

ABSTRACT

WANG, CHUN-JU. Risk Measures and Capital Allocation. (Under the direction of Peter Bloomfield.)

This research first gives a review of risk measures and risk capital allocation, along with the important property of coherency, and the relationships between different coherent risk measures. Secondly, relative accuracy measures are used as model-based criteria to study whether or not bias adjustment by various bootstrap techniques could improve estimates of the expected shortfall (ES) as a risk measure. Thirdly, different tests for backtesting Value-at-Risk (VaR) and ES are investigated as data-based criteria of evaluating risk models. Fourthly, multivariate framework is developed for estimating (conditional) ES and ES risk contributions (ESC), as a principle of capital allocation. Finally, an empirical study of estimating ES and ESC with backtesting is carried out for historical data from Russell Indices.

© Copyright 2010 by Chun-Ju Wang

All Rights Reserved

Risk Measures and Capital Allocation

by
Chun-Ju Wang

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Statistics

Raleigh, North Carolina

2010

APPROVED BY:

Sastry G. Pantula

David A. Dickey

Min Kang

Peter Bloomfield
Chair of Advisory Committee

DEDICATION

*To my parents in Taiwan:
Chuan-Wen Wang and Chiao-Lan Hung*

BIOGRAPHY

Chun-Ju Wang was born in Shulin City, Taipei County, Taiwan, R.O.C. on February 14th, 1977. In 1999, she obtained her B.S. degree in Mathematics from Fu Jen Catholic University, Taipei, Taiwan. After she received her M.B.A. degree concentrating on Statistics from National Taipei University, Taipei, Taiwan, in 2002, she enrolled in Institute of Statistical Science, Academia Sinica as a one-year research student to received advanced training of statistical science. She then decided to prepare for pursuing her Ph.D. abroad and worked as a research assistant there from August 2003 to July 2004. Chun-Ju enrolled in the doctorate program from Department of Statistics at North Carolina State University in Raleigh, NC, USA. Her research is inspired by and under the direction of Dr. Peter Bloomfield. Her research interests are data mining, time series, and statistical models for finance, especially risk measures and capital allocation. From August 2006 to May 2008, she worked as a research assistant on projects in the Office of Assessment, Division of Undergraduate Academic Programs at North Carolina State University. In September 2008, she provided statistical analyses on the project funded by USDA Risk Avoidance and Mitigation for Department of Plant Pathology at North Carolina State University under the supervision of Dr. Consuelo Arellano from Department of Statistics. She will be completing her Ph.D. program as a Spring 2010 graduate.

ACKNOWLEDGEMENTS

My profound appreciation goes to my adviser, Dr. Peter Bloomfield, who has been guiding my research with a lot of brilliant ideas, remarkable advice, and generous help. He has been always encouraging me with his gentle, kind, patient, and optimistic attitude not only for my research but also for my life. I am sincerely grateful that he would help me with going over the detail of my thesis even word by word so that I have been learning much writing skills from him. He has been very nice and definitely professional to work with so that I have been so lucky and very thankful to have him as my adviser.

I would like to thank Dr. Sastry G. Pantula, Dr. David A. Dickey, and Dr. Min Kang for serving as committee members of mine. Their insightful suggestions and comments have been making my work more complete, and I have learned a lot from their excellent lectures as well. I am thankful to Dr. Consuelo Arellano for giving me an opportunity of working with her on the project for Department of Plant Pathology at North Carolina State University. I would also like to thank Dr. Pam Arroway, Ms. Alison McCoy, and Mr. Adrian Blue for their hard work of assisting students in our department. My thanks also go to Dr. Huimei Liu, who advised my master thesis, Dr. Kang C. Jea, and Dr. Nan-Ping Yang, who are both from Fu Jen Catholic University, for their great advice and support on my career. I would like to thank Ms. Hsui-Chuan Wang, who is my High School English Teacher, because her excellent teaching inspired my passion for learning the second language, English.

More thanks go to my family for their unconditional love and consistent support, which have been helping me go through hard times as well as good times. I am really grateful to be the first one of their three children. My special thanks go to Dr. Li Li, one of my best friends, who has been standing by me and helping me get over difficult times since we met in North Carolina State University. I would also like to thank my friends, especially Shih-Wan Kang, Ying-Erh Chen, Tai-Ying Liu, Min Zhang, Kejun Zhang, Wenfo and Wolfgang Witz, Heather Dellinger, and Dr. Andrew Dellinger for their help and support. I would like to thank Taiwanese Student Association at NCSU, whose members have helped me a lot while I first came to USA in 2004.

TABLE OF CONTENTS

List of Tables	viii
List of Figures	x
Chapter 1 Introduction	1
1.1 General Introduction and Motivation	1
1.2 Overview of Risk Measures and Capital Allocation	2
1.2.1 Risk Measures	2
1.2.2 Capital Allocation	10
1.2.3 Summary of Contributions	15
1.3 Overview of Thesis	15
Chapter 2 Bias Adjustment Using Bootstrap Techniques for Estimating ES .	16
2.1 Introduction	16
2.2 Bootstrap Techniques	18
2.2.1 The Ordinary Bootstrap	19
2.2.2 The Exact Bootstrap of L -statistics	21
2.2.3 Blockwise Bootstrap for Dependent Data	22
2.3 Bias Adjustment for Simple ES Estimates	23
2.3.1 For HS Estimates	24
2.3.2 For GPD Estimates	28
2.3.3 For MLE From Elliptical Distributions	31
2.4 Bias Adjustment for Conditional ES Estimates of Time Series	39
2.4.1 Conditional ES of ARMA Processes	40
2.4.2 Conditional ES of GARCH Processes	41
2.4.3 Conditional ES of ARMA-GARCH Processes	46
2.5 Summary of Contributions	47
Chapter 3 Evaluation of the Performance of Risk Models	49
3.1 Model-based Criteria	49
3.1.1 Existent Large Sample Properties of HS Estimates for ES	50
3.1.2 Relative Accuracy Measures	52
3.1.3 Simulation Study	52
3.2 Data-based Criteria (Backtesting)	75
3.2.1 Tests of Unconditional Coverage	76

3.2.2	Tests of Independence	77
3.2.3	Joint Tests of Coverage and Independence	82
3.2.4	The Bootstrap Test of Zero-mean Behaviour	84
3.2.5	Simulation Study of Backtesting	85
3.3	Summary of Contributions	95
Chapter 4 Multivariate Framework for Estimating ES and ESC		96
4.1	Introduction	96
4.2	Methods Without a Volatility Filter	96
4.2.1	Multivariate Version of Historical Simulation	96
4.2.2	Multivariate Elliptical Distributions	97
4.3	Conditional ES and ESC for MGARCH Processes	99
4.3.1	Conditional ES and ESC for MGARCH-Gaussian Process	100
4.3.2	Conditional ES and ESC for MGARCH- t Processes	101
4.3.3	Estimation Strategies	102
4.4	Backtesting of ESC	103
4.5	Summary of Contributions	103
Chapter 5 Empirical Study		104
5.1	Backtesting VaR and ES	114
5.1.1	The Bootstrap Test of Zero-mean Behaviour	114
5.1.2	Joint Tests of Coverage and Independence	114
5.1.3	Test of Independence	115
5.1.4	The Unconditional Coverage Test	115
5.2	Backtesting ESC	116
5.3	Conclusions	116
Chapter 6 Conclusion and Future Topics		117
6.1	Conclusion	117
6.2	Future Topics	118
6.2.1	Sequential Estimation of ESC for Higher Dimensional MGARCH	118
6.2.2	Improvement of GPD Method With Optimal Threshold	118
6.2.3	Other Estimation Methods for GARCH Models	118
References		119
Appendices		124
Appendix A	Proof of Proposition 1.9	125

Appendix B	Proof of Proposition 1.10	127
B.1	To show (B.1) holds if either (B.2) or (B.3) is true	127
B.2	To show either (B.2) or (B.3) is true, if (B.1) holds	129
Appendix C	Proof of Theorem 2.18	131
Appendix D	Proof of Theorem 4.2	132
Appendix E	Proof of Proposition 4.3	134
Appendix F	Proof of Theorem 4.5	136
Appendix G	Proof of Proposition 3.2	139
Appendix H	Proof of Proposition 3.3	142

LIST OF TABLES

Table 3.1	Relative MSE of Original ES Estimates (Sample Size = 50) in %	55
Table 3.2	Relative Accuracy Measures of Simple ES Estimates (Sample Size = 50)	56
Table 3.3	Relative Accuracy Measures of ES Estimates for GARCH-N (Sample Size = 50)	57
Table 3.4	Relative Accuracy Measures of ES Estimates for EWMA (Sample Size = 50)	58
Table 3.5	Relative Accuracy Measures of ES Estimates for GARCH-t (Sample Size = 50)	59
Table 3.6	Relative MSE of Original ES Estimates (Sample Size = 250)	60
Table 3.7	Relative Accuracy Measures of Simple ES Estimates (Sample Size = 250)	61
Table 3.8	Relative Accuracy Measures of ES Estimates for GARCH-N (Sample Size = 250)	62
Table 3.9	Relative Accuracy Measures of ES Estimates for EWMA (Sample Size = 250)	63
Table 3.10	Relative Accuracy Measures of ES Estimates for GARCH-t (Sample Size = 250)	64
Table 3.11	Relative MSE of Original ES Estimates (Sample Size = 500)	65
Table 3.12	Relative Accuracy Measures of Simple ES Estimates (Sample Size = 500)	66
Table 3.13	Relative Accuracy Measures of ES Estimates for GARCH-N (Sample Size = 500)	67
Table 3.14	Relative Accuracy Measures of ES Estimates for EWMA (Sample Size = 500)	68
Table 3.15	Relative Accuracy Measures of ES Estimates for GARCH-t (Sample Size = 500)	69
Table 3.16	Relative MSE of Original ES Estimates (Sample Size = 1000)	70
Table 3.17	Relative Accuracy Measures of Simple ES Estimates (Sample Size = 1000)	71
Table 3.18	Relative Accuracy Measures of ES Estimates for GARCH-N (Sample Size = 1000)	72
Table 3.19	Relative Accuracy Measures of ES Estimates for EWMA (Sample Size=1000)	73
Table 3.20	Relative Accuracy Measures of ES Estimates for GARCH-t (Sample Size = 1000)	74
Table 3.21	Contingency Table for Pearson's χ^2 Test of Independence	79
Table 3.22	Sizes and Powers for Tests of Unconditional Coverage (Pseudo Estimate of VaR)	87

Table 3.23	Sizes and Powers for Tests of Unconditional Coverage (GM Estimate of VaR)	88
Table 3.24	Sizes for Tests of Independence (Pseudo Estimate of VaR)	90
Table 3.25	Powers for Tests of Independence (GM Estimate of VaR)	91
Table 3.26	Sizes and Powers for Joint Tests of Coverage and Independence (Pseudo Estimate of VaR)	93
Table 3.27	Powers for Joint Tests of Coverage and Independence (GM Estimate of VaR)	94
Table 5.1	<i>P</i> -values of the Bootstrap Test of Zero-mean Behaviour for ES	114
Table 5.2	<i>P</i> -values of the Duration-based Test of Coverage and Independence	114
Table 5.3	<i>P</i> -values of Pearson's χ^2 Test of Coverage and Independence	115
Table 5.4	<i>P</i> -values of the Duration-based Test of Independence	115
Table 5.5	<i>P</i> -values of Pearson's χ^2 Test of Independence	115
Table 5.6	<i>P</i> -values of the Unconditional Coverage Test by Binomial Approach	116
Table 5.7	<i>P</i> -values of the Bootstrap Test of Zero-mean Behaviour for ESC	116

LIST OF FIGURES

Figure 2.1	Real World vs. Bootstrap World	18
Figure 3.1	Sizes and Powers for Tests of Unconditional Coverage (Pseudo Estimate of VaR)	87
Figure 3.2	Sizes and Powers for Tests of Unconditional Coverage (GM Estimate of VaR)	88
Figure 3.3	Sizes for Tests of Independence (Pseudo Estimate of VaR)	90
Figure 3.4	Powers for Tests of Independence (GM Estimate of VaR)	91
Figure 3.5	Sizes and Powers for Joint Tests of Coverage and Independence (Pseudo Estimate of VaR)	93
Figure 3.6	Powers for Joint Tests of Coverage and Independence (GM Estimate of VaR)	94
Figure 5.1	Structure of The Empirical Study	105
Figure 5.2	Plot of Daily Values	107
Figure 5.3	Plot of Daily Losses	108
Figure 5.4	Plot of Russell 1000 Index and the Predicted ES	109
Figure 5.5	Plot of Russell 2000 Index and the Predicted ES	110
Figure 5.6	Plot of Aggregate Index and the Predicted ES	111
Figure 5.7	Plot of Russell 1000 Index and the Predicted ESC	112
Figure 5.8	Plot of Russell 2000 Index and the Predicted ESC	113

Chapter 1

Introduction

1.1 General Introduction and Motivation

The development of financial risk management has been of active interest to regulators, researchers, specialists associated with the field, and financial institutions themselves. Much of it originates from the need to quantify, control, and report the risks that a financial institution faces, in the sense of the essential level of uncertainty about the future, comprised of market, credit, liquidity, operational, and systemic risks. The essence is that the amount of capital a financial institution needs should be determined as a buffer to absorb unexpected losses from those risks that could otherwise lead the financial institution to insolvency. Such capital is then called risk capital, and the determination of it has become prominent, since a financial institution is required to hold appropriate risk capital for keeping the possibility of insolvency acceptably low; meanwhile it must still be competitive in the capital market. At the international level, regulators such as the International Association of Insurance Supervisors and the Basel Committee on Banking Supervision (BCBS), to name but two, make many efforts to develop a set of rules for the world's accounting, banking, or insurance systems in order to provide a safer trading environment.

Meanwhile, the concern for capital allocation arises from how the total risk capital for a multi-unit financial institution (or a combination of different portfolios) can be allocated back to each business unit (or each portfolio) within this financial institution, once the total risk capital is calculated based on a specific risk measure. In other words, each business unit requires its own capital to hold, and could benefit from the diversification effect as a reward instead of punishment, which means that the allocated capital for each business unit is less than its stand-alone risk capital. An allocation principle is a function that maps allocation problems associated with a specified risk measure into a unique allocation, and allows coping with the risk capital problem in a consistent way. Once a proper risk measure is determined to use, according to the

nature of the application and the amount of information available, the problem becomes how to estimate the measure of risk and the allocated capital based on the specified risk measure. In order to choose a more suitable estimator, evaluation of the performance of estimates calculated by different methods is investigated as well. For comprehensive reviews of risk measurement, please refer to McNeil et al. (2005) and Dowd (2005b). The former, by McNeil et al. (2005), provides core concepts, techniques and tools of quantitative modeling issues for financial risk management. It also covers the issue of capital allocation. The latter, Dowd (2005b), addresses how to measure the market risk, including estimation of VaR and ES.

This research will first review risk measures and risk capital allocation, along with the important property of coherency, and the relationships between different coherent risk measures. Secondly, relative accuracy measures will be used as model-based criteria to study whether or not bias adjustment by various bootstrap techniques could improve estimates of the expected shortfall (ES) as a risk measure. Thirdly, tests for backtesting ES will be investigated as data-based criteria of evaluating ES estimates. Fourthly, based on multivariate framework, estimation of (conditional) ES and ES risk contributions (ESC), as a principle of capital allocation, will be studied. Finally, an empirical study of estimating ES and ESC with backtesting will be carried out for historical data from Russell Indices.

1.2 Overview of Risk Measures and Capital Allocation

1.2.1 Risk Measures

Interest has been increasing in the subject of how to describe the many and various risks a financial institution faces or a portfolio is exposed to. The measurement of risk gives us a quantitative tool to tell how risky the business of a financial institution (or a portfolio) is. A risk measure, quantitative measurement of risk, is a number which is designed to evaluate the magnitude of risk and capture the primary properties of the risk.

Coherent Risk Measures

Proposed risk measures are judged by whether they satisfy specific desirable properties of risk measures, collected as the axioms of coherence, a concept developed by Artzner et al. (1999).

Definition 1.1. (Axioms of Coherent Risk Measure; Artzner et al. (1999)). *Suppose that \mathcal{G} is the set of all real-valued measurable functions on the sample space Ω of outcomes. A risk measure $\rho : \mathcal{G} \rightarrow \mathbb{R} \cup \{+\infty\}$ is said to be coherent if it satisfies the following four axioms:*

1. **Subadditivity:** For all $X \in \mathcal{G}, Y \in \mathcal{G}$,

$$\rho(X + Y) \leq \rho(X) + \rho(Y)$$

2. **Monotonicity:** For all $X \in \mathcal{G}$ and $Y \in \mathcal{G}$ such that $X \leq Y$,

$$\rho(X) \geq \rho(Y)$$

3. **Positive Homogeneity:** For any number $\lambda \geq 0$ and $X \in \mathcal{G}$,

$$\rho(\lambda X) = \lambda \rho(X)$$

4. **Translation Invariance:** For any constant $k \in \mathbb{R}$,

$$\rho(X + kr) = \rho(X) - k$$

where r is the price of a riskless instrument at the end of the holding period with price 1 today.

Subadditivity is necessary to incorporate the property that diversification of portfolios has a positive impact, which means that combining two portfolios does not create extra risk. Monotonicity says that if a portfolio Y is always worth more than X , then Y should not be riskier than X . Positive homogeneity is a limit case of subadditivity, meaning what happens when no diversification occurs. It also has the interpretation that the risk measure is independent of scale changes (e.g. currency changes) in the unit in which the risk is measured. However, it is debatable, since the risk in the position may become unacceptable in case r is large enough, although it is innocuous for moderate values of r . Translation invariance is a natural requirement and also known as *Risk-free condition*. It represents that adding a sure amount of money decreases the measurement of risk by the same amount. Notice that no probability measure on Ω has been involved.

Note that no probability spaces are involved in the axioms of coherent risk measures (Definition 1.1). (Artzner et al., 1999) not only introduced the definition of coherent risk measure but also generalized the computation of SPAN margin system (SPAN, 1995), developed by the Chicago Merchantile Exchange, to scenario-based risk measures defined below. It provided a general way to construct a coherent risk measure by using a family of probability measures and quantifying risks as the worst-case expected loss.

The definition of scenario-based risk measures (SBRM) is given as follows.

Definition 1.2. (Scenario-based Risk Measures). *The scenario-based risk measure defined by a non-empty set \mathcal{P} of probability measures or generalized scenarios on the space Ω for a random variable X is the function $\rho_{\mathcal{P}}$ on \mathcal{G} , the set of all real-valued functions on Ω , defined by*

$$\rho_{\mathcal{P}}(X) = \sup_{P \in \mathcal{P}} E_P(-X). \tag{1.1}$$

Initially there is no particular probability measure on Ω . Artzner et al. (1999) discussed the relationship between coherence and scenario-based risk measures, along with the following two propositions.

Proposition 1.3. (Artzner et al. (1999)). *Given the non-empty set \mathcal{P} of probability measures, or generalized scenarios, on the set Ω of states of the world, the risk measure $\rho_{\mathcal{P}}$ of Definition 1.2 is a coherent risk measure. It satisfies the relevant axioms if and only if the union of the supports of the probabilities $P \in \mathcal{P}$ is equal to the set Ω .*

Proposition 1.4. (Artzner et al. (1999)). *A risk measure ρ is coherent if and only if there exists a family \mathcal{P} of probability measures on the set of states of nature, such that*

$$\rho_{\mathcal{P}}(X) = \sup_{P \in \mathcal{P}} E_P(-X).$$

Proofs of Propositions 1.3 and 1.4 are provided in Artzner et al. (1999) as well.

Furthermore, Delbaen (2002) extended Definition 1.1 to coherent risk measures on general probability spaces by characterising closed convex sets of probability measures which satisfy the property that every random variable is integrable for at least one probability measure in this set. Please refer to his context for more reference and details.

Value at Risk

The notion of Value-at-Risk or VaR was introduced by Morgan (1995) as a risk management tool, and made a major breakthrough in the evolution of modern risk measures. VaR has been extensively applied to determine the required risk capital which a financial institution should hold to cover its worst case loss. VaR has been employed by regulatory bodies such as the Basel Committee on Bank Supervision (BCBS) which published *the 1996 Amendment* (BCBS, 1996, Amendment to the capital accord to incorporate market risks). VaR is extended straightforwardly from *possible maximum loss* as *possible maximum loss* which is not exceeded within a given confidence level over a given holding period (McNeil et al., 2005). Suppose that X denotes the profit and loss (P&L) of either the business of the underlying financial institution or the specified portfolio, which takes the risk. In other words, X is the change in the value of the business of the underlying financial institution or of the specified portfolio between the date of calculation and a future date. Given the probability distribution of X , the VaR at the confidence level α is the smallest number x such that the probability that the loss will exceed x within a fixed time horizon, is no larger than $1 - \alpha$. The formal definition of VaR is given in Definition 1.5.

Definition 1.5. (Value at Risk; McNeil et al. (2005)). *VaR at the confidence level α is defined as*

$$\text{VaR}_\alpha(X) = \inf\{x \in \mathbb{R} : P(-X > x) \leq 1 - \alpha\}.$$

See McNeil et al. (2005) or Acerbi and Tasche (2002) for more discussion. This definition of VaR presumes that $X(\omega) \in \mathbb{R}$ for all $\omega \in \Omega$ where Ω is the sample space of outcomes. Furthermore,

$$\text{VaR}_\alpha(X) = q_\alpha(-X)$$

(Tasche, 2002b) where $q_\alpha(Y)$ denotes the lower α -th quantile of Y ¹. In the case where X is continuous with a strictly positive density function, the cumulative distribution function (cdf) of X is a continuous and strictly monotonic distribution function and $\text{VaR}(X)$ is the unique solution for x of the equation

$$P(-X \leq x) = \alpha.$$

Alternatively, VaR can be interpreted as the minimum loss incurred in the $(1 - \alpha)\%$ worst cases of the P&L. In practice, the time horizon used is usually the one day or ten days period and the value for level α is quite high and close to 1, typically 95% or 99%. Despite the widespread use of VaR as a standard risk measure among practitioners in the financial market, it has many deficiencies and inconsistencies when applied in real life applications. Particularly, it is heavily criticized for failing to satisfy the subadditivity axiom of coherent risk measures defined in Definition 1.1 and therefore fails to encourage diversification, while clearly satisfying the other three axioms: monotonicity, positive homogeneity, and translation invariance (Artzner et al., 1999, i.c.). Comprehensive discussions on VaR can be found in Jorion (2001) and Holton (2003).

