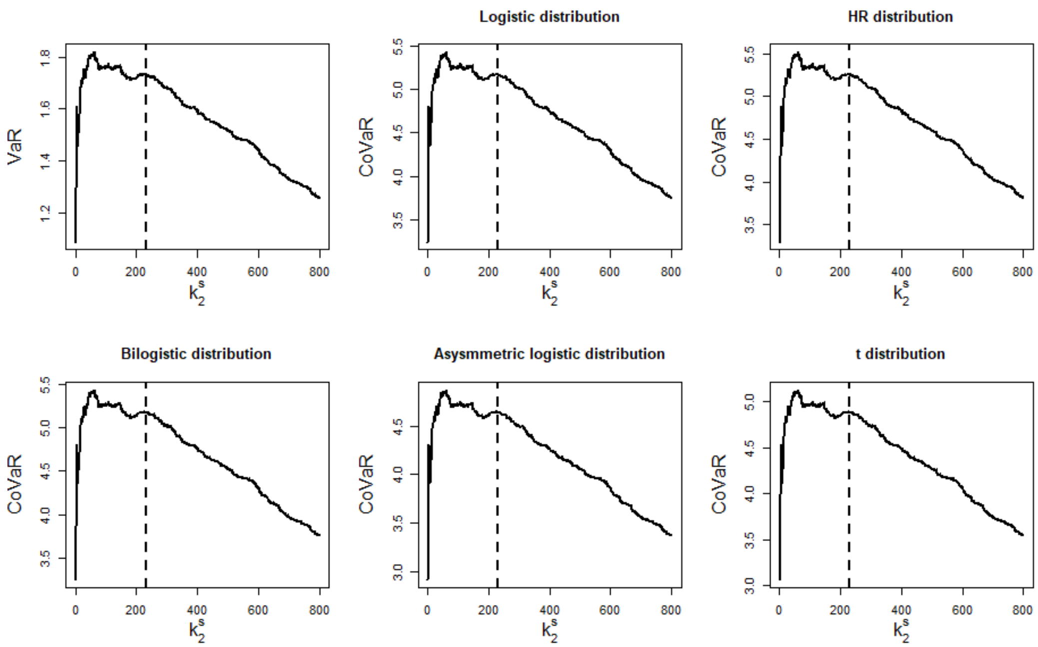


(f) GS-DJUSFN

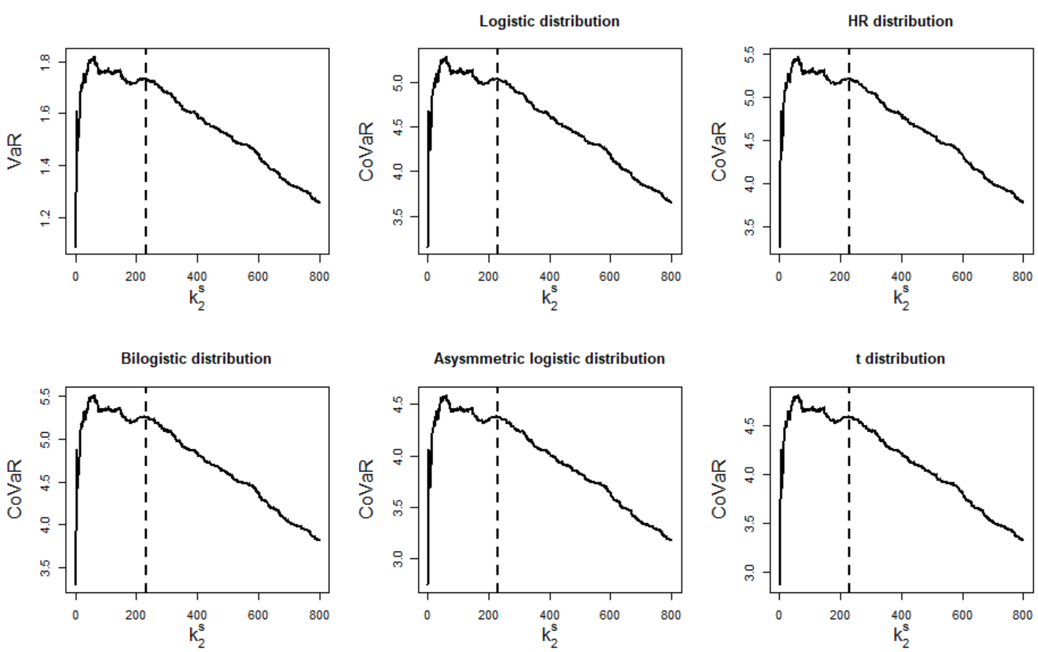


(g) TROW-DJUSFN

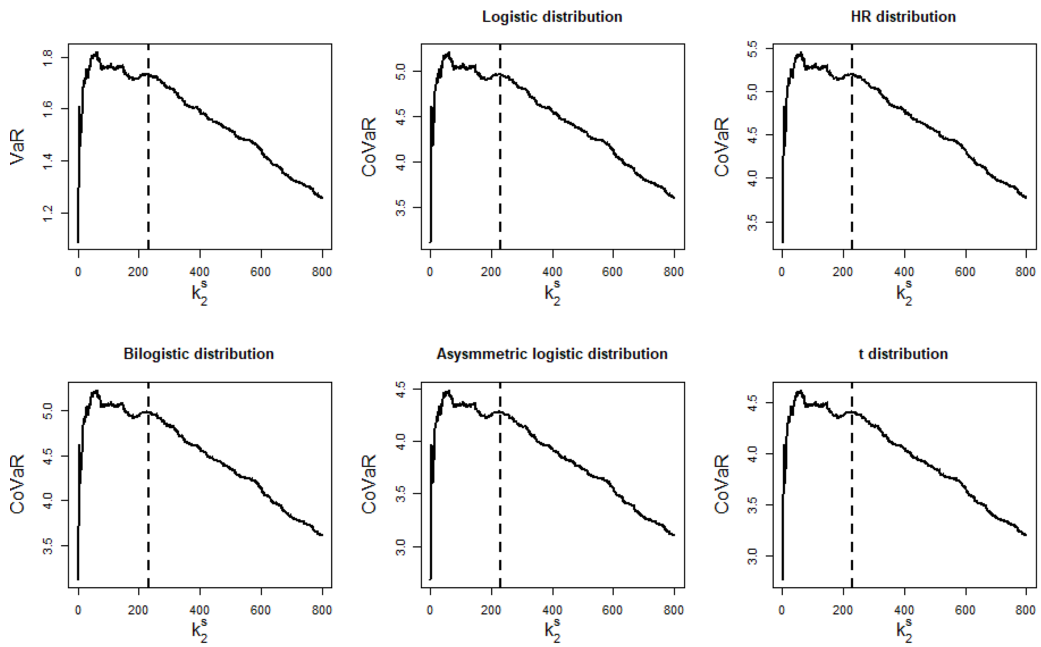
Figure B.5: Estimates of CoVaR as a function of k with raw data at level $\mathrm{p}_n = (0{:}02; 0{:}05)$ for different values of m.