Expected Shortfall

An alternative to VaR is expected shortfall, of which definition is given in 1.6. Discussion on VaR and Tail VaR as risk measures can be found in Overbeck (2000). The most popular and widely-accepted risk measure, expected shortfall, has attracted a lot of attention in current development of risk measures. Under proper conditions, it is also known as tail value at risk, Tail VaR (Artzner et al. (1999)) or CTE, conditional tail expectation (Wirch and Hardy (1999)). Differences between most definitions of expected shortfall were discussed in Tasche (2002a) and Tasche (2002b). The expected shortfall is preferred as a standard risk measure presently, since it satisfies the axioms of coherence proposed by Artzner et al. (1999). Expected shortfall can be generally represented as an average of excess losses or mean shortfall and formally defined at the confidence level α in Definition 1.6.

¹ $q_\alpha(Y) = \inf\{y : P(Y \leq y) \geq \alpha\}$ is the lower α -th quantile of Y

Definition 1.6. (Expected Shortfall; Tasche (2002b)). *Suppose that (Ω, \mathcal{A}, P) is a probability space and a fixed confidence level $\alpha \in (0, 1)$ close to 1. Define X^- as $\max(0, -X)$. Consider a real random variable (r.v.) X on (Ω, \mathcal{A}, P) with $E(X^-) < \infty$ and the indicator function I_A ². Then*

$$ES_\alpha(X) = -(1 - \alpha)^{-1} \{E[X \cdot I_{\{-X \geq \text{VaR}_\alpha(X)\}}] + \text{VaR}_\alpha(X) \cdot (\alpha - P[-X < \text{VaR}_\alpha(X)])\} \quad (1.2)$$

is called Expected Shortfall (ES). The most popular way to define ES (Tasche, 2002b; McNeil et al., 2005) is that

$$ES_\alpha(X) = \frac{1}{1 - \alpha} \int_\alpha^1 \text{VaR}_u(X) du, \quad (1.3)$$

which is equivalent to (1.2).

Equation (1.2) was first introduced in Acerbi et al.. The representation of ES in (1.3) was first established by Wirch and Hardy (1999) for $X \geq 0$ and given by Bertsimas et al. (2000) for continuous distributions. The representation in (1.3) is easy to be described as the *average of the 100(1 - α)% worst losses* of X . It follows that $ES_\alpha(X) \geq \text{VaR}_\alpha$. The proof of the equivalence of (1.2) and (1.3) can be found in Acerbi and Tasche (2002, Proposition 3.2).

The proof of coherency of ES can be found in Acerbi and Tasche (2002). Besides, there are two alternative ways to show the coherency of ES and the detail will be summarized in Section 1.2.1 later. One of them takes advantage of the connection between ES and spectral risk measures, which will be introduced in the next section. The other way takes advantage of the connection between ES and scenario-based risk measures, described in Proposition 1.7 below.

Proposition 1.7. *Suppose that given a model with distribution P and the probability space (Ω, \mathcal{A}, P) . Consider the family \mathcal{Q} of those measures which are boundedly stressed versions of P , in the form of*

$$\mathcal{Q} = \left\{ Q : \frac{dQ}{dP} \leq k \right\}, \quad (1.4)$$

for some constant $k > 1$. Afterwards, by taking $k = 1/(1 - \alpha)$, the scenario-based risk measure $\rho_{\mathcal{Q}}(X) = \sup_{Q \in \mathcal{Q}} E_Q(-X)$ turns out to be

$$\frac{1}{1 - \alpha} \int_\alpha^1 \text{VaR}_u(X) du,$$

which is exactly the ES with respect to P at confidence level α .

² $I_A = \begin{cases} 1, & \text{if event A is true} \\ 0, & \text{if event A is false} \end{cases}$

The discussion and proof of Proposition 1.7 can be found in Delbaen (2002). Proposition 1.7 indicates that for ES as a risk measure, there exists a family of probability measures on the set of states of nature so that ES is coherent by Proposition 1.4. Hence this gives an alternative way to show coherency of ES.

The following introduces *conditional tail expectation* (CTE) or namely *tail conditional expectation* (TCE), which is closely related to ES but does not coincide in general. CTE has received extensive attention due to its identity to ES under suitable conditions such as the continuity of the distribution function $F_X(x)$.

Definition 1.8. (Conditional Tail Expectation (CTE)). *The conditional tail expectation for a real random variable X at confidence level α is given by*

$$\text{CTE}_\alpha(X) = \text{E}[-X | X \leq -\text{VaR}_\alpha(X)]. \quad (1.5)$$

Generally CTE may not be coherent due to violation of subadditivity, as mentioned by Acerbi and Tasche (2002, Example 5.4) and stated by Delbaen (2002, Theorem 6.10).

The relationship between ES and CTE is synthesized as follows. Manifestly ES is continuous in α in (1.3) while it is not in (1.2). This is a distinguishing feature of ES which CTE and VaR do not possess.

Proposition 1.9. *Generally, $\text{CTE}_\alpha(X) \leq \text{ES}_\alpha(X)$.*

Proposition 1.10. *$\text{CTE}_\alpha(X) = \text{ES}_\alpha(X)$ if and only if one of the following conditions holds.*

1. $\text{Pr}[X \leq -\text{VaR}_\alpha(X)] = 1 - \alpha$.
2. $\text{Pr}[X < -\text{VaR}_\alpha(X)] = 0$.

Although similar statements and proofs of Proposition 1.9 and 1.10 can be found in Acerbi and Tasche (2002, Corollary 5.2 and 5.3, respectively), herein more precise and direct proofs of them are provided in Appendix A and Appendix B, respectively. It is observed from the proofs that a first sufficient condition for CTE and ES to coincide is the continuity of the distribution of X .

A series of papers derive analytical CTE formulas for various continuous distributions, in which cases the CTE amounts to the ES. Panjer (2002) derived the CTE formula as a function of mean and variance for the multivariate Gaussian distributions. Landsman and Valdez (2003, 2005) generalized the CTE formula to elliptical distributions and exponential dispersion models, respectively. Hardy (2003) developed the CTE formula for the regime-switching lognormal model. The CTE formula for a multivariate gamma distribution was provided by Furman and Landsman (2005). The CTE formula for the multivariate skew elliptical distributions and phase-type distributions was provided by Cai and Li (2005) and Cai and Tan (2006), respectively.

Kim (2007) derived the CTE formula for exponential family. Furthermore, Jalal and Rockinger (2006) investigated the consequences for VaR and ES of using a GARCH filter on various mis-specified process. McNeil et al. (2005) provides comprehensive formulas of conditional ES for both univariate and multivariate GARCH models. Furman and Landsman (2008) derived analytical CTE formulas for multivariate Poisson distributions, which are not continuous, and hence take notice of that its CTE does not equal to its ES.

Spectral Risk Measures

Employing ES as the basic building block, Acerbi (2002) proposed *spectral risk measures* by generating a new class of risk measures based on ES. It turns out that spectral risk measures take explicit account of a user's degree of subjective risk aversion which neither VaR nor ES does. Please refer to Acerbi (2002) for more details therein. An additional preliminary is needed before introducing it. Let a norm space $\mathcal{L}^1([0, 1])$ be a space, where every element is represented by a class of functions which differ at most on a subset of $[0, 1]$ of zero measure. The norm in this space is defined by

$$\|\phi\| = \int_0^1 |\phi(p)| dp.$$

The Spectral Risk Measure (SRM) is defined in Definition 1.11.

Definition 1.11. (Risk Aversion Function and Spectral Risk Measure; Acerbi (2002)).

The class of Spectral Risk Measure M_ϕ of a r.v. X , generated by an admissible risk spectrum $\phi \in \mathcal{L}^1([0, 1])$, is given by all measures of the type:

$$M_\phi(X) = \int_0^1 \text{VaR}_p(X) \phi(p) dp, \tag{1.6}$$

where $\phi : [0, 1] \rightarrow \mathbb{R}$ is then called the Risk Aversion Function of M_ϕ , satisfying the following conditions.

- *Non-negativity*

$$\phi(p) \geq 0 \quad \text{for all } p \in [0, 1]$$

- *Normalization*

$$\|\phi\| = \int_0^1 |\phi(p)| dp = \int_0^1 \phi(p) dp = 1$$

- *Monotonicity (non-increasingness)*

$$\phi(p_1) \geq \phi(p_2) \quad \text{for all } 0 \leq p_1 \leq p_2 \leq 1.$$

i.e.

$$\phi^{(1)}(p) \leq 0$$

provided that the first derivative of ϕ exists.

The ϕ function is nothing but a weight function in (1.6) which averages the possible outcomes of the random variable X . Besides, different representative functions ϕ_1, ϕ_2 ($\|\phi_1 - \phi_2\| = 0$) of the same element $\phi \in \mathcal{L}^1([0, 1])$ will actually define the same measure M_ϕ . Furthermore, as stated in Theorem 1.12 below M_ϕ defined above is coherent if and only if its risk spectrum ϕ is admissible, which means it satisfies all the conditions therein.

Theorem 1.12. (Acerbi (2002)[Theorem 4.1]) *A spectral risk measure is a coherent risk measure if and only if the risk spectrum, which generates it, is admissible.*³

The *Monotonicity (non-increasingness)* property provides an intuitive insight of the concept of coherence. As a matter of fact, the necessary and sufficient condition for a SRM to be coherent, which states that its risk spectrum ϕ should be admissible, gives a reasonable rule: *a risk measure is coherent if it assigns larger weights to worse cases.* For example, the risk spectra for VaR and ES can be written in the form of

$$\delta(p - \alpha)$$

and

$$\frac{1}{1 - \alpha} I_{\{\alpha \leq p \leq 1\}},$$

respectively, where $\delta(x)$ is the Dirac Delta function defined by

$$\int_a^b f(x)\delta(x - c) dx = f(c), \quad \forall c \in (a, b).$$

As a result, it is easy to see that VaR is not coherent, since its risk spectrum is not admissible. Similarly, ES is coherent, since its risk spectrum is admissible.

By the above proposition, for a spectral risk measure $M_\phi(X)$ with admissible risk spectrum, there exists a family \mathcal{P} of probability measures on the set of states of nature, such that

$$M_\phi(X) = \rho_{\mathcal{P}}(X),$$

where $\rho_{\mathcal{P}}(X)$ is the scenario-based risk measure defined by \mathcal{P} . However, it remains an open question of how to characterize the existing family \mathcal{P} of probability measures for $M_\phi(X)$. Since ES is a special case of scenario-based risk measures and of spectral risk measures, the next sec-

³Notice that however, in Acerbi (2002), a *coherent risk measure* is simply called a *risk measure*

tion will discuss more on coherency and ES by exploiting the connection between ES, scenario-based risk measures and spectral risk measures, and the fact that scenario-based risk measures and spectral risk measures are all coherent.

Supplementary

It has been shown that ES satisfies all properties defined in Definition 1.1 and hence is coherent. By incorporating previous results in the foregoing sections, it is also easy to show the coherency of ES in two alternative ways as follows.

1. Proposition 1.7 indicates that for ES as a risk measure, there exists a family \mathcal{P} of probability measures on the set of states of nature and hence ES is a scenario-based risk measure defined by \mathcal{P} . Thus ES is coherent by Proposition 1.4.
2. We know that ES is a spectral risk measure generated by the admissible risk spectrum $\phi(p) = \frac{1}{1-\alpha} I_{\{\alpha \leq p \leq 1\}}$. Therefore, ES is coherent by Theorem 1.12.

1.2.2 Capital Allocation

The capital allocation problem is that once the total risk capital for a multi-unit financial institution (or a combination of different portfolios) is computed based on a specific risk measure, the risk capital must be allocated back to each business unit (or portfolio) in a consistent way which should recognize the benefit of diversification. Different allocation methods are available; however, only some of them satisfy the nice properties which have economical meanings, while others do not. Those desirable properties were proposed as axioms of *coherent allocation* (Denault, 2001) and *fair allocation* (Kim, 2007) (adapted from Valdez and Chernih (2003) which extended Wang's⁴ idea to the elliptical distribution class for capital allocation problem), respectively. They will be discussed in the next section.

Axioms of Allocation Principles

Consider an investor who can invest in a fixed set of d individual investment possibilities with P&L denoted by the random variables X_1, X_2, \dots, X_d . Let $C = \rho(S)$ and C_i denote the overall risk capital determined by a particular risk measure, and the allocation of C to the i -th portfolio, respectively. The P&L of the actual portfolio is of course $S = \sum_{i=1}^d X_i$. An allocation principle is a function that maps allocation problems associated with a specified risk measure into a unique allocation. Next section will cover axioms of coherent allocation principle and those of fair allocation principle.

⁴See Wang (2002) for reference.

Definition 1.13. (Axioms of Coherent Allocation; Denault (2001)). *An allocation principle is coherent if it satisfies the following axioms:*

1. **Full allocation:** *The allocated risk capitals add up to the total risk capital.*

$$\rho(S) = \sum_{i=1}^d C_i \quad (1.7)$$

2. **No undercut:** *For any subset M of $\{1, 2, \dots, d\}$,*

$$\sum_{k \in M} C_k \leq \sum_{k \in M} \rho(X_k) \quad (1.8)$$

3. **Symmetry:** *If by joining any subset M of $\{1, 2, \dots, d\} \setminus \{i, j\}$, portfolios i and j have identical contribution to the risk capital, then $C_i = C_j$.*

4. **Riskless allocation:** *If the d^{th} portfolio is a riskless instrument, then*

$$C_d = \rho(\alpha r) = -\alpha$$

where r is the price, at the end of the holding period, of a reference riskless instrument whose price is 1 at the beginning of the period.

Full allocation means that the overall risk capital is fully allocated back to the individual portfolio positions. No undercut addresses the diversification effect as a reward. That is, an investor holding the portfolio could benefit from the diversification of combination of portfolios. Symmetry ensures that the allocation of a portfolio depends only on its contribution to risk and nothing else. Riskless allocation means that when all other things are equal, a portfolio that increases its cash position results in decrease of its allocated capital by the same amount.

Definition 1.14. (Axioms of Fair Allocation). *An allocation principle is fair if it satisfies the following axioms:*

1. **Full allocation:** *The allocated risk capitals add up to the total risk capital.*

$$\rho(X) = \sum_{i=1}^d C_i \quad (1.9)$$

2. **No undercut:** *For any subset M of $\{1, 2, \dots, d\}$,*

$$\sum_{k \in M} C_k \leq \sum_{k \in M} \rho(X_k) \quad (1.10)$$

3. **Symmetry:** *If by joining any subset M of $\{1, 2, \dots, d\} \setminus \{i, j\}$, portfolios i and j have identical contribution to the risk capital, then $C_i = C_j$.*
4. **Consistency:** *For any subset M of $\{1, 2, \dots, d\}$ with $|M| = m$, let $X^* = \sum_{i \in M} X_i$ and $\mathbf{X}_{d-m} = (X_{j_1}, \dots, X_{j_m})'$ for all $j_k \in \{1, 2, \dots, d\} \setminus M$, where $k = 1, \dots, d - m$. Then the new hierarchical structure of the d individual portfolios becomes X^*, \mathbf{X}'_{d-m} as a combination of individual $d - m + 1$ portfolios. Let C_1^* denote the allocated capital of the portfolio X^* as an individual one, computed on X^*, \mathbf{X}'_{d-m} . Afterward*

$$\sum_{i \in M} C_i = C_1^*. \quad (1.11)$$

Obviously, the definition of fair allocation adopts the first three axioms of coherent allocation but replaces the last one, riskless allocation, by *consistency*. Consistency ensures that the allocation of an individual portfolio cannot depend on the level at which allocation is calculated. For instance, consider a company consisting of multiple units of business at the top level and sub-units of business at lower level. According to the consistency axiom, the capital allocation for a unit of business at the top level must be identical whether the allocation of its sub-units has been directly computed to the company or indirectly first to the top level and later aggregated for the company. In sum, the allocated capital for one unit of business should be independent of the hierarchical structure of the company.

Euler Capital Allocation Principle

Consider the random variables X_1, X_2, \dots, X_d on a common probability space (Ω, \mathcal{A}, P) as the P&L for d individual investment possibilities. Let $C = \rho(S)$ and C_i denote the total risk capital determined by a specified risk measure, and the allocation of C to the i -th portfolio, respectively. Furthermore, consider a set $U \subset \mathbb{R}^d \setminus \{(0, \dots, 0)\}$ of portfolio weights and define $S(u) = \sum_{i=1}^d u_i X_i$ for $u = (u_1, \dots, u_d) \in U$. The total P&L of the actual portfolio is of course $S(I) = \sum_{i=1}^d X_i$. Let ρ denote some risk measure defined on a set \mathcal{M} which contains the random variables $\{S(u) : u \in U\}$. Then introduce the associated risk-measure function $f_\rho : U \rightarrow \mathbb{R}$

$$f_\rho(u) = \rho(S(u)). \quad (1.12)$$

Thus $f_\rho(u)$ is the required total risk capital for a position u in the set of investment possibilities.

Definition 1.15. *Let f_ρ be a risk-measure function on some set $U \subset \mathbb{R}^d \setminus \{(0, \dots, 0)\}$ such that $(1, \dots, 1) \in U$. A mapping $\pi_{f_\rho} : U \rightarrow \mathbb{R}^d$ is called a per-unit capital allocation principle*

associated with f_ρ , if

$$\sum_{i=1}^d u_i \pi_i = f_\rho(u), \quad (1.13)$$

where π_i represents the i -th element of $\pi_{f_\rho}(u)$.

This definition interprets that the i -th element of $\pi_{f_\rho}(u)$ gives the amount of capital to one unit of X_i , when the overall position has P&L $S(u)$. The amount of capital allocated to the position $u_i X_i$ is $u_i \pi_i$ and the equality (1.13) is simply the full allocation property.

Definition 1.16. (Positive Homogeneity). *A function $f : U \subset \mathbb{R}^d \rightarrow \mathbb{R}$ is called homogeneous of degree k if for any $\gamma > 0$ and $u \in U$ with $\gamma u \in U$ the following equation holds:*

$$f(\gamma u) = \gamma^k f(u)$$

where k is a positive integer.

Note that *positive homogeneity* in McNeil et al. (2005) is in fact *positive homogeneity of degree 1*. If the risk measure ρ is homogeneous of degree k , then the risk-measure function f_ρ corresponding by (1.12) to ρ is homogeneous of degree k . In the case of continuously differentiability, homogeneous functions can be characterized by Euler's theorem as follows.

Theorem 1.17. (Euler's theorem). *Let $U \subset \mathbb{R}^d$ be an open set and $f : U \rightarrow \mathbb{R}$ be a continuously differentiable function. Then f is homogeneous of degree k if and only if it satisfies*

$$k f(u) = \sum_{i=1}^d u_i \frac{\partial f}{\partial u_i}(u).$$

Therefore, by Euler's theorem, the risk-measure function $f_\rho(u)$ satisfies

$$f_\rho(u) = \sum_{i=1}^d u_i \frac{\partial f_\rho}{\partial u_i}(u) \quad (1.14)$$

if and only if it is homogeneous of degree 1. Hence comparison of (1.13) and (1.14) gives *Euler capital allocation principle* defined in Definition 1.18.

Definition 1.18. (Euler capital allocation principle). *Suppose that a risk-measure function f_ρ is positive homogeneous of degree 1 on some set $U \subset \mathbb{R}^d \setminus \{(0, \dots, 0)\}$ such that $(1, \dots, 1) \in U$. Assume that f is continuously differentiable; then the per-unit Euler capital allocation principle associated with f_ρ is the mapping*

$$\pi_{f_\rho} : U \rightarrow \mathbb{R}^d, \quad \pi_{f_\rho}(u) = \left(\frac{\partial f_\rho}{\partial u_1}(u), \frac{\partial f_\rho}{\partial u_2}(u), \dots, \frac{\partial f_\rho}{\partial u_d}(u) \right). \quad (1.15)$$

The Euler capital allocation principle is also called *allocation by the gradients* and satisfies the full allocation property, which means that the overall risk capital $f_\rho(u)$ is fully allocated back to the individual portfolio positions. Tasche (2007, 2008) showed that the risk contributions to some risk measure ρ with continuously differentiable f_ρ are uniquely determined by Euler capital allocation with $u = (1, \dots, 1)$ if and only if f_ρ is positive homogeneous of degree 1. Then those risk contributions, uniquely determined by the Euler capital allocation with $u = (1, \dots, 1)$, are called Euler contributions. A comprehensive description of the Euler capital allocation principle and more reference of it can be found in Tasche (2007, 2008). The Euler allocation principle was also justified by other authors by different approaches. Denault (2001) derived the Euler allocation principle by game theory and it was regarded as *Aumann-Shapley per-unit allocation* for homogeneous risk measures of degree one. That is, the *Aumann-Shapley per-unit allocation* is equivalent to the *Euler capital allocation*. Additionally, it is shown as a coherent allocation of risk capital as well in his paper.

In the continuous case, by employing the risk-measure function

$$\rho_{\text{ES}_\alpha}(u) = (1 - \alpha)^{-1} \int_\alpha^1 \rho_{\text{VaR}_p}(u) dp$$

where $\rho_{\text{VaR}_p}(u) = \text{VaR}_p(u)$, the Euler risk contributions corresponding to the risk measure ES is called expected shortfall contributions (ESC) and given by

$$\text{ESC}_\alpha(X_k) = -E[X_k | S \leq -\text{VaR}_\alpha(S)], \quad k = 1, \dots, n. \quad (1.16)$$

The natural extension of CTE for capital allocation is coherent Acerbi and Tasche (1999), was also discussed in Tasche (1999), and Overbeck (2000) and briefly mentioned in Denault (2001). The capital allocation derived from the particular risk measure, CTE, is then called *CTE based capital allocation rule* and abbreviated to *CTE allocation* thereby. Notice that CTE allocation amounts to ESC given above while the two risk measures, CTE and ES coincide. It lies in the category of Euler allocation principle and is shown as not only a coherent allocation principle (Denault, 2001) but also a fair allocation (Kim, 2007). Several papers have derived analytic expressions of CTE allocation for various parametric models. Panjer (2002) proved that CTE allocation coincides with the allocation of the capital asset pricing model (CAPM), which is sometimes referred to as *covariance-based* allocation, in the special case of multivariate Gaussian distributions. Solcá (2000) derived CTE allocation as well as allocation based on covariance principle for Gaussian distributions. CTE allocation for elliptical distribution class is extended in Landsman and Valdez (2003) and Dhaene et al. (2007) where two other cases, log-elliptical distribution class and comonotonic random vector, were also discussed. Furman and Landsman (2008) derived analytical expressions for multivariate Poisson distributions.

1.2.3 Summary of Contributions

- Comprehensive literature review
- Direct and clear statements in Proposition 1.9 and 1.10 and their proofs

1.3 Overview of Thesis

All studies hereafter are started from the losses as the data, which is merely the negative value of the P&L, for convenience. In other words, suppose that a random sample $\mathbf{S} = (S_1, \dots, S_n)'$ of potential total losses such that $S_i = \sum_{j=1}^m X_{i,j}$, $i = 1, 2, \dots, n$. The formulas of expected shortfall and its contributions are then given by

$$\text{ES}_\alpha(S) = \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_u(S) du \quad (1.17)$$

and

$$\text{ESC}_\alpha(X_k) = \text{E}[X_k | S \geq \text{VaR}_\alpha(S)]; \quad k = 1, \dots, m, \quad (1.18)$$

respectively. Chapter 2 studies bias adjustment using bootstrap techniques for estimating ES. Chapter 3 researches evaluation criteria for risk models, divided into two aspects consisting of model-based criteria and data-based criteria, along with a simulation study of examining how the bias adjustment works for improving ES estimates using model-based criteria, and another simulation study of backtesting VaR and ES using data-based criteria. Chapter 4 introduces a multivariate framework for estimating ES and ESC. Chapter 5 carries out an empirical study of estimating ES and ESC with backtesting examination, using multivariate framework, for historical data from Russell Indexes. The final chapter gives a final summary and discussion, followed by future topics.

Chapter 2

Bias Adjustment Using Bootstrap Techniques for Estimating ES

2.1 Introduction

Dowd (2005a) and Hardy (2006) suggested creating confidence intervals for ES by the non-parametric bootstrap, and Baysal and Staum (2007) implemented it. Cotter and Dowd (2007) applied the non-parametric bootstrap to estimate financial risk measures, including ES, for future positions. Scaillet (2004) applied kernel density estimators to estimate ES and ESC along with the asymptotic properties of the kernel estimators of ES. The notion was extended to estimate conditional ES in Scaillet (2005) and Cai and Wang (2006) where conditional VaR was estimated as well. However, Chen (2008) revealed that the extra kernel smoothing (Scaillet (2004)) does not improve the HS estimate of ES, while it does in VaR case.

There are two major concerns in the recent history of ES estimation. One is heavy-tailed distributions and the other is changing volatility models. *Extreme value theory* (EVT) is widely applied in estimating heavy-tailed distributions, and hence for VaR and ES. Generalized autoregressive conditional heteroscedastic (GARCH) and *exponentially weighted moving average* (EWMA) models are popular volatility models (or volatility filters) for daily risk-factor return series, and ES derived from changing volatility models are called *conditional ES* since it is defined conditionally on the past. McNeil and Frey (2000) showed that EVT is preferable to model the tail of a heavy-tailed distribution such as the innovation distribution of a GARCH model, and that *generalized Pareto Distribution* (GPD) method is preferred to the approach based on the Hill method (Hill, 1975) for estimating high quantiles such as VaR. McNeil et al. (2005) provides comprehensive formulas of conditional ES for both univariate and multivariate GARCH models and methods of fitting an appropriate (M)GARCH model. Pattarathammas et al. (2008) showed that GARCH and EWMA are both useful in VaR and ES estimation, in

an empirical application of ten East Asian MSCI¹ Country Indices, and that the conditional extreme value theory models tend to be better than other models with normality assumption, while they are not trivially different from the filtered historical simulation model (conditional HS). Jalal and Rockinger (2006) investigated the consequences for VaR and ES of using a GARCH filter on various mis-specified processes.

Accurate estimation of the ES and its variability becomes the top priority. Although the empirical ES will converge to the true one as the sample size increases, it is systematically biased and can be improved by statistical methods. For example, Kim (2007) used the bootstrap techniques to adjust the empirical estimates of VaR and CTE for bias and to estimate its variance; meanwhile, an alternative technique, non-parametric delta method employing influence function, was also provided in this material to explore the variability of the estimated risk measures. The issue of bias becomes a more serious statistical problem, especially when the sample size is quite small, although the bias tends to zero as the sample size gets larger. However, blindly adjusting an estimate for bias can be dangerous in practice due to high variability in the estimated bias. Bias adjustment may be accompanied by larger increase on the standard error, which in turn causes a larger mean square error. Thus there is a practical guideline in (Efron and Tibshirani, 1993, section 10.6) stated as follows. In case the estimated bias is small compared to the estimated standard error, it is safer to use the unadjusted estimate rather than its bias-adjusted estimate. Otherwise, if the estimated bias is large compared to the estimated standard error, then it may be better to use the bias-adjusted estimate rather than the original estimate. There arises another concern that the estimated bias should have the same sign (negative or positive) as the true bias has, in order to adjust the bias in the right direction. Therefore, the bias adjustment procedure will work well generally if the following two conditions are both satisfied: (1) The ratio of the estimated bias to the estimated standard error is large. (2) The true bias and its estimate are either both negative or both positive.²

Section 2.2 will introduce the most common resampling methods, bootstrap techniques, and describe how to apply them to improve estimates by adjusting for bias. Then Section 2.3 will discuss how to apply bias adjustment for improving ES estimates for i.i.d. models. Finally, Section 2.4 will introduce the application of bias adjustment for ES estimates of several common types of time series models. The last section will summarize the contributions of this chapter.

¹A stock market index of 1500 'world' stocks. It is maintained by MSCI Inc., formerly Morgan Stanley Capital International and is often used as a common benchmark for 'world' or 'global' stock funds.

²Although the true bias is usually unknown, it is possible to know whether the true quantity of interest is larger than its estimate or not and hence the sign of the true bias is known.

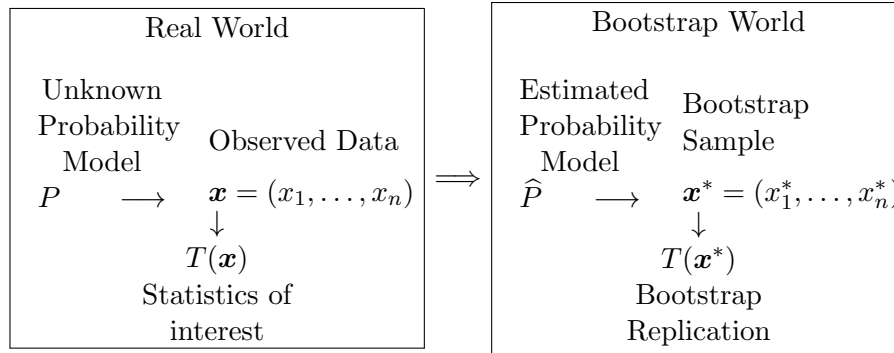


Figure 2.1: Real World vs. Bootstrap World

2.2 Bootstrap Techniques

The bootstrap is particularly useful for non-parametric statistical inference since it was originally proposed to estimate or approximate the sampling distribution of a statistic and its characteristics. The essential idea is to use available data at hand to estimate or approximate some accuracy measures of a statistic (estimator) of interest, such as the bias, the variance, and the mean square error. The block-wise bootstrap (Carlstein, 1986; Shi, 1986) is designed to capture the dependence structure of the data in the resampling process under no specific model assumption made on the data and hence are less *model-dependent* than other residual-based modifications of bootstrap. It can work well for general stationary process with short-range dependence. Standard textbooks are available as reference for comprehensive methodology such as Davison and Hinkley (1997), Efron and Tibshirani (1993), and Shao and Tu (1995).

The bootstrap, one of the resampling methods, was introduced by Efron (1979) as a crucial step in statistical analysis. The basic idea is to model sampling from a population by the process of resampling from the sample. In other words, bootstrap mimics the relationship from the population to the sample by the relationship from the original sample or the data to the bootstrap sample. The ordinary bootstrap generates bootstrap samples by resampling simulation (with replacement) from the original sample. Then most features of the relationship of the original sample to its underlying distribution are replicated by the relationship of the bootstrap samples to the original sample. This can be demonstrated by Figure 2.1. The non-parametric bootstrap can be applied properly for both parametric and non-parametric models. However, the non-parametric bootstrap may fail in some examples such as those in which the parameter of interest depends on the smoothness of the true distribution (see Efron and Tibshirani, 1993, section 7.4 for more information therein).

2.2.1 The Ordinary Bootstrap

The fundamental procedure of the ordinary bootstrap (OB) can be briefly illustrated as follows. Suppose that the random sample $\mathbf{X} = (X_1, \dots, X_n)'$ of size n is i.i.d.³ from an unknown distribution with cumulative distribution function F which is estimated by \widehat{F}_n . The empirical distribution function \widehat{F}_n is a simple non-parametric estimator of F and defined by

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq x\}}$$

where $I_{\{A\}}$ is the indicator function of the set A . Let $t(F)$ be the true parameter of interest and suppose that it can be estimated by the statistic $T = t(\widehat{F}_n)$, the *plug-in* estimate. A pseudo sample of the same size as the original one is generated from the e.d.f. \widehat{F}_n by resampling with replacement from data $(\mathbf{X} = (X_1, \dots, X_n))$ and the generated sample, denoted by \mathbf{X}^* , is called a bootstrap sample. Note that the capital letters indicate that it is a random sample as well, but from the e.d.f., indicated by superscript $*$. The statistic of interest computed from this bootstrap sample is then expressed by $T^* = T(\mathbf{X}^*)$. The different bootstrap samples $\mathbf{X}_1^*, \dots, \mathbf{X}_B^*$ are generated by repeating the above resampling simulation B times. For each sample, the statistic of interest T_k^* is generated from the k -th bootstrap sample, giving $T_k^* = T(\mathbf{X}_k^*)$, $k = 1, \dots, B$. Eventually the bootstrap estimate of the expected value of a statistic T is given by

$$E(T|\widehat{F}_n) = E^*(T^*) \approx \frac{1}{B} \sum_{b=1}^B T_b^*. \quad (2.1)$$

Here the condition \widehat{F}_n in the first term states that the expectation is taken with respect to the e.d.f. while the true value $t(F)$ can be written by $E(T|F)$. Let standard bootstrap estimator \overline{T}_B^* denote $B^{-1} \sum T_b^*$.

Algorithm 2.1 below is an explicit description of the bootstrap procedure, provided by Efron and Tibshirani (1993).

Algorithm 2.1. (The bootstrap algorithm for estimating bias and standard error of T).

1. Generate B independent bootstrap samples $\mathbf{X}_1^*, \mathbf{X}_2^*, \dots, \mathbf{X}_B^*$, each consisting of n data values drawn with replacement from \mathbf{X} .
2. Compute the statistic of interest corresponding to each bootstrap sample: $T_b^* = T(\mathbf{X}_b^*)$, $b = 1, \dots, B$

³independent and identically distributed

3. • Estimate the bias by the sample bias of the B replications

$$\widehat{\text{bias}}_B^{OB} = \bar{T}_B^* - t(\widehat{F}_n) \quad (2.2)$$

- Estimate the standard error by the sample standard deviation of the B replications

$$\widehat{\text{se}}_B^{OB} = \left\{ \sum_{b=1}^B [T_b^* - \bar{T}_B^*]^2 / (B - 1) \right\}^{1/2}. \quad (2.3)$$

In other words, the unknown bias $E(T|F) - t(F)$ is approximated by its bootstrap estimate (2.2) under B resampling along with the bootstrap estimate of the standard error for T given in (2.3). Sometimes it is possible to calculate $E(T|\widehat{F}_n)$ analytically without actually performing the resampling simulation. The bootstrap estimate will be close to the true estimated one as the resampling size B increases; that is, the bootstrap estimate converges to the true estimated one as $B \rightarrow \infty$. The difference between them is called the resampling (simulation) error. If this is achieved, then the remaining uncertainty is only ascribed to the original statistical error. The limit of $\widehat{\text{bias}}_B^{OB}$ and $\widehat{\text{se}}_B^{OB}$ are then the ideal bootstrap estimate of $\text{bias}_F(T)$ and $\text{se}_F(T)$ as B goes to infinity. Ultimately the bias-adjusted estimate is yielded by

$$t(\widehat{F}_n) - \widehat{\text{bias}}_B^{OB} = 2 \cdot t(\widehat{F}_n) - \bar{T}_B^*. \quad (2.4)$$

Furthermore, if $T = t(\widehat{F}_n)$, then (2.4) can be rewritten as

$$t(\widehat{F}_n) - \widehat{\text{bias}}_B^{OB} = 2T - \bar{T}_B^*. \quad (2.5)$$

Before applying bias adjustment by bootstrap techniques, one thing to note is that the bootstrap estimate of bias defined by

$$E(T|\widehat{F}_n) - t(\widehat{F}_n)$$

can be replaced by

$$\bar{T}_B^* - T$$

only when T is the *plug-in* estimate of θ , $t(\widehat{F}_n)$. Otherwise, the true bias will not be approximated correctly by $\bar{T}_B^* - T$. It is a misconduct of bias adjustment by using $\bar{T}_B^* - T$ as the bootstrap estimate of bias, because it fails to detect the bias if T is not the plug-in estimate $t(\widehat{F}_n)$. That is, the bias-adjusted estimate $2 \cdot t(\widehat{F}_n) - \bar{T}_B^*$ can not be substituted by $2T - \bar{T}_B^*$ if $T \neq t(\widehat{F}_n)$. This can be demonstrated further in Example 2.2.

Example 2.2. (An example of misconducted bias adjustment)

Suppose that a random sample is given from the distribution function F with mean μ_F and that the statistic of interest T is used to estimate the parameter $\theta = t(F)$, which can be written as a function of the distribution F . Furthermore, suppose that T is not exactly the plug-in estimate $t(\widehat{F})$ for θ , where \widehat{F} is the empirical estimate of F . For instance, if T is a linear function of $t(\widehat{F}_n)$ given by

$$T = a \cdot t(\widehat{F}_n) + b = a \cdot \bar{x}_n + b$$

for some constants $a \neq 0$ and $b \neq 0$ then obviously, $T \neq t(\widehat{F}_n)$. Now here is the point as follows. The true bias defined by $E(T|F) - t(F)$ can be approximated by the bootstrap estimate $E(T|\widehat{F}_n) - t(\widehat{F}_n)$ rather than $\bar{T}_B^* - T$. It is clear that the latter quantity

$$\begin{aligned} \bar{T}_B^* - T &= (a \cdot \bar{x}_{n,B}^* + b) - (a \cdot \bar{x}_n + b) \\ &= a \cdot [\bar{x}_{n,B}^* - \bar{x}_n] \end{aligned}$$

will fail to approximate the true bias, because it tends to be zero especially as B is getting large. In other words, it is not able to approximate the true bias by the estimate $\bar{T}_B^* - T$ in the case of $T \neq t(\widehat{F}_n)$, and even ends up with nothing as B increases. Therefore, the bias-adjusted estimate $2 \cdot t(\widehat{F}_n) - \bar{T}_B^*$ can not be substituted by $2T - \bar{T}_B^*$ if $T \neq t(\widehat{F}_n)$.

2.2.2 The Exact Bootstrap of L -statistics

In practice, it is not always possible that the replicating size B of bootstrap samples gets large enough to completely eliminate the resampling error. Hutson and Ernst (2000) established the formulas for the exact bootstrap (EB) mean and variance of the L -statistic which is a linear function of order statistics with constant coefficients. It gives the direct calculation of bootstrap estimates without resampling process. Then the bootstrap is exact in the sense that the resampling error is entirely removed from the process or equivalently in the sense of bootstrapping with infinitely many bootstrap samples.

The formal definition of the L -statistic is given in Definition 2.3 and the applicable formulas are given in Theorem 2.4.

Definition 2.3. *L -statistics.* Let $X_{1:n} \leq X_{2:n} \leq \dots \leq X_{n:n}$ denote the order statistics of an independent and identically distributed sample of size n from a continuous distribution F with support over the entire real line. An L -statistic (or L -estimator) is defined as

$$T_n = \sum_{i=1}^n c_i X_{i:n},$$

where the choice of constants c_i determines the properties of T_n .

Theorem 2.4. (Exact bootstrap of L -statistic; Hutson and Ernst (2000)) *The exact bootstrap (EB) of the L -statistic, $E(X_{(r)}|F), 1 \leq r \leq n$ is given by*

$$E(X_{r:n}|\widehat{F}_n) = \sum_{j=1}^n \omega_{j(r)} X_{(j)},$$

where $\omega_{j(r)} = r \binom{n}{r} [B(\frac{j}{n}; r, n-r+1) - B(\frac{j-1}{n}; r, n-r+1)]$, and $B(x; a, b) = \int_0^x t^{a-1} (1-t)^{b-1} dt$ is the incomplete beta function.

For convenience, the weight for each element of the ordered sample $\mathbf{X}_{:n}$, is rewritten in terms of matrix by \mathbf{w} with elements $w_{i,j} = \omega_{i(j)}$; $i = 1, \dots, n$ and $j = 1, \dots, n$. Then using Theorem 2.4, any L -statistic can be written as $T = \mathbf{c}' \mathbf{X}_{:n}$ where \mathbf{c} is a column vector of size n . Hence the sample EB of an L -statistic T is obtained as

$$E(T|\widehat{F}_n) = E(\mathbf{c}' \mathbf{X}_{:n}|\widehat{F}_n) = \mathbf{c}' E(\mathbf{X}_{:n}|\widehat{F}_n) = \mathbf{c}' \mathbf{w}' \mathbf{X}_{:n}. \quad (2.6)$$

Then the EB bias estimate is expressed by

$$\widehat{\text{bias}}^{EB} = E(T|\widehat{F}_n) - t(\widehat{F}_n) = \mathbf{c}' \mathbf{w}' \mathbf{X}_{:n} - \mathbf{c}' \mathbf{X}_{:n}. \quad (2.7)$$

It follows that the bias-adjusted estimate becomes

$$t(\widehat{F}_n) - \widehat{\text{bias}}^{EB} = \mathbf{c}' \mathbf{X}_{:n} - (\mathbf{c}' \mathbf{w}' \mathbf{X}_{:n} - \mathbf{c}' \mathbf{X}_{:n}) = \mathbf{c}' (2I - \mathbf{w}') \mathbf{X}_{:n}. \quad (2.8)$$

2.2.3 Blockwise Bootstrap for Dependent Data

Blindly applying the bootstrap for dependent data may lead to incorrect results such as inconsistency, since some properties that the bootstrap possesses for i.i.d. case may collapse for non-i.i.d. case. To obtain the correct result and retain the properties that the bootstrap possesses, the resampling process should be modified for dependent data.

The block-wise bootstrap (BB) is proposed by modifying the resampling process for dependent data, but no specific model assumption is made on the data. The block-wise bootstrap, first proposed by Carlstein (1986) and Shi (1986), is to treat blocks of consecutive observations, instead of single observations, as the primary units of the data. Without loss of generality, suppose that $n = kh$ for two integers k and h such that $k \rightarrow \infty$ and $h \rightarrow \infty$ as $n \rightarrow \infty$. The data are grouped into a set of k non-overlapping blocks with h observations in each block: $\{X_{(j-1)h+1}, X_{(j-1)h+2}, \dots, X_{jh}\}, j = 1, \dots, k$. With the block-wise bootstrap, the idea is to choose the block size h large enough so that those blocks will be nearly independent (or observations more than h units apart will be nearly independent). Hence the ordinary bootstrap can be applied to the set of blocks as to the data set, the set of single observations.

2.3 Bias Adjustment for Simple ES Estimates

Consider a random sample $\mathbf{X} = (X_1, \dots, X_n)'$ of potential total losses. Recall that the formula of expected shortfall is given by

$$\text{ES}_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_u(X) du = \frac{1}{1-\alpha} \int_\alpha^1 q_u(X) du,$$

where $q_u(X) = \inf\{x \in R : P(X \leq x) \geq u\} = \text{VaR}_u(X)$ is the u -th quantile (VaR at confidence level u) of X .

If the above formula can be derived in terms of the parameters of a particular model, then the estimation of ES can be easily implemented as long as the estimation of parameters is carried out first. Hence ES would be well-estimated if the model assumption is correct. Therefore, the basic strategy of estimating ES for a specific model is briefly summarized as follows.

- Step 1: Derive ES as a function of model parameters from the above formula. That is,

$$\text{ES}_\alpha = h(\boldsymbol{\theta}; \alpha)$$

where $\boldsymbol{\theta}$ is the parameter vector of the assumed model.