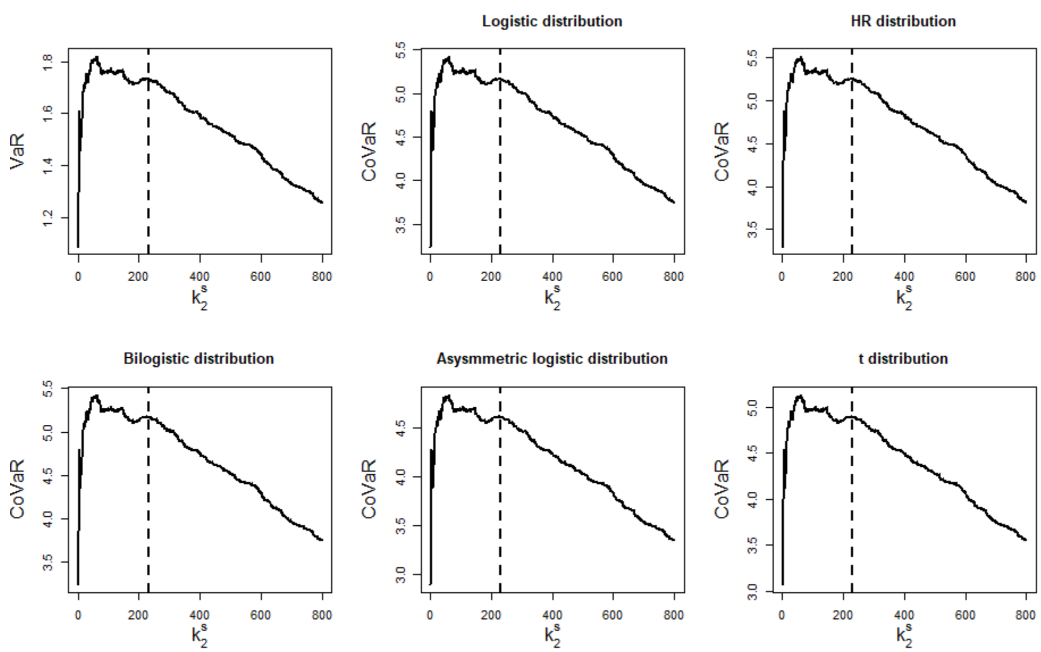
(a) BK-DJUSFN



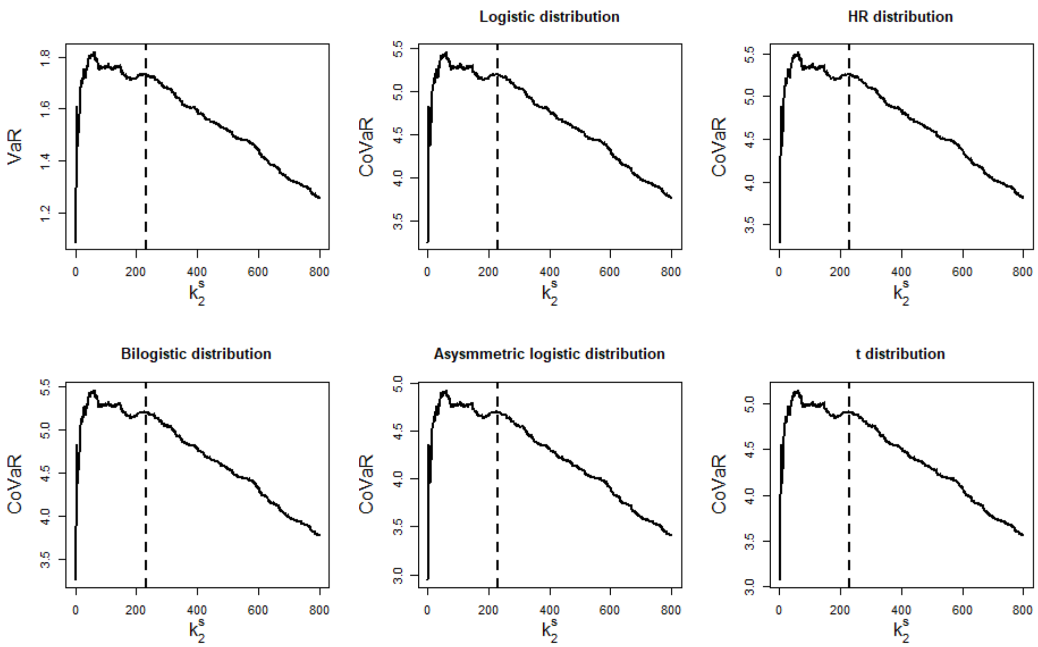
(b) AFL-DJUSFN



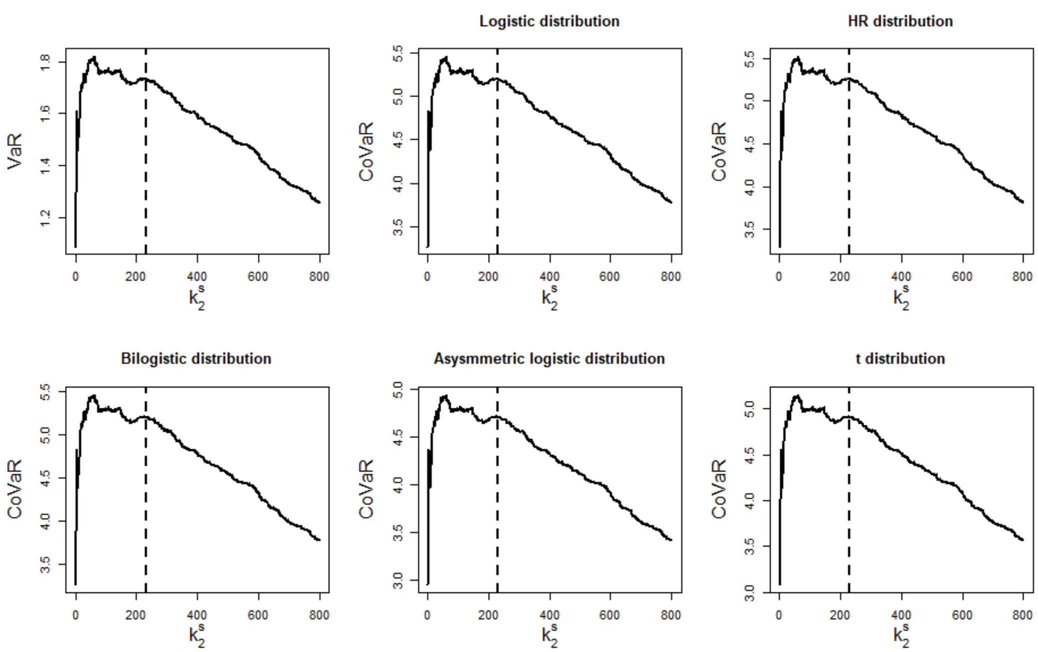
(c) ALL-DJUSFN



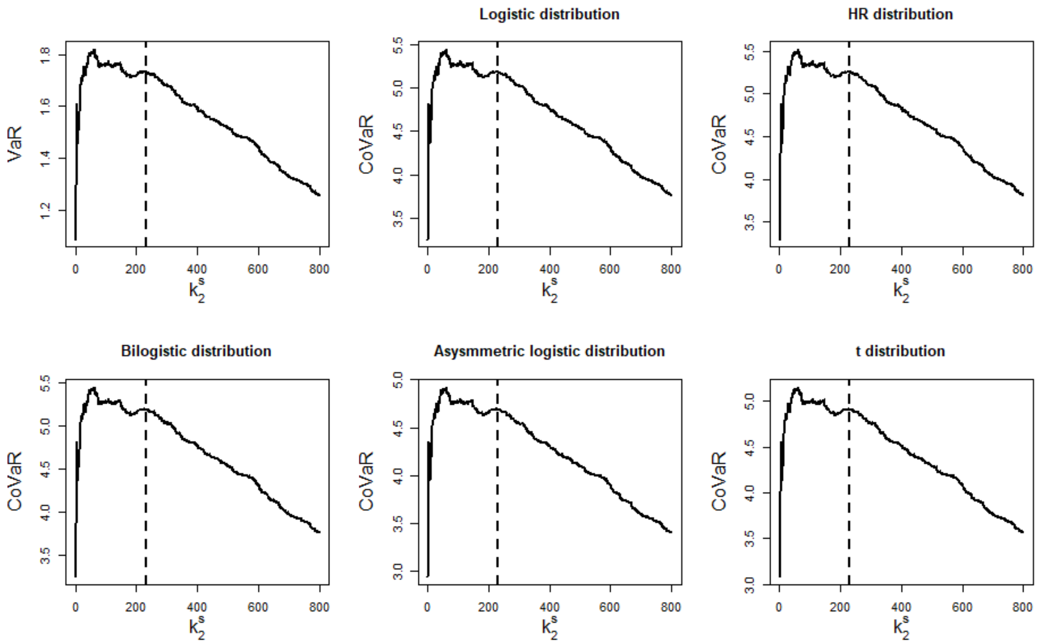
(d) AXP-DJUSFN



(e) BEN-DJUSFN



(f) GS-DJUSFN



(g) TROW-DJUSFN

Figure B.6: Estimates of CoVaR as a function of k_2^s with realized residuals at level $\mathbf{p}_n = (0.02, 0.05)$. The vertical line represents the $k_2^s = k_1^s = 230$.