- Step 2: Estimate the model parameters by fitting the assumed model to sample data.
- Step 3: Estimate ES by computing the function, h at $\hat{\boldsymbol{\theta}}$. That is,

$$\text{ES}_\alpha = h(\hat{\boldsymbol{\theta}}; \alpha).$$

Suppose that $\widehat{\text{ES}}_\alpha$ first estimates ES at level α by common methods above. Then it is easy to implement bias adjustment for $\widehat{\text{ES}}_\alpha$ using variants of bootstrap techniques, since they vary only with the resampling process but keep the same basic idea of adjusting $\widehat{\text{ES}}_\alpha$ for bias. The central idea of bias adjustment for improving $\widehat{\text{ES}}_\alpha$ is given by

$$\widehat{\text{ES}}_\alpha^{new} = \widehat{\text{ES}}_\alpha - \widehat{\text{bias}}(\widehat{\text{ES}}_\alpha) \tag{2.9}$$

where $\widehat{\text{ES}}_\alpha^{new}$ denotes the bias-adjusted estimate and $\widehat{\text{bias}}(\widehat{\text{ES}}_\alpha)$ represents the bootstrapping estimate of the true bias of $\widehat{\text{ES}}_\alpha$.

The following two sections will introduce, in turn, historical simulation and GPD methods to obtain a simple estimate of ES, $\widehat{\text{ES}}_\alpha$, and then for each of them, describe how to implement bias adjustment through (2.9) to get various bias-adjusted estimates $\widehat{\text{ES}}_\alpha^{new}$ by bootstrap techniques described in Section 2.2.

2.3.1 For HS Estimates

Historical simulation (HS) is a non-parametric approach which uses historical data for estimation by simulating from the empirical distribution of data. It is conceptually simple and easy to implement and quite extensively used. The empirical distribution is a building block in historical simulation. Suppose that the distribution function $F(x)$ (also called the cumulative distribution function (cdf) or probability distribution function), describes the probability that a variate X takes on a value less than or equal to a number x . Statistically, it is defined by

$$F(x) = \Pr(X \leq x).$$

The empirical distribution function (e.d.f.) $\hat{F}_n(x)$ is a simple estimate of the true distribution function $F(x)$ and is defined as

$$\hat{F}_n(x) = \frac{1}{n} \sum_{k=1}^n I_{\{x_k \leq x\}}, \quad x \in \mathbb{R}$$

where x_1, \dots, x_n are outcomes of random variables X_1, \dots, X_n whose cumulative distribution function (c.d.f.) is $F(x)$. Intuitively, $\hat{F}_n(x)$ and $F(x)$ are very closely related. In fact, it has been justified in the following theorem.

Theorem 2.5. (Glivenko-Cantelli Theorem). *Assume that X_1, \dots, X_n are i.i.d. random variables with c.d.f. $F(x)$ and that $\hat{F}_n(x)$ is the e.d.f. Define*

$$D_n = \sup_{-\infty < x < \infty} |\hat{F}_n(x) - F(x)|,$$

which is a random variable. Then

$$P(\lim_{n \rightarrow \infty} D_n = 0) = 1.$$

Suppose that X^* is a random variable with the distribution function $\hat{F}_n(x)$ and observed data x_1, \dots, x_n for a fixed n . Thus X^* is a discrete random variable with the probability distribution function

$$P(X^* = x_k) = \frac{\sum_{i=1}^n I_{\{x_i = x_k\}}}{n}, \quad k = 1, \dots, n_m$$

where $1 \leq n_m \leq n$ and n_m is the number of distinct numbers in the data set. Consequently, we have

$$E(X^*) = \sum_{k=1}^{n_m} x_k P(X^* = x_k) = \sum_{k=1}^n \frac{x_k}{n} = \bar{x},$$

and

$$\text{Var}(X^*) = \frac{1}{n} \sum_{k=1}^n (x_k - \bar{x})^2.$$

In general, for any function $g(\cdot)$,

$$\mathbb{E}[g(X^*)] = \int g(x) d\widehat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n g(x_i).$$

A parameter $\theta = t(F)$, a function of the distribution function F , can be estimated as $\widehat{\theta} = t(\widehat{F}_n)$ by the plug-in principle, which estimates parameters from samples without any assumption of the distribution function F .

Recall that the formula of expected shortfall is given by

$$\text{ES}_\alpha(X) = \frac{1}{1-\alpha} \int_\alpha^1 \text{VaR}_u(X) du = \frac{1}{1-\alpha} \int_\alpha^1 q_u(x) du = \frac{1}{1-\alpha} \int_\alpha^1 F^{-1}(u) du,$$

where $F^{-1}(u) = q_u(X) = \inf\{x \in R : F(x) \geq u\} = \text{VaR}_u(X)$ is the lower u -th quantile (VaR at confidence level u) of X as an inverse function of F . Furthermore, define the inverse function of \widehat{F} as

$$\widehat{F}_n^{-1}(u) = \begin{cases} x_{(1)}, & \text{if } 0 \leq u \leq 1/n \\ x_{(j)}, & \text{if } \frac{j-1}{n} < u \leq \frac{j}{n}; j = 2, \dots, n. \end{cases} \quad (2.10)$$

Or using indicator functions, it can be rewritten as

$$\widehat{F}_n^{-1}(u) = s_{(1)} \cdot I_{\{0 \leq u \leq 1/n\}} + \sum_{j=2}^n s_{(j)} \cdot I_{\{\frac{j-1}{n} < u \leq \frac{j}{n}\}}. \quad (2.11)$$

Then by the plug-in principle, the empirical estimator of ES for a given confidence level α is straightforwardly given by

$$\begin{aligned} \widehat{\text{ES}}_\alpha(X) &= \frac{1}{1-\alpha} \int_\alpha^1 \widehat{F}_n^{-1}(u) du \\ &= \frac{1}{1-\alpha} \left[\int_\alpha^{\frac{\lfloor n\alpha \rfloor + 1}{n}} \widehat{F}_n^{-1}(u) du + \sum_{j=\lfloor n\alpha \rfloor + 2}^n \int_{\frac{j-1}{n}}^{\frac{j}{n}} \widehat{F}_n^{-1}(u) dp \right] \\ &= \frac{1}{1-\alpha} \left[\left(\frac{\lfloor n\alpha \rfloor + 1}{n} - \alpha \right) \cdot X_{(\lfloor n\alpha \rfloor + 1)} + \sum_{j=\lfloor n\alpha \rfloor + 2}^n \frac{1}{n} \cdot X_{(j)} \right] \\ &= \frac{(\lfloor n\alpha \rfloor + 1) - n \cdot \alpha}{n(1-\alpha)} \cdot X_{(\lfloor n\alpha \rfloor + 1)} + \frac{1}{n(1-\alpha)} \cdot \sum_{j=\lfloor n\alpha \rfloor + 2}^n X_{(j)}. \end{aligned}$$

It gives the plug-in version of empirical estimator for ES as follows.

Proposition 2.6. (Plug-in version of empirical estimator for ES). *Define the empirical inverse function of F as in (2.10). Then it yields the empirical estimator for ES as*

$$\widehat{\text{ES}}_{\alpha}(X) = \frac{(\lfloor n\alpha \rfloor + 1) - n \cdot \alpha}{n(1 - \alpha)} \cdot X_{(\lfloor n\alpha \rfloor + 1)} + \frac{1}{n(1 - \alpha)} \cdot \sum_{j=\lfloor n\alpha \rfloor + 2}^n X_{(j)}. \quad (2.12)$$

If $\lfloor n\alpha \rfloor = n\alpha$, then (2.12) turns out to be

$$\widehat{\text{ES}}_{\alpha}(X) = \frac{1}{n(1 - \alpha)} \sum_{i=\lfloor n\alpha \rfloor + 1}^n X_{(i)} \quad (2.13)$$

where $\lfloor \cdot \rfloor$ is the floor function⁴.

It is clear that equations (2.12) and (2.13) are fairly close when the computation is made for large n . In vector format, the random sample $\mathbf{X} = (X_1, \dots, X_n)'$ is given along with $\mathbf{X}_{:n} = (X_{(1)}, \dots, X_{(n)})'$ as its ordered sample. Let $\mathbf{c} = (n(1 - \alpha))^{-1}(0, \dots, 0, 1, \dots, 1)'$ with zeros for the first $\lfloor n\alpha \rfloor$ elements. Then (2.13) can be re-written in the form

$$\widehat{\text{ES}}_{\alpha}(X) = \mathbf{c}' \mathbf{X}_{:n}. \quad (2.14)$$

The denominator, $n(1 - \alpha)$, was replaced by $n - \lfloor n\alpha \rfloor$ in Acerbi and Tasche (2002) in the sense of that ES is an average of the tail (excess losses).

For a random loss X with distribution function F , recall that

$$\text{CTE}_{\alpha}(X) = \text{E}[-X | X \leq -\text{VaR}_{\alpha}(X)],$$

and it can be re-written as

$$\frac{\int_A x dF(x)}{\int_A dF(x)}$$

where $A := \{\omega : X(\omega) \geq \text{VaR}_{\alpha}(X)\}$. For the case where X is continuous with strictly monotonic distribution function F , it is easy to show that $\text{CTE}_{\alpha}(X) = [\int_A x dF(x)] / (1 - \alpha)$ and $\text{CTE} = \text{ES}$. Then in this case, a simple Monte Carlo estimate of CTE also gives us a simple Monte Carlo estimate of ES by

$$\widehat{\text{ES}}_{\alpha} = \frac{1}{1 - \alpha} \left(\frac{1}{n} \sum_{x_i \in A} f(x_i) \right) \quad (2.15)$$

with $f(x) = x$ where x_i 's are realizations drawn from F . It follows that the so-called *HS estimate*

⁴The floor function is defined as $\lfloor x \rfloor = \max\{n \in \mathbb{Z} : n \leq x\}$ for $x \in \mathbb{R}$.

of ES amounts to the form given in (2.13) by drawing x_i 's from the empirical distribution function \widehat{F} and plugging into (2.15).

Bias Adjustment for HS Estimates

Recall the central idea of bias adjustment in (2.9). To implement bias adjustment for HS estimates of ES, first compute HS estimates of ES, $\widehat{ES}_\alpha^{\text{HS}}$, by (2.13) to replace \widehat{ES}_α in (2.9), and then compute the bootstrapping estimate of bias for $\widehat{ES}_\alpha^{\text{HS}}$, $\widehat{\text{bias}}(\widehat{ES}_\alpha^{\text{HS}})$, using variants of bootstrap techniques. Hence the bias-adjusted estimate of ES is given by

$$\widehat{ES}_\alpha^{\text{new}} = \widehat{ES}_\alpha^{\text{HS}} - \widehat{\text{bias}}(\widehat{ES}_\alpha^{\text{HS}}). \quad (2.16)$$

Kim (2007, pp.69) had provided that for i.i.d. samples, the true bias of the HS estimate of ES is negative and so is the EB estimate of bias, so that bias adjustment will work in the right direction.

Take the ordinary bootstrap as an example below to illustrate how to apply bias adjustment for HS estimate of ES. Applying the ordinary bootstrap to estimate ES at the confidence level α by simply substituting the statistic of interest T by the HS estimate of ES \widehat{ES}_α , we have the bias-adjusted estimate for ES_α as follows.

$$\widehat{ES}_\alpha^{\text{OB}} = \widehat{ES}_\alpha - \widehat{\text{bias}} = 2\widehat{ES}_\alpha - \widehat{ES}_\alpha^*, \quad (2.17)$$

where \widehat{ES}_α^* is the OB estimate of \widehat{ES}_α , and the bootstrapping estimate of bias, $\widehat{\text{bias}}$, equals $\widehat{ES}_\alpha^* - \widehat{ES}_\alpha$.

Furthermore, since the HS estimate of ES is an L -statistic, the EB bias-adjusted estimates of it can be obtained directly by applying Theorem 2.4. That is,

$$\widehat{ES}_\alpha^{\text{EB}} = \mathbf{c}'(2I - \mathbf{w}')\mathbf{X}_{:n}, \quad (2.18)$$

where $\mathbf{c} = (n(1 - \alpha))^{-1}(0, \dots, 0, 1, \dots, 1)'$ with zeros for the first $[n\alpha]$ elements. To adjust the HS estimate of ES for bias by EB, it is worth mentioning that for efficient computation, only $(n - [n\alpha])$ columns of the matrix \mathbf{w} from the $(n - [n\alpha] + 1)$ th column to the end need to be computed due to linear algebra of (2.18). Besides, for dependent data, the HS estimate of ES may better be improved by adjusting for bias using block-wise bootstrap. In other words, the bias of $\widehat{ES}_\alpha^{\text{HS}}$ in (2.16) is better to be estimated by block-wise bootstrap to obtain BB bias-adjusted estimate of ES for dependent data. Eventually, it is simple to compute the HS estimate and then bias adjustment can be implemented to improve the HS estimate by OB and EB.

2.3.2 For GPD Estimates

Another approach could be extended from extreme value theory (EVT) originated from Embrechts et al. (1997) for modelling the tails of heavy-tailed distributions. For heavy-tailed distributions, Generalized Pareto Distribution method is applicable to estimation of ES and will be described in this section.

Traditional extreme value theory investigates limiting distributions of normalized maxima $M_n := \max(X_1, \dots, X_n)$ of i.i.d. random variables and concludes that the only possible non-degenerate limiting distributions for M_n belong to the generalized extreme value (GEV) family.

Definition 2.7. (Generalized extreme value (GEV) distribution).

The distribution function of the (standard) GEV distribution is defined by

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

where $1 + \xi x > 0$. A three-parameter family is defined by $H_{\xi, \mu, \sigma}(x) := H_\xi((x - \mu)/\sigma)$ for a shape parameter, a location parameter $\mu \in \mathbb{R}$, and a scale parameter $\sigma > 0$.

H_ξ defines a type of distribution according to ξ and forms a family of distributions specified up to location and scaling.

Definition 2.8. (Maximum domain of attraction).

F is said to be in the maximum domain of attraction of H, written $F \in \text{MDA}(H)$, for some non-degenerate distribution function H, if there exist sequences of real numbers $\{d_n\}$ and $\{c_n\}$, such that

$$\lim_{n \rightarrow \infty} P\{(M_n - d_n)/c_n \leq x\} = \lim_{n \rightarrow \infty} F^n(c_n x + d_n) = H(x)$$

where $c_n > 0$ for all n .

Theorem 2.9. *H must be a distribution of type H_ξ if $F \in \text{MDA}(H)$ for some non-degenerate distribution function H. In other words, H is a GEV distribution.*

It has been shown that the distribution function F_ν of a standard univariate Student's t distribution with degree of freedom ν is in the maximum domain of attraction of H (i.e. $F_\nu \in \text{MDA}(H_\xi)$) for some $\xi \in \mathbb{R}$ and non-degenerate distribution function H . This implies that the distribution function of a Gaussian distribution is also in the maximum domain of attraction of H (i.e. $F_\nu \in \text{MDA}(H_\xi)$) for some $\xi \in \mathbb{R}$ and non-degenerate distribution function H , since the limiting distribution Student's t distribution is Gaussian as ν goes to ∞ .

The generalized Pareto Distribution (GPD) plays an important role in extreme value theory as a natural model for the excess distribution over a high threshold. Precisely, GPD is the *canonical distribution* for modelling exceedances over a high threshold.

Definition 2.10. (Generalized Pareto Distribution (GPD)). *The distribution function (df) of the GPD is defined as*

$$G_{\xi,\beta}(x) = \begin{cases} 1 - (1 + \xi/\beta)^{-1/\xi}, & \text{if } \xi \neq 0 \\ 1 - \exp(-x/\beta), & \text{if } \xi = 0 \end{cases} \quad (2.19)$$

where $\beta > 0$, and $x \geq 0$ when $\xi \geq 0$ and $0 \leq x \leq -\beta/\xi$ when $\xi < 0$.

The parameter ξ and β are referred to as the *shape* and *scale* parameters, respectively. Define the excess distribution function F_u corresponding to F as

$$F_u(x) = P(X - u \leq x | X > u) = \frac{F(x+u) - F(u)}{1 - F(u)} \quad (2.20)$$

for $0 \leq x < x_F - u$, where $x_F \leq \infty$ is the right endpoint of F , defined by $\sup\{x \in \mathbb{R} : F(x) < 1\}$. If a random variable has distribution function $F = G_{\xi,\beta}$, then the excess distribution function $F_u(x)$ over the threshold u is of the form

$$F_u(x) = G_{\xi,\beta(u)}(x), \quad \beta(u) = \beta + \xi u, \quad (2.21)$$

where $0 \leq x < \infty$ if $\xi \geq 0$ and $0 \leq x \leq -(\beta/\xi) - u$ if $\xi < 0$.

Theorem 2.11. (Pickands (1975) and Balkema and de Haan (1974))

A distribution function $F \in \text{MDA}(H_\xi)$, $\xi \in \mathbb{R}$ if and only if there exists a positive-measurable function $\beta(u)$ such that the corresponding excess distribution function F_u

$$\lim_{u \rightarrow x_F} \sup_{0 \leq x < x_F - u} |F_u(x) - G_{\xi,\beta(u)}(x)| = 0.$$

Theorem 2.11 gives a mathematical result that essentially shows that GPD is a natural limiting excess distribution for a large class of underlying loss distributions. Note that ideally the excess distribution is GPD but in practice it will not exactly be GPD.

To exploit Theorem 2.11, assume that $F \in \text{MDA}(H_\xi)$ so as to model F_u by a GPD for some suitably chosen high threshold u . That is, suppose that the loss distribution F has endpoint x_F and for some high threshold u , $F_u(x) = G_{\xi,\beta(u)}(x)$ for $0 \leq x \leq -(\beta/\xi) - u$ for some $\xi \in \mathbb{R}$ and $\beta > 0$. Define the tail probability by $\bar{F}(x) = 1 - F(x)$. Observe that for $x \geq u$,

$$\begin{aligned} \bar{F}(x) &= P(X > u)P(X > x | X > u) \\ &= \bar{F}(u)P(X - u > x - u | X > u) \\ &= \bar{F}(u)\bar{F}_u(x - u) \\ &= \bar{F}(u)\left(1 + \xi \frac{x - u}{\beta}\right)^{-1/\xi}. \end{aligned} \quad (2.22)$$

Solving $\alpha = 1 - \overline{F}(x)$ via (2.22) for x yields a high quantile for the underlying distribution, which can be interpreted as a VaR if F is a continuous and strictly monotonic distribution function. Hence for $\alpha \geq F(u)$, the analytical expression of VaR is given by

$$\text{VaR}_\alpha(X) = u + \frac{\beta}{\xi} \left[\left(\frac{1 - \alpha}{\overline{F}(x)} \right)^{-\xi} - 1 \right]. \quad (2.23)$$

Assuming that $\xi < 1$, the analytical expression of ES is given by

$$\text{ES}_\alpha(X) = \frac{\text{VaR}_\alpha(X)}{1 - \xi} + \frac{\beta - \xi u}{1 - \xi}. \quad (2.24)$$

The following describes the practical process to estimate the tail, and hence VaR and ES, using the GPD model. Suppose that the data X_1, \dots, X_n are an observed sequence of losses with distribution function $F(x) = P(X \leq x)$. Let N_u denote the random number of X_i 's that exceed the threshold u and relabel these data as $\tilde{X}_1, \dots, \tilde{X}_{N_u}$. For each exceedance, calculate the amount $Y_j = \tilde{X}_j - u$ of the the excess loss over the threshold u . It follows that writing $g_{\xi, \beta}$ as the density of the GPD, the log-likelihood may be easily calculated to be

$$\begin{aligned} \ln L(\xi, \beta; Y_1, \dots, Y_{N_u}) &= \sum_{j=1}^{N_u} g_{\xi, \beta}(Y_j) \\ &= -N_u \ln \beta - \left(1 + \frac{1}{\xi}\right) \sum_{j=1}^{N_u} \ln\left(1 + \xi \frac{Y_j}{\beta}\right). \end{aligned} \quad (2.25)$$

Maximizing (2.25) subject to the parameter constrains that $\beta > 0$ and $1 + \xi X_j/\beta > 0$ for all j yields MLE of ξ and β , $\hat{\xi}$ and $\hat{\beta}$, respectively. Furthermore, Smith (1987) first proposed an estimator for tail probabilities $\overline{F}(x)$, which is given by

$$\widehat{\overline{F}}(x) = \frac{N_u}{n} \left(1 + \hat{\xi} \frac{x - u}{\hat{\beta}} \right)^{-1/\hat{\xi}} \quad (2.26)$$

Therefore, by fixing the number of data in the tail to be $N_u \ll n$, VaR and ES can be estimated in turn by

$$\text{VaR}_\alpha(X) = X_{(N_u+1)} + \frac{\hat{\beta}}{\hat{\xi}} \left[\left(\frac{n}{N_u} (1 - \alpha) \right)^{-\hat{\xi}} - 1 \right], \quad (2.27)$$

where $X_{(N_u+1)}$ is the $N_u + 1$ -th order statistic as the effectively random threshold, and

$$\text{ES}_\alpha(X) = \frac{\text{VaR}_\alpha(X)}{1 - \hat{\xi}} + \frac{\hat{\beta} - \hat{\xi} u}{1 - \hat{\xi}}. \quad (2.28)$$

The estimate of ES in (2.28) is then called the GPD estimate of ES.

Write $X_{(1)} \leq \dots \leq X_{(n)}$ for the order statistics of the data. The Hill method gives an alternative estimator of the tail $\bar{F}(x)$ (Hill, 1975), given by

$$\widehat{\bar{F}}(x) = \frac{N_u}{n} \left(1 + \widehat{\xi} \frac{x - u}{\widehat{\xi} u} \right)^{-1/\widehat{\xi}}, \quad (2.29)$$

where

$$\widehat{\xi} = \frac{1}{k} \left(\sum_{j=1}^k \log X_{(j)} - \log X_{(k)} \right) \quad (2.30)$$

for sufficiently small k . McNeil and Frey (2000) suggested that the GPD method was preferable since the estimator in (2.29) lacks the additional scale parameter β in (2.26) and tends to not perform well in simulated examples therein. Therefore, we will not consider the Hill method but the GPD method for the following research.

Bias Adjustment for GPD Estimates

Again recall the central idea of bias adjustment in (2.9). To implement bias adjustment for GPD estimates of ES, first compute GPD estimate of ES, $\widehat{\text{ES}}_{\alpha}^{\text{GPD}}$, by (2.28) to replace $\widehat{\text{ES}}_{\alpha}$ in (2.9), and then compute the bootstrapping estimate of bias, $\widehat{\text{bias}}(\widehat{\text{ES}}_{\alpha}^{\text{GPD}})$, for $\widehat{\text{ES}}_{\alpha}^{\text{GPD}}$ using variants of bootstrap techniques. Hence the bias-adjusted estimate of ES is given by

$$\widehat{\text{ES}}_{\alpha}^{\text{new}} = \widehat{\text{ES}}_{\alpha}^{\text{GPD}} - \widehat{\text{bias}}(\widehat{\text{ES}}_{\alpha}^{\text{GPD}}).$$

Nevertheless, all bootstrap techniques except EB method, described in Section 2.2, can be applied in suitable situations to adjust GPD estimate of ES for bias, since GPD estimate is not an L -statistic. In addition, GPD method can be applied to Student's t case since a Student's t distribution with a small degree of freedom has heavy tail.

2.3.3 For MLE From Elliptical Distributions

There has been an abundance of literature so far that provided explicit and analytical formulas of ES and ESC in closed forms for a variety of common models (see for more reference in Section 1.2.1 and 1.2.2). In the midst of them, Dhaene et al. (2007) synthesized some results for elliptical distributions encompassing Gaussian and Student's t distributions.

Definition 2.12. (Multivariate elliptical distributions). Consider the vector $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)'$ and the positive semi-definite matrix Σ with elements σ_{ij} ($i, j = 1, 2, \dots, d$). The random vector $\mathbf{X} = (X_1, \dots, X_d)'$ is said to have an elliptical distribution with parameters $\boldsymbol{\mu}$ and Σ if

its characteristic function $E[\exp(it'\mathbf{X})]$ can be expressed as

$$\psi_{\mathbf{X}}(\mathbf{t}) = E[\exp(it'\mathbf{X})] = \exp(it'\boldsymbol{\mu})\phi\left(\frac{1}{2}\mathbf{t}'\Sigma\mathbf{t}\right), \quad (2.31)$$

for some scalar function ϕ , where $\mathbf{t}' = (t_1, t_2, \dots, t_d)$.

Denote the elliptical distribution by $\mathbf{X} \sim E_d^*(\boldsymbol{\mu}, \Sigma, \phi)$ and refer to $\boldsymbol{\mu}$ as the location vector, Σ as the dispersion matrix and ϕ as the *characteristic generator* of the distribution. Note that the moments of $\mathbf{X} \sim E_d^*(\boldsymbol{\mu}, \Sigma, \phi)$ do not necessarily exist and that not every multivariate elliptical distribution has a probability density function (pdf). Theorem 2.13 is well-known for providing necessary and sufficient conditions for the existence of an elliptical density.

Theorem 2.13. (Elliptical densities; Dhaene et al. (2007)). *Suppose that $\mathbf{X} \sim E_d^*(\boldsymbol{\mu}, \Sigma, \phi)$ and $\Gamma(\cdot)$ is the Gamma function. Then it has a density if and only if the pdf of \mathbf{X} is given by*

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{c_d}{\sqrt{|\Sigma|}} g_d\left(\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right) \quad (2.32)$$

for some non-negative function g_d satisfying the condition

$$0 < \int_0^\infty u^{d/2-1} g_d(u) du < \infty, \quad (2.33)$$

a normalizing constant c_d given by

$$c_d = \frac{\Gamma(d/2)}{(2\pi)^{d/2}} \left[\int_0^\infty u^{d/2-1} g_d(u) du \right]^{-1}, \quad (2.34)$$

and positive definite Σ . If $\mathbf{X} \sim E_d^*(\boldsymbol{\mu}, \Sigma, \phi)$ has a density, the notation is given by $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, g_d)$ and g_d is the so-called density generator.

The following properties of elliptical distributions can be found in Landsman and Valdez (2003). Suppose that $\mathbf{X} \sim E_d^*(\boldsymbol{\mu}, \Sigma, \phi)$ where the (i, j) -th component of Σ is denoted by σ_{ij} . Then the properties below hold well under appropriate conditions.

1. *Marginal distributions.* Each component X_k of $\mathbf{X} \sim E_d^*(\boldsymbol{\mu}, \Sigma, \phi)$ has one-dimensional elliptical distribution with the same characteristic generator ϕ and the same density generator g_1 :

$$X_k \sim E_1(\mu_k, \sigma_k^2, \phi), \quad k = 1, \dots, d, \quad (2.35)$$

where $\sigma_k^2 = \sigma_{kk}$. In particular, $X_k \sim E_1(\mu_k, \sigma_k^2, g_1)$ for $k = 1, 2, \dots, d$.

2. *Linear combinations.* Linear combinations of elliptical random vectors remain elliptical

with the same characteristic generator ϕ . Let $B \in \mathbb{R}^{k \times d}$ and $\mathbf{b} \in \mathbb{R}^d$. Then

$$B\mathbf{X} + \mathbf{b} \sim E_k^*(B\boldsymbol{\mu} + \mathbf{b}, B\Sigma B', \phi).$$

As a special case, if $\mathbf{a} \in \mathbb{R}^d$, then

$$\mathbf{a}'\mathbf{X} \sim E_1^*(\mathbf{a}'\boldsymbol{\mu}, \mathbf{a}'\Sigma\mathbf{a}, \phi)$$

3. *Sum of random variables (Special case of linear combinations).* The random variable S defined by $S = \sum_{j=1}^d X_j = \mathbf{e}'\mathbf{X}$, where $\mathbf{e} = (1, \dots, 1)'$, is elliptically distributed with the same characteristic generator:

$$S \sim E_1^*(\mu_S, \sigma_S^2, \phi), \quad (2.36)$$

with parameters given by

$$\mu_S = \mathbf{e}'\boldsymbol{\mu} = \sum_{j=1}^d \mu_j \quad \text{and} \quad \sigma_S^2 = \mathbf{e}'\Sigma\mathbf{e} = \sum_{j=1}^d \sum_{k=1}^d \sigma_{jk},$$

respectively.

4. (a) It has been shown that the condition

$$\int_0^\infty g_1(u)du < \infty \quad (2.37)$$

guarantees the existence of the mean and hence the mean vector for $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, g_d)$ is $E(\mathbf{X}) = \boldsymbol{\mu}$.

- (b) Suppose that the first partial derivative of function ϕ with respect to t exists and let $\phi^{(1)}(t_0)$ denotes $\frac{\partial \phi(t)}{\partial t}|_{t=t_0}$ which is the first partial derivative of function ϕ with respect to t evaluated at $t = t_0$. If in addition the following condition

$$|\phi^{(1)}(0)| < \infty, \quad (2.38)$$

holds, the covariance matrix exists and equals (Cambanis et al., 1981)

$$\text{Cov}(\mathbf{X}) = -\phi^{(1)}(0)\Sigma. \quad (2.39)$$

- (c) Then the characteristic generator can be chosen such that

$$\phi^{(1)}(0) = -1 \quad (2.40)$$

so that the covariance matrix above becomes

$$\text{Cov}(\mathbf{X}) = \Sigma. \quad (2.41)$$

(d) Note that condition (2.38) is equivalent to the condition (2.37).

5. In the case of $d = 1$, write $X \sim E^*(\mu, \sigma^2, \phi)$, or $X \sim E(\mu, \sigma^2, g)$ if the condition (2.33) in Theorem 2.13 holds, and the standardized random variable $Z = (X - \mu)/\sigma$ has a standard elliptical (oftentimes called spherical) distribution function

$$F_z(z) = c_1 \int_{-\infty}^z g_1\left(\frac{1}{2}u^2\right) du. \quad (2.42)$$

with mean 0 and variance $\sigma_Z^2 = 2c_1 \int_0^\infty u^2 g\left(\frac{1}{2}u^2\right) du = -\phi^{(1)}(0)$. In addition, if the characteristic generator is chosen such that condition (2.40) holds, then $\sigma_Z^2 = 1$.

For convenience, the notation d for dimension would be dropped for univariate cases hereafter.

Multivariate Gaussian and Student's t distributions are two important subclasses of elliptical distributions, as described in the following two examples.

Example 2.14. (Multivariate Gaussian distribution). Consider the vector $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)'$ and the positive semi-definite matrix Σ . The d -dimensional random vector $\mathbf{X} = (X_1, \dots, X_d)'$ is said to have a multivariate Gaussian distribution with parameters $\boldsymbol{\mu}$ and Σ , denoted by $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$, if its characteristic function $E[\exp(it' \mathbf{X})]$ is given by

$$\psi_{\mathbf{X}}(\mathbf{t}) = E[\exp(it' \mathbf{X})] = \exp(it' \boldsymbol{\mu}) \exp\left(-\frac{1}{2} \mathbf{t}' \Sigma \mathbf{t}\right). \quad (2.43)$$

It is easy to verify that $N_d(\boldsymbol{\mu}, \Sigma)$ is an elliptical distribution with characteristic generator given by

$$\phi(t) = e^{-t}. \quad (2.44)$$

Furthermore, the density generator g_d and the normalizing constant c_d of $N_d(\boldsymbol{\mu}, \Sigma)$ are given by

$$g_d(u) = e^{-u} \quad (2.45)$$

and

$$c_d = \frac{1}{(2\pi)^{d/2}}, \quad (2.46)$$

respectively. Since the characteristic generator chosen in (2.44) satisfies (2.40), the covariance matrix $\text{Cov}(X)$ amounts to the second parameter Σ of definition.

Example 2.15. (Multivariate Student's t distribution). The random vector \mathbf{X} is said to have a Student's- t distribution, denoted by $\mathbf{X} \sim t_d(\boldsymbol{\mu}, \Sigma; r)$ with degree of freedom $r > 0$, if its

density generator can be expressed as

$$g_d(u) = \left(1 + \frac{2u}{r}\right)^{-(r+d)/2}. \quad (2.47)$$

Its joint density could be simplified in the form below

$$f(\mathbf{x}) = \frac{\Gamma(\frac{1}{2}(r+d))}{\Gamma(\frac{1}{2}r)(\pi r)^{d/2}|\Sigma|^{1/2}} \left(1 + \frac{(\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})}{\nu}\right)^{-(r+d)/2},$$

where $\Gamma(\cdot)$ is the Gamma function. Furthermore, by (2.34) it can be shown that the normalizing constant is

$$c_d = \frac{\Gamma((r+d)/2)}{\Gamma(r/2)} (\pi r)^{-d/2}. \quad (2.48)$$

Note that $\text{Cov}(\mathbf{X}) = (\nu/(\nu-2))\Sigma$ for $\nu > 2$; that is, Σ is not exactly the covariance matrix of \mathbf{X} in the above definition of the multivariate Student's t distribution. Please refer to Sutradhar (1986) and Hurst (1995) for explicit forms of characteristic functions of multivariate and univariate Student's t distributions. For the univariate case where $d = 1$, it is the univariate Student's t distribution with density

$$\begin{aligned} f_X(x) &= \frac{c_1}{\sigma} g_1 \left(\frac{1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right) \\ &= \frac{1}{\sigma \sqrt{r} B(1/2, r/2)} \left[1 + \frac{(x - \mu)^2 / \sigma^2}{r} \right]^{-(r+d)/2}, \end{aligned}$$

where $c_1 = [\sqrt{r} B(1/2, r/2)]^{-1}$ with the beta function $B(\cdot, \cdot)$ defined by $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ and $g_1(u) = (1 + 2u/r)^{-(r+d)/2}$.

Formulas of ES for Elliptical Distributions

Landsman and Valdez (2003) employed the cumulative generator defined in Definition 2.16 to derive ES formulas for elliptical distributions.

Definition 2.16. (Cumulative generator; Landsman and Valdez (2003)). Suppose that $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, g_d)$. The cumulative generator is the function given by

$$G(x) = c_1 \int_0^x g_1(u) du. \quad (2.49)$$

Now the λ -factor, $\lambda_{g,\alpha}$, is to be first introduced and defined below.

Definition 2.17. Consider a univariate elliptical distribution denoted by $E(\mu, \sigma^2, g)$, where g is the density generator, providing that it exists. Let $\bar{G}(x) = G(\infty) - G(x)$ where $G(\cdot)$ is the

cumulative generator defined by (2.49). Then the λ -factor $\lambda_{g,\alpha}$ is defined by

$$\lambda_{g,\alpha} = \frac{\overline{G}(\frac{1}{2}z_\alpha^2)}{1 - \alpha}, \quad (2.50)$$

where $z_\alpha = (\text{VaR}_\alpha(Z) - \mu)/\sigma$ or equivalently $z_\alpha = \text{VaR}_\alpha(Z)$ for $Z \sim E(0, 1, g)$.

The λ -factor plays an essential role for ES formulas as demonstrated in the following and in fact, it is the ES for standardized elliptical distributions $E(0, 1, g)$. The general formulas of ES for the class of elliptical distributions is adapted in Theorem 2.18 below, which is originated from CTE formulas given in Landsman and Valdez (2003), since ES and CTE coincide for elliptical distributions.

Theorem 2.18. (Formulas of ES for elliptical distributions). *Suppose that $X \sim E(\mu, \sigma^2, g)$. Then under condition (2.37), we have ES of X*

$$\text{ES}_\alpha(X) = \mu + \sigma \cdot \lambda_{g,\alpha} \quad (2.51)$$

where $\lambda_{g,\alpha}$ is defined in Definition 2.17.

The proof of Theorem 2.18 is given in Appendix C. First observe that Equation (2.51) holds well even if $\text{Var}(X)$ does not exist, since σ is merely a parameter of an elliptical distribution and not necessarily equal to $\text{Var}(X)$. Therefore the condition for existence of $\text{Var}(X)$ is not necessary for Equation (2.51). Noticeably the enhanced formula of ES at level α given in Theorem 2.18 depends on the same $\lambda_{g,\alpha}$ for any elliptical distribution with the same density generator g because the value of $\lambda_{g,\alpha}$ varies solely with the confidence level α and the density generator g . Hence the ES formula is known and simple as a linear function of parameters, μ and σ , given the density generator g and the confidence level α . This is the essence of Theorem 2.18. In other words, for a subclass of elliptical distributions with the same density generator, the value of ES at a specified confidence level α does not vary with $\lambda_{g,\alpha}$ but only with those two parameters, μ and σ . By exploiting Theorem 2.18, the value of $\lambda_{g,\alpha}$ has to be calculated only once as the confidence level α and the density generator g is given and then the value of ES is very easy to be computed, once the two parameters have been estimated.

Compare formula (2.51) above with the original formula given by Landsman and Valdez (2003) as in the following form

$$\text{ES}_\alpha(X) = \mu + \sigma^2 \cdot \lambda \quad (2.52)$$

where

$$\lambda = \frac{\overline{G}(\frac{1}{2}z_\alpha^2)}{\sigma(1 - F_X(\text{VaR}_\alpha(X)))} \quad (2.53)$$

with the cdf of X , F_X and

$$z_\alpha = \frac{\text{VaR}_\alpha(X) - \mu}{\sigma}.$$

Evidently, the value of λ given in the original version (2.53) depends not only on g and α but also on σ , while $\lambda_{g,\alpha}$ in Theorem 2.18 only varies with g and α . Precisely, given the assumption of distribution and confidence level α , $\lambda_{g,\alpha}$ is known and fixed but λ is not. Besides, for elliptical distributions where X is certainly a continuous random variable, $1 - F_X(\text{VaR}_\alpha(X))$ is merely $1 - \alpha$ and $z_\alpha = \text{VaR}_\alpha(Z)$ is determined only by g and α as is $\lambda_{g,\alpha}$ in Theorem 2.18, where $Z \sim E(0, 1, g)$ is the standardized random variable. Hence additional calculation of $\text{VaR}_\alpha(X)$ is not necessary in the new formula. It is quite clear that the enhanced version of ES formula given in Theorem 2.18 herein is more appealing and powerful than the original one.

For simplicity, $\lambda_{g,\alpha}$ will be denoted by λ for the same g and α , or λ_α when it varies only with α for a particular subclass of elliptical distributions such as Gaussian. Taking Gaussian distributions as an example, they share the same value of λ for a particular confidence level α since their generators are all the same as g . Therefore, it is worthwhile to have explicit and analytic forms of λ for certain elliptical distributions subsuming Gaussian and t , and λ can be interpreted as the ES for standardized elliptical distribution $E(0, 1, g)$. The simplified forms of λ for Gaussian and Student's t distributions are summarized as follows.

1. *Gaussian distribution.*

$$\lambda_\alpha = \frac{\varphi(\Phi^{-1}(\alpha))}{1 - \alpha} \quad (2.54)$$

where $\varphi(\cdot)$ and $\Phi(\cdot)$ denote respectively the density and distribution functions of a standard Gaussian distribution.

2. *Student's t distribution.* Recall that for Students' t distributions, σ in (2.51) may not be the standard deviation (square root of variance) but definitely the parameter in the definition of elliptical distributions. For the case where $r = 1$, it is the Cauchy distribution, for which the mean does not exist so that CTE does not exist and neither does ES. ES does not exist for $r < 1$ either because the integral in its definition diverges. Then considering $r > 1$, we have

$$\lambda_{r,\alpha} = \frac{\bar{G}\left(\frac{1}{2}z_\alpha^2\right)}{1 - \alpha}, \quad (2.55)$$

where

$$\begin{aligned} \bar{G}\left(\frac{1}{2}z_\alpha^2\right) &= c_1 \frac{r}{r-1} \left(1 + \frac{z_\alpha^2}{r}\right)^{-(r-1)/2} \\ &= \frac{\sqrt{r}}{(r-1)B\left(\frac{1}{2}, \frac{r}{2}\right)} \left(1 + \frac{z_\alpha^2}{r}\right)^{-(r-1)/2} \end{aligned}$$

with $B(\cdot, \cdot)$ is the beta function⁵. Furthermore, $\overline{G}(\frac{1}{2}z_\alpha^2)$ can be rewritten for $r > 2$ in the simple form

$$\begin{aligned} & \frac{c_1}{c^*} \frac{r}{r-1} f_Z \left(\sqrt{\frac{r-2}{r}} z_\alpha; r-2 \right) \\ &= \sqrt{\frac{r}{r-2}} \cdot f_Z \left(\sqrt{\frac{r-2}{r}} z_\alpha; r-2 \right) \end{aligned} \quad (2.56)$$

where $f_Z(z; m)$ is the density function of the standard t , $Z \sim t(0, 1; m)$.

In sum, providing that the density generator g exists, Theorem 2.18 not only provides a shortcut to compute ES for elliptical distributions with known parameters, but also gives a convenient way to estimate ES, since only parameters μ and σ are left to be estimated, and $\lambda_{g,\alpha}$ is already determined by the density generator g and the confidence level α . The well-known ES formula for Gaussian distributions could also be found in McNeil et al. (2005, Example 2.18), and the ES formula for t distributions in in McNeil et al. (2005, Example 2.19). Although the ES formula for a standard t distribution in McNeil et al. (2005), given below, looks different from either (2.55) or (2.56), their values amount to the same thing from numerical experiments.

$$\frac{f_r(F_r^{-1}(\alpha))}{1-\alpha} \cdot \frac{r + (F_r^{-1}(\alpha))^2}{r-1}$$

where f_r and F_r are the pdf and cdf of $t(0, 1; r)$, respectively.

Bias Adjustment

Consider an i.i.d. random sample from an elliptical distribution. Then λ is already determined once the model is assumed as well as the confidence level α specified and hence there is no need to estimate λ but to calculate it. Therefore the parametric estimation problem of ES pulls back to be the ordinary estimation problem of model parameters, in the sense that once the model parameters are estimated then the estimate of ES can be yielded immediately through Equation (2.51) in Theorem 2.18 by substituting corresponding estimates for parameters. It can then be foreseen that the performance of parametric estimators of ES for elliptical distributions will be mostly driven by the performance of estimated parameters from model fitting, since ES of elliptical distributions can be explicitly expressed by a linear function of parameters of the underlying distribution.

As a benchmark, MLE of ES for elliptical distributions can be easily calculated by plugging in MLE of parameters into formula (2.51), with λ given in (2.54) for Gaussian distributions or

⁵ $B(a, b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$.

λ given in (2.55) for t distributions.

2.4 Bias Adjustment for Conditional ES Estimates of Time Series

For time series, the concept of conditional ES is generally adopted as the ES conditional on the past. In general, suppose that a time series $\{X_t\}_{t \in \mathbb{Z}}$ follows a particular model with a parameter vector $\boldsymbol{\theta}$ of all involved parameters, and innovations which are a series of white noise (WN) process $\{\varepsilon_t\}_{t \in \mathbb{Z}}$. The goal is to predict the conditional ES one-step ahead based on available data up to the current time point. It is usually able to ideally assume that the conditional ES of X_{T+1} at level α , $\text{ES}_\alpha(X_{T+1}|\mathcal{F}_T)$, can be derived as a function h of the available data ($\{X_t\}_{t=1,\dots,T}$), parameters, and the ES of innovations. That is,

$$\text{ES}_\alpha(X_{T+1}|\mathcal{F}_T) = h(\boldsymbol{\theta}, \{X_t\}_{t=1,\dots,T}, \text{ES}_\alpha(\varepsilon_t)), \quad (2.57)$$

where $\boldsymbol{\theta}$ denotes the parameter vector comprising all parameters of the time series model including parameters of the distribution of ε_t , and $\text{ES}_\alpha(\varepsilon_t)$ represents the ES of its innovations. Let

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{pmatrix}$$

where $\boldsymbol{\theta}_1$ represents the parameter vector comprising all parameters of the time series model excluding parameters of the distribution of ε_t and $\boldsymbol{\theta}_2$ the remaining parameters which are those of distributions of $\{\varepsilon_t\}_{t \in \mathbb{Z}}$. Alternatively, (2.57)

$$\text{ES}_\alpha(X_{T+1}|\mathcal{F}_T) = h(\boldsymbol{\theta}_1, \{X_t\}_{t=1,\dots,T}, \text{ES}_\alpha(\varepsilon_t)). \quad (2.58)$$

Then employing the estimation strategy proposed above, it will be easy to estimate $\text{ES}_\alpha(X_{T+1}|\mathcal{F}_T)$ through (2.57) by plugging in the data values $\{X_t\}_{t=1,\dots,T}$ and substituting estimates of $\boldsymbol{\theta}$. Similarly, $\text{ES}_\alpha(X_{T+1}|\mathcal{F}_T)$ can also be estimated through (2.58) by plugging in the data values $\{X_t\}_{t=1,\dots,T}$ and substituting estimates of $\boldsymbol{\theta}_1$ and $\text{ES}_\alpha(\varepsilon_t)$. In general, the value of ES estimated by (2.57) and (2.58) should be very close if the postulated model is correct. If the distribution of the innovation is known, then either (2.57) or (2.58) is appropriate to use; otherwise (2.58) is preferred.

Note that maximum likelihood estimators (MLE) are usually preferred as a benchmark in Step 2 above, because in Step 3, the proposed scheme yields an estimate of ES, a function of MLE of model parameters. Then by the invariance property, this estimate of ES is also an MLE and hence will preserve nice properties of MLE. Besides MLE of ES, the following sections will

discuss in detail how to apply bias adjustment to improve HS or GPD estimates of it under particular models.

2.4.1 Conditional ES of ARMA Processes

Let $\text{WN}(0, \sigma^2)$ represent a white noise (WN) process centred with mean zero and variance σ . Define the backshift operator or lag operator B by $B^k X_t = X_{t-k}$. Now suppose that the time series model is an ARMA process in Definition 2.19.

Definition 2.19. (ARMA processes). *Let $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ be $\text{WN}(0, \sigma_\varepsilon^2)$. The time series $\{X_t\}_{t \in \mathbb{Z}}$ is a zero-mean ARMA(p, q) process if it is a covariance stationary process such that*

$$\phi(B)X_t = \psi(B)\varepsilon_t, \quad \forall t \in \mathbb{Z} \quad (2.59)$$

where $\phi(B) = 1 - \phi_1 B^1 - \dots - \phi_p B^p$ and $\psi(B) = 1 + \psi_1 B^1 + \dots + \psi_q B^q$. Furthermore, $\{X_t\}_{t \in \mathbb{Z}}$ is an ARMA process with mean μ if the centered series $\{X_t - \mu\}_{t \in \mathbb{Z}}$ is a zero-mean ARMA(p, q) process.

It follows that for ARMA processes, the function $h(\cdot, \cdot, \cdot)$ in Equation (2.57) is determined by

$$h(\{X_t\}_{t=1, \dots, T}, \boldsymbol{\theta}, \text{ES}_\alpha(\varepsilon)) = \mu + \sum_{i=1}^p \phi_i (X_{T+1-i} - \mu) + \sum_{j=1}^q \psi_j \varepsilon_{T+1-j} + \text{ES}_\alpha(\varepsilon_t) \quad (2.60)$$

such that $\text{ES}_\alpha(X_{T+1} | \mathcal{F}_T) = h(\{X_t\}_{t=1, \dots, T}, \boldsymbol{\theta}, \text{ES}_\alpha(\varepsilon_t))$. Let $\mu_{T+1} = \sum_{i=1}^p \phi_i (X_{T+1-i} - \mu) + \sum_{j=1}^q \psi_j \varepsilon_{T+1-j}$ to be a function of $\{X_t\}_{t=1, \dots, T}$ and $\boldsymbol{\theta}$. Thus (2.60) can be simplified into

$$h(\mu_{T+1}, \text{ES}_\alpha(\varepsilon)) = \mu_{T+1} + \text{ES}_\alpha(\varepsilon_t). \quad (2.61)$$

Prediction of Conditional ES for ARMA Processes

Now by observing either Equation (2.60) or (2.61), prediction of conditional ES one step ahead turns out to be estimation of the quantity $\mu_{T+1} + \text{ES}_\alpha(\varepsilon_t)$ for ARMA processes. The estimation of μ_{T+1} can be obtained from the model fitting.

On the parametric side, the estimation of $\text{ES}_\alpha(\varepsilon_t)$ depends on the assumption of the distribution of $\{\varepsilon_t\}_{t \in \mathbb{Z}}$. The general idea is to derive the formula of ES for the innovations under the specified distribution conditional on the past and then to plug the estimated parameters into Equation (2.61) to obtain the estimate of $\text{ES}_\alpha(\varepsilon_t)$. With regard to the assumption of distribution, Gaussian or (scaled) Student's t distributions with mean zero and variance σ_ε^2 are considered as candidates. The estimation methods for i.i.d. elliptical distributions, discussed

in Section 2.3.3, are immediately applicable since both Gaussian and Student's t are elliptical distributions. Recall that formulas of ES for elliptical distributions from Equation (2.51) is

$$\text{ES}_\alpha(\varepsilon_t) = \mu + \sigma \cdot \lambda_{g,\alpha} \quad (2.62)$$

where

$$\lambda_{g,\alpha} = \frac{\overline{G}(\frac{1}{2}z_\alpha^2)}{1 - \alpha} \quad (2.63)$$

with $z_\alpha = \text{VaR}_\alpha(Z)$ for $Z \sim E(0, 1, g)$. Therefore, for $N(0, 1)$, simply put $\mu = 0$ and $\sigma = \sigma_\varepsilon$ in Equation (2.62), and then $\text{ES}_\alpha(\varepsilon_t)$ can be calculated through

$$\text{ES}_\alpha(\varepsilon_t) = \sigma_\varepsilon \cdot \lambda_\alpha \quad \text{for } N(0, 1), \quad (2.64)$$

where λ is determined by Equation (2.54). Parallel to Gaussian distributions, put $\mu = 0$ and $\sigma = \sqrt{(r-2)/r}\sigma_\varepsilon$ in Equation (2.62) for $t(0, \sigma^2; r)$ with degree of freedom $r > 2$ and thus get

$$\text{ES}_\alpha(\varepsilon_t) = \sigma_\varepsilon \lambda_{r,\alpha} \sqrt{(r-2)/r} \quad (2.65)$$

where $\lambda_{r,\alpha}$ is determined by (2.55). Again in this case, variance of ε_t is not σ^2 itself but $\sigma_\varepsilon^2 = r\sigma^2/(r-2)$. Finally substitute estimates of σ in (2.62) and obtain an estimate of $\text{ES}_\alpha(\varepsilon_t)$.

On the non-parametric side of estimating $\text{ES}_\alpha(\varepsilon_t)$, the sequence of model residuals plays a major and important role, because it could be thought of as a sequence of realizations of the unobserved i.i.d. sequence $\{\varepsilon_t\}_{t \in \mathbb{Z}}$. Hence the HS estimate of $\text{ES}_\alpha(\varepsilon_t)$ could be adopted straightforwardly by computing the HS estimate of ES for the residuals treated as the underlying data at this moment. Therefore, bias adjustment by OB, EB, or IB in Chapter 2 can be applied as well to improve the HS estimate of ES for the residuals and gives hope to improve the estimate of ES for the ARMA process. In addition, if the sequence $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ has heavy-tailed distribution, then GPD method is applicable to estimate $\text{ES}_\alpha(\varepsilon_t)$ as well. Hence bias-adjusted GPD estimates of $\text{ES}_\alpha(\varepsilon_t)$ by OB or IB techniques may also gives hope to further improve the estimate of $\text{ES}_\alpha(X_{t+1}|\mathcal{F}_t)$ for the original data.

The following sections will describe the same idea in (2.57) for GARCH and ARMA-GARCH in detail since they are popular models of changing volatility for daily risk-factor return series.

2.4.2 Conditional ES of GARCH Processes

Generalized autoregressive conditional heteroscedastic (GARCH) processes are prominent models for daily risk-factor return series.

Definition 2.20. (GARCH processes). Let $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ be strict white noise (SWN) with mean zero and variance one. The time series $\{X_t\}_{t \in \mathbb{Z}}$ is a GARCH(p, q) process if it is strictly

stationary and if it satisfies, for all $t \in \mathbb{Z}$ and some strictly positive-valued process $\{\sigma_t\}_{t \in \mathbb{Z}}$, the equations of the form

$$X_t = \sigma_t \varepsilon_t \quad \text{and} \quad \sigma_t^2 = c_0 + \sum_{i=1}^p a_i X_{t-i}^2 + \sum_{j=1}^q b_j \sigma_{t-j}^2 \quad (2.66)$$

where $c_0 > 0$, $a_i \geq 0, \forall i$ and $b_j \geq 0, \forall j$ are constants, and $\{\varepsilon_t\} \sim i.i.d.(0, 1)$ are innovations satisfying $\{\varepsilon_t\} \perp \{X_{t-k}, k \geq 1\}, \forall t$.

The σ_t 's are the so-called *volatilities*. A GARCH($p, 0$) process is the so-called ARCH(p) process. Let $\mathcal{F}_t = \sigma(\{X_s : s \leq t\})$ denote the sigma algebra representing the history of the process up to time t so that $\{\mathcal{F}_t\}_{t \in \mathbb{Z}}$ is the natural filtration. It follows that the conditional ES is formulated by

$$\text{ES}_\alpha(X_{t+1} | \mathcal{F}_t) = \sigma_{t+1} \text{ES}_\alpha(\varepsilon_t), \quad (2.67)$$

where $\text{ES}_\alpha(\varepsilon)$ is the ES of $\{\varepsilon_t\}$ at level α . Hence it can be estimated through

$$\widehat{\text{ES}}_\alpha(X_{t+1} | \mathcal{F}_t) = \widehat{\sigma}_{t+1} \widehat{\text{ES}}_\alpha(\varepsilon_t) \quad (2.68)$$

by plugging in the estimates of σ_{t+1} and $\text{ES}_\alpha(\varepsilon_t)$, $\widehat{\sigma}_{t+1}$ and $\widehat{\text{ES}}_\alpha(\varepsilon_t)$, respectively. Now the problem becomes how to predict one-step forward the volatility at time $t + 1$, σ_{t+1} , for a GARCH(p, q) model and estimate ES of the innovations at confidence level α . This involves fitting a GARCH model to data, volatility forecasting, and calculating ES of the innovations.

Three model fitting strategies are suggested in McNeil et al. (2005, section 4.4.2, pp.162):

- (1) Fit a GARCH model with an appropriate innovation distribution to the data by (conditional) maximum likelihood (ML), and use a prediction methodology to estimate σ_{t+1} . Any further parameters of the innovation distribution ought to be estimated simultaneously in the ML estimation process.
- (2) Fit a GARCH model by (conditional) quasi-maximum likelihood (QML) and use a prediction methodology as in strategy (1) to estimate σ_{t+1} . Afterwards use the model residuals to find an appropriate innovation distribution and estimate its parameters in a separate second step.
- (3) Estimate σ_{t+1} by EWMA (RiskMetrics, 1996). The innovation distribution could be estimated as in strategy (2) from the sequence of residuals calculated by dividing the observations by volatility estimates computed by EWMA.

Generally, the model residuals are inferred values for the unobserved innovations and recursively

calculated from the data and fitted model by

$$\widehat{R}_t^{model} = X_t/\widehat{\sigma}_t, \quad \text{and } \widehat{\sigma}_t = \sqrt{\widehat{c}_0 + \sum_{i=1}^p \widehat{a}_i X_{t-i}^2 + \sum_{j=1}^q \widehat{b}_j \widehat{\sigma}_{t-j}^2}. \quad (2.69)$$

Note that the first two strategies build the likelihood functions based on a GARCH model and the last one is a sensible model-free approach to volatility forecasting based on the classical technique of exponential smoothing.

Fitting GARCH Models to Data

For a GARCH(p, q) model of the form in (2.66), assume that the data is of size $n + p$ and labelled as $X_{-p+1}, \dots, X_0, X_1, \dots, X_n$ and let $g(\cdot)$ denote the density function of the innovations. Let $\boldsymbol{\theta} = (\alpha_0, \dots, \alpha_p, \beta_1, \dots, \beta_q)'$, $\mathbf{X} = (X_1, \dots, X_n)'$, $\mathbf{X}_0 = (X_{-p+1}, \dots, X_0)'$, and $\boldsymbol{\sigma}_0 = (\sigma_{-q+1}, \dots, \sigma_{-1}, \sigma_0)'$. Then the likelihood can be built conditionally on the initial values of X_{-p+1}, \dots, X_0 (observed) and $\sigma_{-q+1}, \dots, \sigma_{-1}, \sigma_0$ (unobserved) as follows. Since

$$f_{\mathbf{X}|\mathbf{X}_0, \boldsymbol{\sigma}_0}(\mathbf{X}|\mathbf{X}_0, \boldsymbol{\sigma}_0) = \prod_{t=1}^n f_{X_t|X_{t-1}, \dots, X_0}(X_t|X_{t-1}, \dots, X_0), \quad (2.70)$$

a conditional likelihood given $\mathbf{X}_0, \boldsymbol{\sigma}_0$ is of the form

$$L(\boldsymbol{\theta}; \mathbf{X}|\mathbf{X}_0, \boldsymbol{\sigma}_0) = \prod_{t=1}^n \frac{1}{\sigma_t} g\left(\frac{X_t}{\sigma_t}\right), \quad (2.71)$$

where $\sigma_t = \sqrt{c_0 + \sum_{i=1}^p b_i X_{t-i}^2 + \sum_{j=1}^q a_j \sigma_{t-j}^2}$. Since the required values of $\sigma_0, \sigma_{-1}, \dots, \sigma_{-q+1}$ are not actually observed, this is usually implemented by choosing starting values such as the sample variance of X_1, \dots, X_n , or even simply zero. It follows that the log-likelihood is of the form

$$\ln L(\boldsymbol{\theta}; \mathbf{X}|\mathbf{X}_0, \boldsymbol{\sigma}_0) = \sum_{t=1}^n l_t(\boldsymbol{\theta}|\mathbf{X}_0, \boldsymbol{\sigma}_0), \quad (2.72)$$

where l_t denotes the log-likelihood contribution arising from the t -th observation. Hence the MLE of $\boldsymbol{\theta}$ is a solution that maximizes the conditional log-likelihood in (2.72) and as a local maximum in general, solves the likelihood equations

$$\frac{\partial}{\partial \boldsymbol{\theta}} \ln L(\boldsymbol{\theta}; \mathbf{X}|\mathbf{X}_0, \boldsymbol{\sigma}_0) = \sum_{t=1}^n \frac{\partial l_t(\boldsymbol{\theta}|\mathbf{X}_0, \boldsymbol{\sigma}_0)}{\partial \boldsymbol{\theta}}, \quad (2.73)$$

where the left-hand side is also known as the *score vector* of the conditional likelihood. For Gaussian innovations, (2.71) is the so-called Gaussian likelihood and its conditional log-likelihood is simply

$$\sum_{t=1}^n (\ln \sigma_t^2 + X_t^2 / \sigma_t^2), \quad (2.74)$$

where $\sigma_t^2 = c_0 + \sum_{i=1}^p b_i X_{t-i}^2 + \sum_{j=1}^q a_j \sigma_{t-j}^2$. Apart from Gaussian distributions, scaled t distributions with mean zero and variance one are also frequently used forms of $g(\cdot)$. The Equations (2.72) are usually solved numerically using modified Newton-Raphson procedures. A particular method which is applied for GARCH models is the BHHH method of Berndt et al. (1974).

The last step is to check the validity of the whole model assumption used in the fitting procedure. Since the model residuals are supposed to behave like a strict white noise, correlograms of raw and absolute values can be constructed and portmanteau tests of strict white noise (see McNeil et al., 2005, Section 4.2.3) applied to investigate it. Assuming that the hypothesis of strict white noise is not rejected, so that the model has been satisfactorily fitted, the validity of the innovation distribution used in the ML fitting can be examined by QQ plots and goodness-of-fit tests for the Gaussian or scaled t distributions. Essentially, the Gaussian likelihood, for which $g(\cdot)$ is assumed as the Gaussian density function, is treated as an objective function to be maximized, and may still give reasonable parameter estimates if Gaussian is not the true innovation distribution. If the Gaussian MLEs seem to estimate the dynamics reasonably but the residuals do not behave like i.i.d. Gaussian observations, then *quasi-maximum likelihood* (QML) method comes in and the estimates obtained are QMLEs. This is sometimes called *pre-whitening* (McNeil et al., 2005) of the data and a separate second stage needs to be done by modelling the innovation distribution using the residuals from the GARCH model as data. This is the so-called *two-stage analysis* (McNeil et al., 2005) and may be a useful approach in higher-dimensional modelling.

Volatility Forecasting for GARCH Processes

Suppose that the model has been fitted from a sample of size n and its parameters estimated. To estimate σ_{t+1} in Equation (2.67) requires a prediction methodology for volatility forecasting to enable the volatility prediction one step ahead. Take the GARCH(1,1) model as an example, from which general procedure for more complex models could be analogized. A natural volatility prediction based on GARCH(1,1) models is given by

$$E(X_{n+1}^2 | \mathcal{F}_n) = \sigma_{n+1}^2 = c_0 + aX_n^2 + b\sigma_n^2. \quad (2.75)$$

In practice, an approximation has to be made based on formula (2.75) because the infinite series of past values that would allow the calculation of σ_{n+1}^2 is unavailable. It is intuitive to approximate σ_n^2 by an estimate of squared volatility $\widehat{\sigma}_n^2$ calculated recursively by Equations (2.69). Therefore, the approximate prediction of the squared volatility at time $n + 1$ can be made by

$$\widehat{\sigma}_{n+1}^2 = \widehat{\mathbb{E}}(X_{n+1}^2 | \mathcal{F}_n) = \widehat{c}_0 + \widehat{a}X_n^2 + \widehat{b}\widehat{\sigma}_n^2. \quad (2.76)$$

Alternatively, the exponentially weighted moving-average (EWMA) can be applied as a sensible model-free approach to volatility forecasting, based on the classical technique of exponential smoothing. A recursive scheme for one-step-ahead volatility forecasting could thus be defined by

$$\widehat{\sigma}_{t+1}^2 = (1 - \lambda)X_t^2 + \lambda\widehat{\sigma}_t^2 \quad (2.77)$$

for the decay factor λ , typically taken to be 0.94 by RiskMetrics for daily VaR estimation.

Approaches to Estimating ES of GARCH Innovations

The innovations of a GARCH model $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ are herein treated as the target data. Similar to ARMA process but subject to the constraint that $\{\varepsilon_t\}$ has mean zero and variance one, estimation methods are described in the following. On the parametric side, the estimation of ES for the innovations depends on the assumption of its distribution. The general idea remains to derive the formula of ES for the innovations under the specified distribution conditional on the past and then to plug the estimated parameters into this formula to obtain the estimated ES for the innovations. With regard to the assumption of innovation distribution, a Gaussian or (scaled) Student's t distributions with mean zero and variance one are usually candidates. The estimation methods for i.i.d. elliptical distributions, discussed in Section 2.3.3 is applicable since both Gaussian and Student's t distributions are special elliptical distributions. Recall that formula of ES for elliptical distributions from Equation (2.51) is

$$\text{ES}_\alpha(\varepsilon_t) = \mu + \sigma \cdot \lambda_\alpha \quad (2.78)$$

where

$$\lambda_\alpha = \frac{\overline{G}(\frac{1}{2}z_\alpha^2)}{1 - \alpha} \quad (2.79)$$

with $z_\alpha = \text{VaR}_\alpha(Z)$ for $Z \sim \text{E}(0, 1, g)$. Therefore, simply put $\mu = 0$ and $\sigma = 1$ in Equation (2.78), and then ES of Gaussian innovations under a GARCH model can be calculated by

$$\text{ES}_\alpha(\varepsilon_t) = \lambda_\alpha \quad \text{for } N(0, 1), \quad (2.80)$$

where λ_α is determined by Equation (2.54). Parallel to Gaussian distributions, put $\mu = 0$ and $\sigma = \sqrt{(r-2)/r}$ in Equation (2.78) for scaled Student's t , $t(0, \frac{r-2}{r}; r)$ with degree of freedom $r > 2$, such that its variance would be 1 instead of $r/(r-2)$ for standard Student's t . Thus

$$\begin{aligned} \text{ES}_\alpha(\varepsilon_t) &= \sigma \lambda_{r,\alpha} \\ &= \sqrt{\frac{r-2}{r}} \lambda_{r,\alpha} \quad \text{for } t(0, \frac{r-2}{r}; r), \end{aligned} \tag{2.81}$$

where $\lambda_{r,\alpha}$ is determined by (2.55). Recall that for $r > 2$,

$$\lambda_{r,\alpha} = \frac{1}{1-\alpha} \sqrt{\frac{r}{r-2}} \cdot f_Z \left(\sqrt{\frac{r-2}{r}} z_\alpha; r-2 \right).$$

Hence (2.81) turns out to be of the simple form

$$\text{ES}_\alpha(\varepsilon_t) = f_z \left(\sqrt{\frac{r-2}{r}} z_\alpha; r-2 \right) / (1-\alpha) := \eta_{r,\alpha}, \tag{2.82}$$

where f_z is a density of standard t distribution, $t(0, 1; r)$, and $\eta_{r,\alpha}$ represents the ES of scaled t with mean zero and variance one.

On the non-parametric side of estimating $\text{ES}_\alpha(\varepsilon_t)$, the sequence of model residuals still plays a major and more important role because it could be thought of as a sequence of realizations of the unobserved i.i.d. innovations. Hence non-parametric estimation of $\text{ES}_\alpha(\varepsilon_t)$ for GARCH processes is the same as that for ARMA processes discussed previously. Furthermore, if the GARCH process is fitted to data by two-stage analysis via QML when the model residuals do not behave like i.i.d. Gaussian, then non-parametric methods of estimating $\text{ES}_\alpha(\varepsilon_t)$ based on model residuals may give better estimate of conditional ES, $\text{ES}_\alpha(X_{t+1}|\mathcal{F}_t)$.

2.4.3 Conditional ES of ARMA-GARCH Processes

An approach similar to that described in the foregoing can be used to predict the conditional ES one-step forward for an ARMA models with GARCH errors, namely ARMA-GARCH process.

Definition 2.21. (ARMA-GARCH processes). Let $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ be strict white noise with mean zero and variance one. The time series $\{X_t\}_{t \in \mathbb{Z}}$ is said to be an ARMA(p_1, q_1) process with GARCH(p_2, q_2) errors $\{Z_t\}_{t \in \mathbb{Z}}$, if it is covariance stationary and if it satisfies, for all $t \in \mathbb{Z}$ and some strictly positive-valued process $\{\sigma_t\}_{t \in \mathbb{Z}}$, the equations of the form

$$\phi(B)(X_t - \mu) = \psi(B)Z_t \quad \text{and} \quad Z_t = \sigma_t \varepsilon_t \tag{2.83}$$

with

$$\sigma_t^2 = c_0 + \sum_{i=1}^{p_2} a_i (X_{t-i} - \mu_{t-i})^2 + \sum_{j=1}^{q_2} b_j \sigma_{t-j}^2$$

where $\phi(B) = 1 - \phi_1 B^1 - \dots - \phi_{p_1} B^{p_1}$, $\psi(B) = 1 - \psi_1 B^1 - \dots - \psi_{q_1} B^{q_1}$, $c_0 > 0$, $a_i \geq 0$, $i = 1, \dots, p_2$, $b_j \geq 0$, $j = 1, \dots, q_2$, $\sum_{i=1}^{p_2} a_i + \sum_{j=1}^{q_2} b_j < 1$, and $\mu_t = \mu + \sum_{i=1}^{p_1} \phi_i (X_{t-i} - \mu) + \sum_{j=1}^{q_1} \psi_j \sigma_{t-j} \varepsilon_{t-j}$.

Let ARMA(p_1, q_1)-GARCH(p_2, q_2) represents the ARMA-GARCH model with particular p_1 , q_1 , p_2 , and q_2 . Equation (2.83) can also be written as

$$X_t = \mu_t + \sigma_t \varepsilon_t, \quad (2.84)$$

where

$$\mu_t = \mu + \sum_{i=1}^{p_1} \phi_i (X_{t-i} - \mu) + \sum_{j=1}^{q_1} \psi_j \sigma_{t-j} \varepsilon_{t-j},$$

and

$$\sigma_t^2 = c_0 + \sum_{i=1}^{p_2} a_i (X_{t-i} - \mu_{t-i})^2 + \sum_{j=1}^{q_2} b_j \sigma_{t-j}^2.$$

Therefore, incorporating concepts of previous sections with (2.57) for ARMA-GARCH processes, the ES of X_{t+1} at level α conditional on the past history up to time t is of the form

$$\text{ES}_\alpha(X_{t+1}|\mathcal{F}_t) = \mu_t + \sigma_{t+1} \text{ES}_\alpha(\varepsilon_t), \quad (2.85)$$

where $\text{ES}_\alpha(\varepsilon_t)$ is the ES of $\{\varepsilon_t\}$ at level α . Similarly, the estimation methods of conditional ES for GARCH processes can also be applied for ARMA-GARCH processes. That is, in Equation (2.85), μ_t can be estimated by fitting ARMA-GARCH models to data, then σ_{t+1} predicted from the estimated model, and finally $\text{ES}_\alpha(\varepsilon_t)$ predicted by the standardized residuals from model.

2.5 Summary of Contributions

The contributions in the chapter are summarized below.

- Historical simulation (HS) estimate of ES
- ES formula for elliptical distributions (Landsman and Valdez, 2003) through λ -factor
- Applications of the above formula:
 - ES formula for Student's t distributions (Landsman and Valdez, 2003)

- ES formulas for ARMA-GARCH processes (special cases: ARMA and GARCH processes)
- Application of bootstrap techniques for bias adjustment.

Chapter 3

Evaluation of the Performance of Risk Models

To evaluate different estimators of the same parameter of interest, there are two major aspects to be investigated: model-based criteria and data-based criteria. Bias, mean square error, and consistency lies in the category of model-based criteria. Nonetheless, relative accuracy measures will be considered, because the true value of ES varies with volatilities as volatilities vary with time.

3.1 Model-based Criteria

Proposed estimators are usually evaluated by consistency and their accuracy measures, including bias and mean square error (MSE). However, under changing volatility models based on GARCH models, the true value of ES is not a constant, but varies with volatilities while volatilities vary with time. Due to this special feature, relative accuracy measures have to be taken into account.

Note that MLEs are typically consistent under certain regularity conditions, asymptotically efficient in general, and asymptotically normally distributed. The most useful property of MLE is known as the invariance property of maximum likelihood estimators. Therefore, once ES or ESC can be written as a function of the parameters, thereafter it can be estimated by the function of MLEs of those parameters. In this way, the estimated ES turns out to be an MLE of ES and can be regarded as a benchmark to compare other estimators of ES (or ESC) with it.

3.1.1 Existent Large Sample Properties of HS Estimates for ES

In Acerbi and Tasche (2002, Proposition 4.1) and Acerbi (2004), it has been shown that at the confidence level α , the HS estimator of ES, given as

$$\widehat{\text{ES}}_\alpha = \frac{1}{n(1-\alpha)} \sum_{t=\lfloor n\alpha \rfloor + 1}^n X_{(t)}, \quad (3.1)$$

is consistent for an independent sequence $\{X_t\}_{t=1}^n$ of random variables and that it also converges to the true value of ES in L_1 if the random variable X of the potential loss is integrable. For a large class of time series, it may be inferred from Puri and Ruymgaart (1993) that the HS estimator of ES as a linear function of L -statistics is consistent provided that its variance exists.

In Chen (2008), the non-parametric estimator of ES is given by

$$\frac{\sum_{i=1}^n Y_i 1_{\{Y_i \geq \text{VaR}_\alpha\}}}{\sum_{i=1}^n 1_{\{Y_i \geq \text{VaR}_\alpha\}}} = (n - \lfloor n\alpha \rfloor)^{-1} \sum_{t=1}^n Y_t 1_{\{Y_t \geq \text{VaR}_\alpha\}}, \quad (3.2)$$

where Y is defined as the negative log return (or loss) and VaR_α is the VaR of Y , and shown that under proper conditions, the asymptotic normality holds. Precisely, under proper conditions, we have

$$\sqrt{n}(1-\alpha)\sigma_0^{-1}(\alpha; n)(\widehat{\text{ES}}_\alpha - \text{ES}_\alpha) \xrightarrow{d} N(0, 1) \quad (3.3)$$

where $\sigma_0^2(\alpha; n) = \{Var[(Y_1 - \text{VaR}_\alpha)1_{\{Y_1 \leq \text{VaR}_\alpha\}}] + 2 \sum_{k=1}^{n-1} \gamma(k)\}$ with

$$\gamma(k) = Cov[(Y_1 - \text{VaR}_\alpha)1_{\{Y_1 \leq \text{VaR}_\alpha\}}, (Y_{k+1} - \text{VaR}_\alpha)1_{\{Y_{k+1} \leq \text{VaR}_\alpha\}}]$$

for positive integer k . Note that the non-parametric estimator of ES (3.2) used in Chen (2008) required the estimation of VaR prior to it but the HS estimator (3.1) used in this study does not require any estimation of VaR. As a matter of fact, the non-parametric estimator of ES (3.2) used in Chen (2008) is the empirical estimator of the so-called CTE (conditional tail expectation)

$$E[Y|Y \geq \text{VaR}_\alpha(Y)]$$

rather than of ES in general. See Section 1.2.1 for more information about the difference between ES and CTE.

For an independent sequence $\{X_t\}_{t=1}^n$ of random variables with its ordered statistics $X_{i:n}; i = 1, \dots, n$, define the L -statistic as

$$L_n = \sum_{i=1}^n c_{in} X_{i:n} = \frac{1}{n} \sum_{i=1}^n J\left(\frac{i}{n}\right) X_{i:n} = \int_{-\infty}^{\infty} x J[\widehat{F}_n(x)] d\widehat{F}_n(x) = \int_0^1 \widehat{F}_n^{-1}(u) J(u) du \quad (3.4)$$

where $J(u)$ is a function of u ($0 \leq u \leq 1$) such that $J(i/n) = nc_{in}$ ($1 \leq i \leq n$) and \widehat{F}_n is the empirical c.d.f. It is a linear function of order statistics. Define

$$\mu(J, F) = \int_{-\infty}^{\infty} xJ(F(x))dF(x) = \int_0^1 J(u)F^{-1}(u)du \quad (3.5)$$

and

$$\begin{aligned} \sigma^2(J, F) &= 2 \int \int_{\infty < x < y < \infty} J(F(x))J(F(y))[F(x)(1 - F(y))]dx dy \\ &= 2 \int \int_{0 < u < v < 1} J(u)J(v)u(1 - v)dF^{-1}(u)dF^{-1}(v). \end{aligned} \quad (3.6)$$

The following theorem in David and Nagaraja (2003, section 11.4) presents the asymptotic normality of the L -statistic under some regularity conditions.

Theorem 3.1. (David and Nagaraja (2003)) *Assume that $\text{Var}(X_1) < \infty$ and that $J(u)$ is bounded and continuous a.e. F^{-1} . Then*

$$\lim_{n \rightarrow \infty} n\text{E}(L_n) = \mu(J, F), \quad \lim_{n \rightarrow \infty} n \cdot \text{Var}(L_n) = \sigma^2(J, F),$$

and if $\sigma^2(J, F) > 0$,

$$L_n^* = \frac{L_n - \text{E}(L_n)}{[\text{Var}(L_n)]^{\frac{1}{2}}} \xrightarrow{d} N(0, 1). \quad (3.7)$$

Furthermore, L^* in (3.7) can be replaced by

$$L_n^0 = n^{1/2} \frac{[L_n - \mu(J, F)]}{\sigma(J, F)}$$

when the two conditions below hold:

- (a) $\int_0^1 [u(1 - u)]^{1/2} dF^{-1}(u) < \infty$, a condition almost equivalent to having finite variance.
- (b) J is bounded and satisfies a Lipschitz¹ condition of order $p > \frac{1}{2}$ except perhaps at a finite number of discontinuity points of F^{-1} ,

$$\lim_{n \rightarrow \infty} n^{1/2} [\text{E}(L_n) - \mu(J, F)] = 0.$$

Hence for an i.i.d. random sample under some regularity conditions, the HS estimator of ES at confidence level α , given in (3.1), is asymptotically unbiased and its asymptotic variance is $\sigma^2(J, F)$ defined in (3.6) with $J(u) = (1 - \alpha)^{-1} \forall u \in [0, 1]$.

For a large class of time series (non i.i.d.) under some regularity conditions, it can be

¹A function $f(x)$ satisfies a Lipschitz condition of order p on an interval if there is a constant such that $|f(x) - f(y)| \leq c|x - y|^\alpha$ for all x and y in the interval.

inferred from Puri and Ruymgaart (1993) that the HS estimator of ES at confidence level α remains asymptotically unbiased.

3.1.2 Relative Accuracy Measures

First of all, note that due to the GARCH model assumption, the value of the ES at a fixed time for each simulation run will not be identical, because it depends on the value of the volatility at the time where it is predicted. Therefore the traditional accuracy measures, bias and mean square error (MSE), are no longer reliable and then the relative accuracy measures will be introduced below to evaluate estimates of ES. Firstly, for a fixed time point, the relative bias is defined by

$$ReBias = E \left(\frac{\widehat{ES}_\alpha - ES_\alpha}{ES_\alpha} \right) \quad (3.8)$$

where X , ES_α , \widehat{ES}_α , are the observation, true value of ES, and estimates of ES, respectively. For example, if the model is fitted by a GARCH model, then $X = X_t$, $ES_\alpha = ES_\alpha(X_t|\mathcal{F}_{t-1})$, and $\widehat{ES}_\alpha = \widehat{ES}_\alpha(X_t|\mathcal{F}_{t-1})$. Secondly, the relative MSE is defined in the similar fashion by

$$ReMSE = E \left[\left(\frac{\widehat{ES}_\alpha - ES_\alpha}{ES_\alpha} \right)^2 \right]. \quad (3.9)$$

3.1.3 Simulation Study

The following simulation study has been conducted to investigate the performance of those estimators of ES proposed in the this chapter through examining their relative accuracy measures. Four different sizes, 50, 250, 500 and 1,000, of samples were simulated from a stationary GARCH(1,1) process with scaled Student's t innovations in the form of

$$X_t = \sigma_t \varepsilon_t \quad \text{and} \quad \sigma_t^2 = c_0 + aX_{t-1}^2 + b\sigma_{t-1}^2$$

where $b = 0.75$, $a = 4(1 - b)/5$ and $c_0 = v(1 - a - b)$ with $v = 0.00004$, which is the long-term variance such that the long-term volatility is square root of it, 0.63%. Precisely, the innovations are generated from the scaled Student's t with degree of freedom 4; that is, $\varepsilon_t \sim t(0, 1/2; 4)$ such that the variance of ε_t would be one. The simulation was replicated for 1,000 times. The confidence level of ES was set to be $\alpha = 0.95$.

As mentioned in Section 2.3.3, it can be foreseen that the performance of parametric estimates of ES for elliptical distributions will be mostly driven by the goodness of model fitting, since ES of elliptical distributions can be explicitly expressed by a linear function of parameters of the underlying distribution. It is also true for conditional ES estimates of time series with

errors of elliptical distributions, since the underlying distribution conditional on the past is to be elliptical as well, and hence conditional ES can be estimated by a function of past realizations and estimated model parameters by appropriately applying Theorem 2.18.

Here is a summary of methods of estimating ES considered:

1. Without volatility filter:

- Gaussian model (GM)
- HS
- GPD.

All above estimates may be improved by adjusting for bias by applying appropriate bootstrap techniques: (1)ordinary bootstrap (OB), (2)block-wise bootstrap (BB). The bootstrap sample sizes are 50, 250, 400, and 400 for sample size 50, 250, 500, 1000, respectively. Note that the exact bootstrap (EB) is only applicable to HS estimates of ES.

2. With volatility filter: Model fitting and volatility prediction:

- GARCH
- EWMA.

ES estimation of GARCH model errors:

- one-staged: Gaussian and Student's t
- two-staged: HS and GPD.

All above estimates may be improved by adjusting for bias by applying OB. Again, EB is only applicable to HS estimates of ES for model errors.

Hence the relative accuracy measures defined in Section 3.1.2 will be reported from Monte Carlo simulation. That is, the relative bias is approximated by

$$ReBias = \frac{1}{N} \sum_{i=1}^N \frac{\widehat{ES}_{\alpha,i} - ES_{\alpha,i}}{ES_{\alpha,i}}, \quad (3.10)$$

and the relative MSE is approximated by

$$ReMSE = \frac{1}{N-1} \sum_{i=1}^N \left(\frac{\widehat{ES}_{\alpha,i} - ES_{\alpha,i}}{ES_{\alpha,i}} \right)^2, \quad (3.11)$$

where X_i , $\widehat{ES}_{\alpha,i}$, $ES_{\alpha,i}$, are the observation, true value of ES, and estimates of ES, for the i -th simulation respectively, and N is the total number of simulations. For example, if the model

is fitted by a GARCH model, then for each simulation, $X = X_t$, $ES_\alpha^t = ES_\alpha(X_t|\mathcal{F}_{t-1})$, and $\widehat{ES}_\alpha^t = \widehat{ES}_\alpha(X_t|\mathcal{F}_{t-1})$, respectively.

The following sections are simulation results for different sample sizes: 50, 250, 500, and 1000. The results include absolute values of relative bias and values of relative MSE along with their standard errors. For convenience, “Simple”, “GARCH-N”, “EWMA-N” and “GARCH-t” denote the simple estimating methods without volatility filter, GARCH model with Gaussian innovations, EWMA method of volatility prediction for GARCH-N, and GARCH model with Student’s t innovations, respectively. The combination of a specific model and a specific estimating method means ES is estimated by plugging estimates of this model parameters, estimated by this method, in the ES formula for that model. For example, the combination of Simple and ML means ES is estimated by substituting parameters by corresponding estimates in the ES formula for Gaussian models. Furthermore, for the combination of a changing volatility model, such as GARCH, and HS (or GPD) method means that the model is estimated by two-stage QML and the ES of innovations is estimated using model residuals by HS (or GPD). Besides, let “ReBias” and “ReMSE” denote relative bias and relative MSE, respectively. Those ES estimates without bias adjustment are called original estimates herein. In addition, values are multiplied by 100 for better reading; that is, the unit is %.

Results for Sample of Size 50

Now let’s take a look at results of the simulation study when sample size is 50. The relative MSE of original ES estimates for sample of size 50 are shown in Table 3.1. From the plot of the table, it is easy to see that ML is likely to outperform HS and GPD with smaller ReMSE for each model. For ML, GARCH-t has the smallest ReMSE as expected, GARCH-N is in the second place, and Simple has the largest ReMSE. Additionally, GPD is likely to outperform HS for EWMA-N and GARCH-t and otherwise for Simple and GARCH-N. From Table 3.2 for Simple, it shows that OB and BB both reduce ReBias for ML insignificantly, according to the standard errors, and none of them reduces ReMSE simultaneously. Table 3.3 for GARCH-N, Table 3.4 for EWMA-N and 3.5 for GARCH-t all show that OB reduce ReBias for GARCH-N fitted by ML insignificantly, according to the standard errors, and it does not reduce ReMSE simultaneously. In sum, bias adjustment does not work well when sample size is 50.

Results for Sample of Size 250

Let us take a look at results of the simulation study when sample size is 250. Table 3.6 shows that for the original ES estimates overall, ML is likely to dominate GPD with smaller ReMSE, and GPD is likely to dominate HS. It is as same as the result for sample size of 50 that GARCH-t still has the smallest ReMSE, that GARCH-N has second smallest ReMSE, and that Simple

Table 3.1: Relative MSE of Original ES Estimates (Sample Size = 50) in %



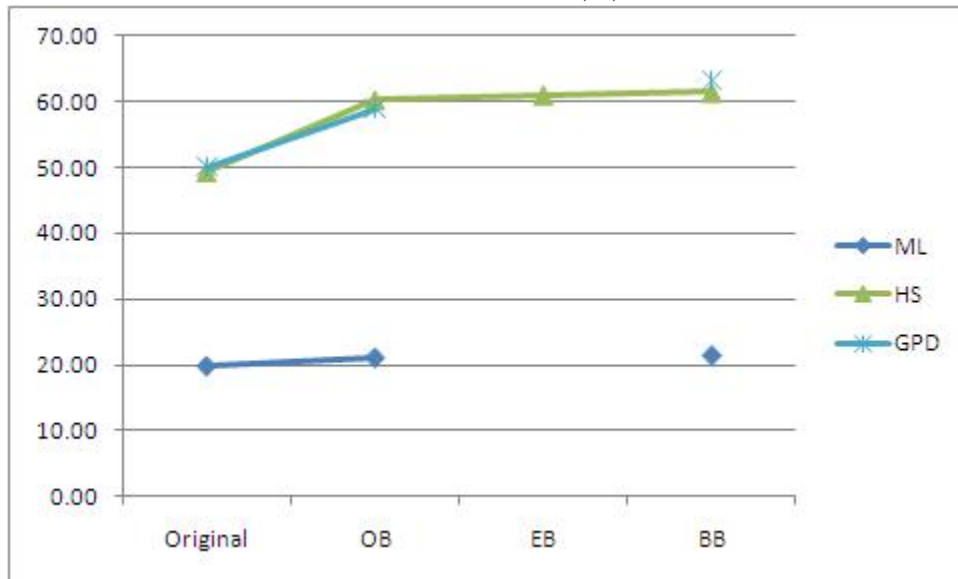
%	ML	HS	GPD
Simple	19.87 (2.63)	49.27 (8.44)	50.08 (9.39)
GARCH-N	7.87 (0.53)	14.06 (1.16)	14.07 (1.21)
EWMA-N	8.55 (0.68)	26.85 (2.68)	25.76 (2.59)
GARCH-t	7.34 (0.37)	15.63 (1.27)	14.76 (1.18)

Table 3.2: Relative Accuracy Measures of Simple ES Estimates (Sample Size = 50)

1. Relative Bias (%)

%	Original	OB	EB	BB
ML	-3.85 (1.41)	-1.84 (1.45)		-1.41 (1.46)
HS	19.20 (2.14)	25.78 (2.32)	20.33 (2.39)	25.86 (2.34)
GPD	16.60 (2.18)	21.67 (2.33)		22.69 (2.41)

2. Relative MSE (%)

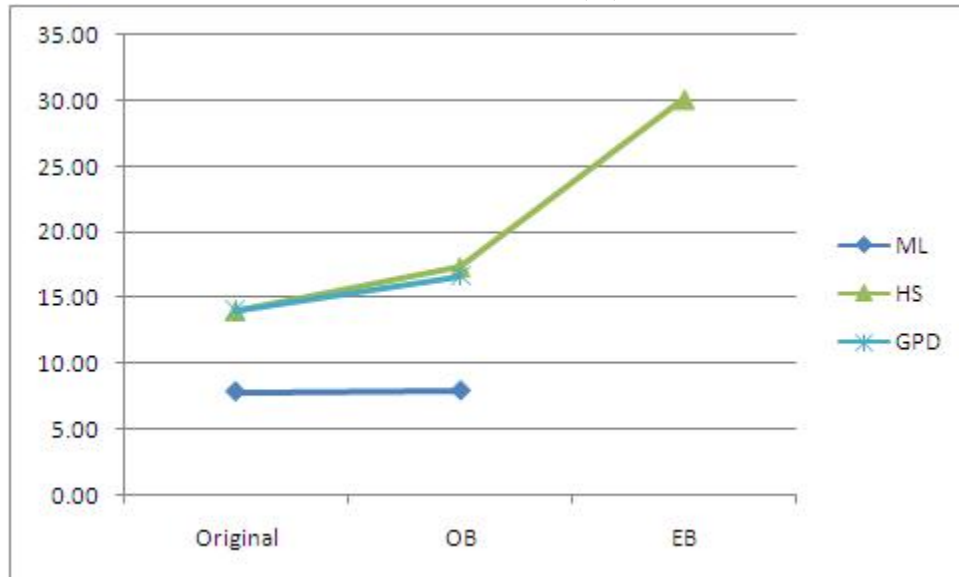


%	Original	OB	EB	BB
ML	19.87 (2.63)	21.09 (2.89)		21.44 (2.86)
HS	49.27 (8.44)	60.28 (10.20)	61.03 (11.80)	61.46 (10.18)
GPD	50.08 (9.39)	58.92 (11.17)		63.32 (11.90)

Table 3.3: Relative Accuracy Measures of ES Estimates for GARCH-N (Sample Size = 50)

		Original		OB		EB	
1. Relative Bias (%)	ML	-13.00	(0.79)	<u>-11.33</u>	(0.82)	-19.22	(1.63)
	HS	5.27	(1.17)	10.70	(1.28)		
	GPD	2.19	(1.19)	6.44	(1.28)		

2. Relative MSE (%)

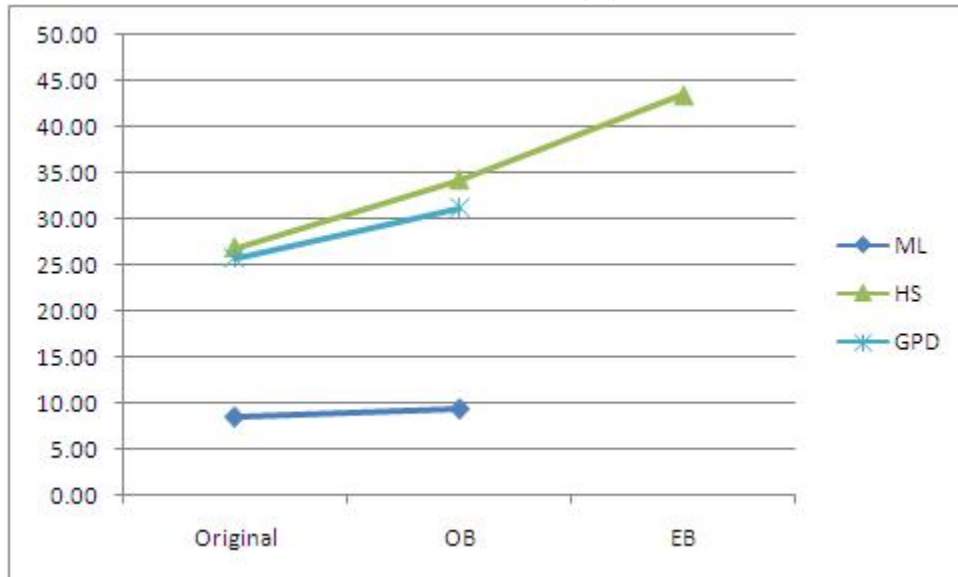


	Original		OB		EB	
ML	7.87	(0.53)	7.94	(0.62)	30.14	(2.56)
HS	14.06	(1.16)	17.40	(1.52)		
GPD	14.07	(1.21)	16.66	(1.48)		

Table 3.4: Relative Accuracy Measures of ES Estimates for EWMA (Sample Size = 50)

		Original		OB		EB	
1. Relative Bias (%)	ML	-4.77	(0.91)	-2.50	(0.97)		
	HS	18.16	(1.54)	24.55	(1.68)	-22.50	(1.96)
	GPD	15.07	(1.53)	20.10	(1.65)		

2. Relative MSE (%)

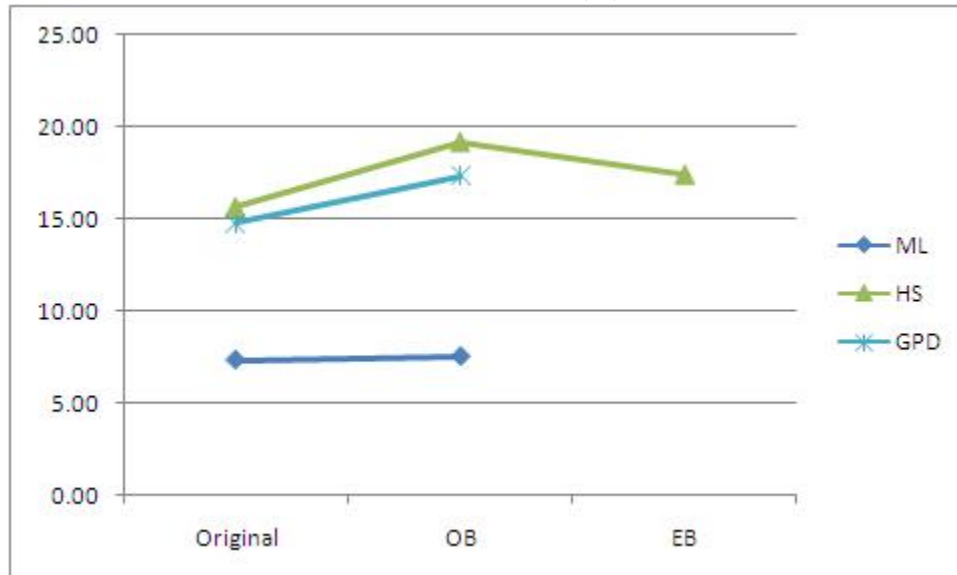


		Original		OB		EB	
ML		8.55	(0.68)	9.37	(0.82)		
HS		26.85	(2.68)	34.29	(3.59)	43.50	(3.85)
GPD		25.76	(2.59)	31.17	(3.14)		

Table 3.5: Relative Accuracy Measures of ES Estimates for GARCH-t (Sample Size = 50)

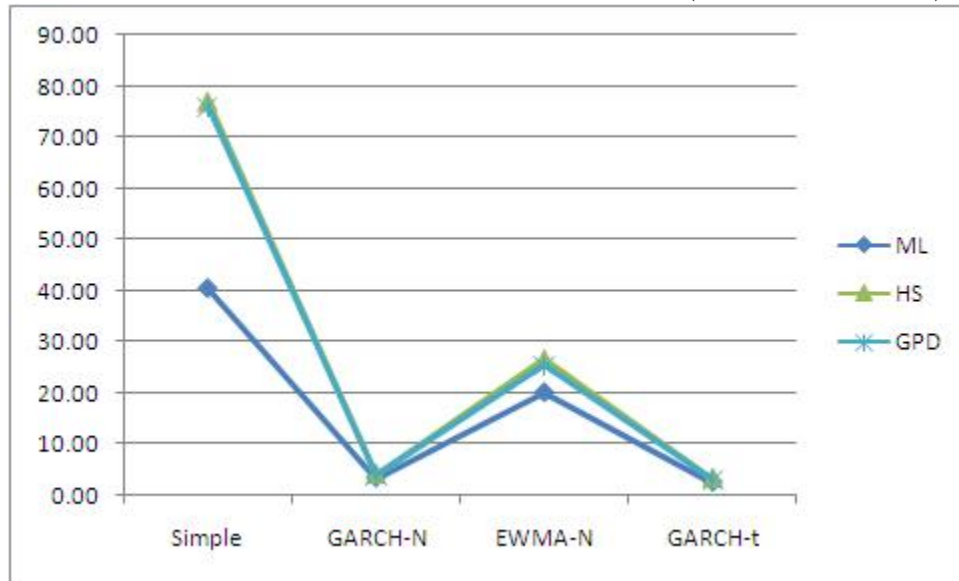
	Original	OB	EB	
1. Relative Bias (%)	ML	-11.28 (0.78)	<u>-9.30</u> (0.82)	5.20 (1.31)
	HS	7.29 (1.23)	12.58 (1.33)	
	GPD	3.61 (1.21)	7.61 (1.30)	

2. Relative MSE (%)



	Original	OB	EB
ML	7.34 (0.37)	7.54 (0.49)	17.39 (1.43)
HS	15.63 (1.27)	19.15 (1.62)	
GPD	14.76 (1.18)	17.35 (1.44)	

Table 3.6: Relative MSE of Original ES Estimates (Sample Size = 250)



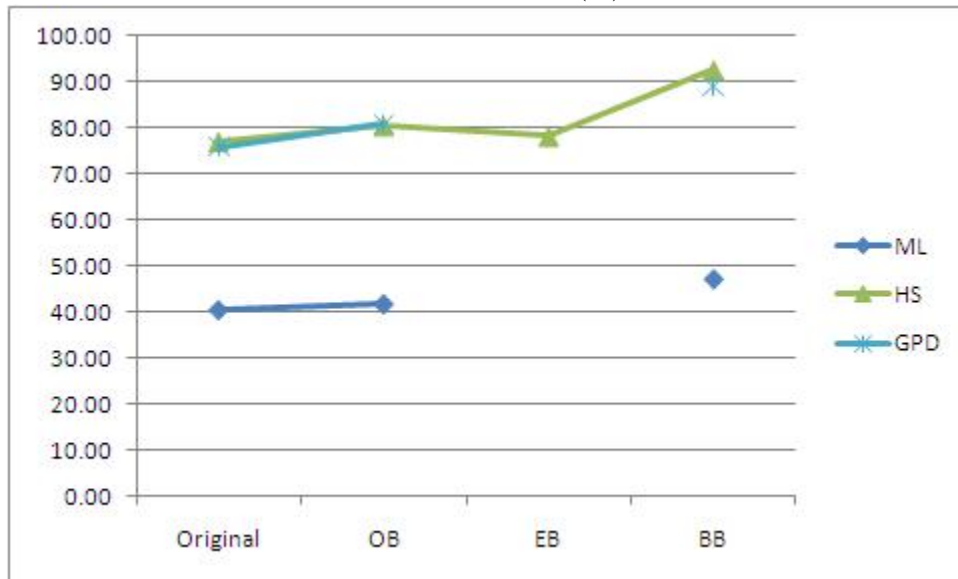
%	ML	HS	GPD
Simple	40.57 (10.09)	77.08 (21.90)	75.90 (22.27)
GARCH-N	3.33 (0.20)	4.09 (0.37)	3.94 (0.33)
EWMA-N	20.10 (6.45)	26.53 (8.13)	25.40 (7.97)
GARCH-t	2.49 (0.10)	3.29 (0.20)	3.13 (0.17)

Table 3.7: Relative Accuracy Measures of Simple ES Estimates (Sample Size = 250)

1. Relative Bias (%)

%	Original	OB	EB	BB
ML	8.99 (1.99)	9.76 (2.02)		11.23 (2.14)
HS	27.41 (2.64)	28.94 (2.69)	<u>24.10</u> (2.69)	31.08 (2.88)
GPD	24.99 (2.64)	26.30 (2.72)		28.01 (2.85)

2. Relative MSE (%)

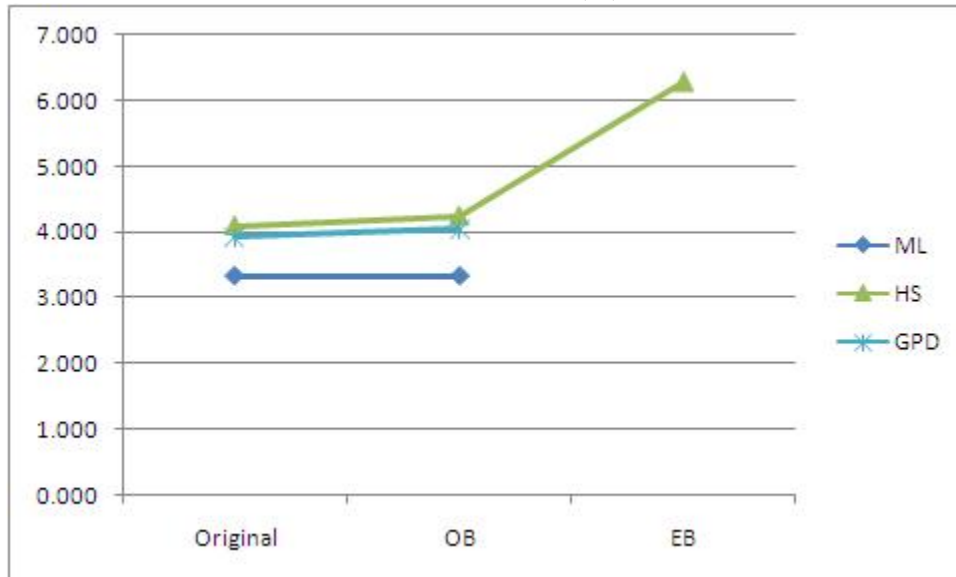


%	Original	OB	EB	BB
ML	40.57 (10.09)	41.79 (10.33)		47.07 (12.05)
HS	77.08 (21.90)	80.50 (22.98)	78.23 (23.69)	92.63 (27.58)
GPD	75.90 (22.27)	80.89 (24.39)		88.90 (27.15)

Table 3.8: Relative Accuracy Measures of ES Estimates for GARCH-N (Sample Size = 250)

		Original		OB		EB	
1. Relative Bias (%)	ML	-9.99	(0.48)	<u>-9.52</u>	(0.49)	-9.40	(0.74)
	HS	1.39	(0.64)	2.49	(0.65)		
	GPD	-1.25	(0.63)	<u>-0.45</u>	(0.64)		

2. Relative MSE (%)

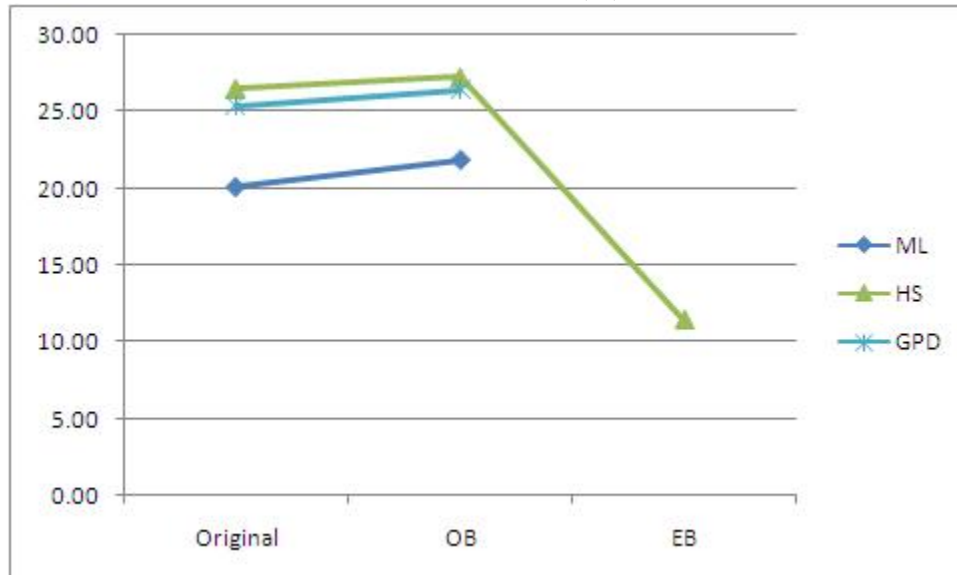


	Original		OB		EB	
ML	3.33	(0.20)	3.33	(0.21)	6.29	(0.71)
HS	4.09	(0.37)	4.24	(0.39)		
GPD	3.94	(0.33)	4.05	(0.34)		

Table 3.9: Relative Accuracy Measures of ES Estimates for EWMA (Sample Size = 250)

		Original		OB		EB	
1. Relative Bias (%)	ML	-0.93	(1.42)	<u>-0.22</u>	(1.48)	-19.85	(0.87)
	HS	12.83	(1.58)	13.99	(1.59)		
	GPD	10.15	(1.56)	11.09	(1.59)		

2. Relative MSE (%)

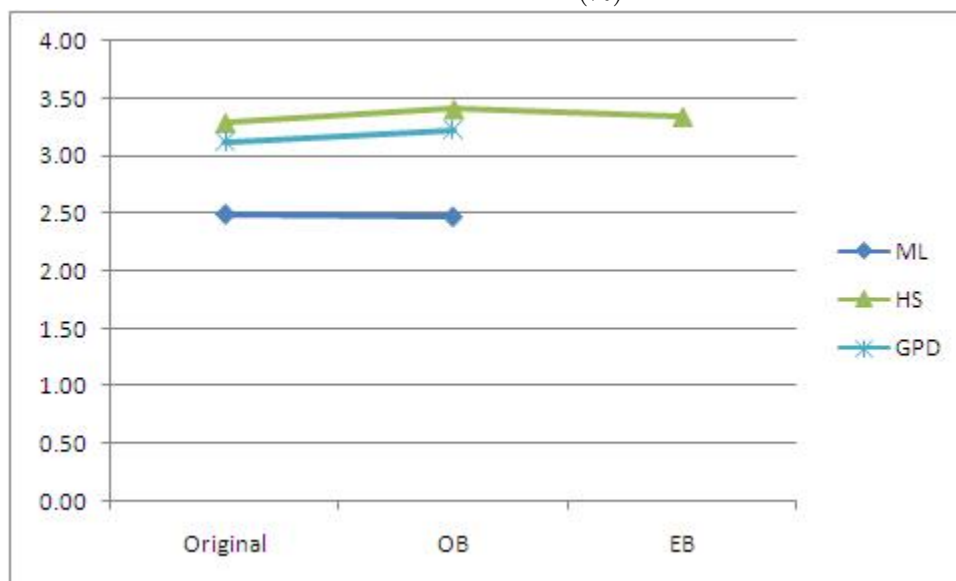


	Original		OB		EB	
ML	20.10	(6.45)	21.86	(7.41)	<u>11.44</u>	(0.95)
HS	26.53	(8.13)	27.26	(8.22)		
GPD	25.40	(7.97)	26.45	(8.39)		

Table 3.10: Relative Accuracy Measures of ES Estimates for GARCH-t (Sample Size = 250)

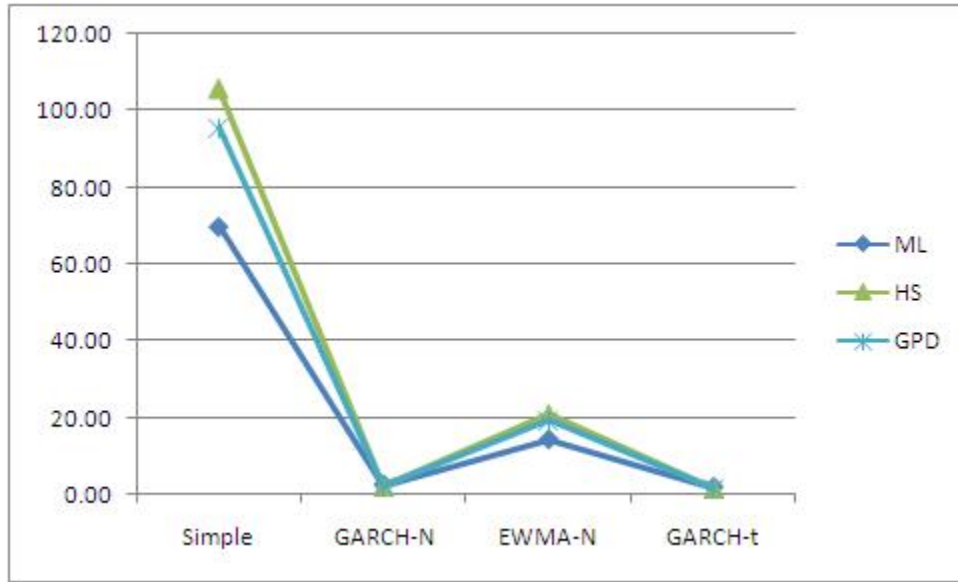
		Original		OB		EB	
1. Relative Bias (%)	ML	-9.74	(0.39)	<u>-9.21</u>	(0.40)	-3.30	(0.57)
	HS	1.46	(0.57)	2.51	(0.58)		
	GPD	-1.46	(0.56)	<u>-0.68</u>	(0.57)		

2. Relative MSE (%)



	Original		OB		EB	
ML	2.49	(0.10)	<u>2.47</u>	(0.11)	3.34	(0.16)
HS	3.29	(0.20)	3.41	(0.21)		
GPD	3.13	(0.17)	3.23	(0.17)		

Table 3.11: Relative MSE of Original ES Estimates (Sample Size = 500)



%	ML	HS	GPD
Simple	69.96 (44.20)	105.79 (60.57)	95.40 (54.98)
GARCH-N	2.46 (0.27)	2.30 (0.23)	2.13 (0.19)
EWMA-N	14.30 (3.74)	21.10 (5.35)	19.15 (5.23)
GARCH-t	1.85 (0.15)	1.68 (0.12)	1.56 (0.10)

has the largest ReMSE. Table 3.7 shows that for Simple, EB reduces ReBias for HS, but not significantly according to the standard errors, and it does not reduce ReMSE simultaneously. Table 3.8 shows that for GARCH-N, OB slightly reduces ReBias for ML and GPD insignificantly, according to the standard errors, and it reduce an insignificant amount of ReMSE only for ML simultaneously. Table 3.9 shows that for EWMA-N, OB reduces ReBias for ML insignificantly, according to the standard errors, and it does not reduce ReMSE simultaneously; however, EB reduces ReMSE for HS insignificantly, according to the standard errors. Table 3.10 shows that for GARCH-t, OB reduces ReBias for ML and GPD insignificantly, according to the standard errors, and it reduce ReMSE only for ML simultaneously but insignificantly, according to the standard errors.

Results for Sample of Size 500

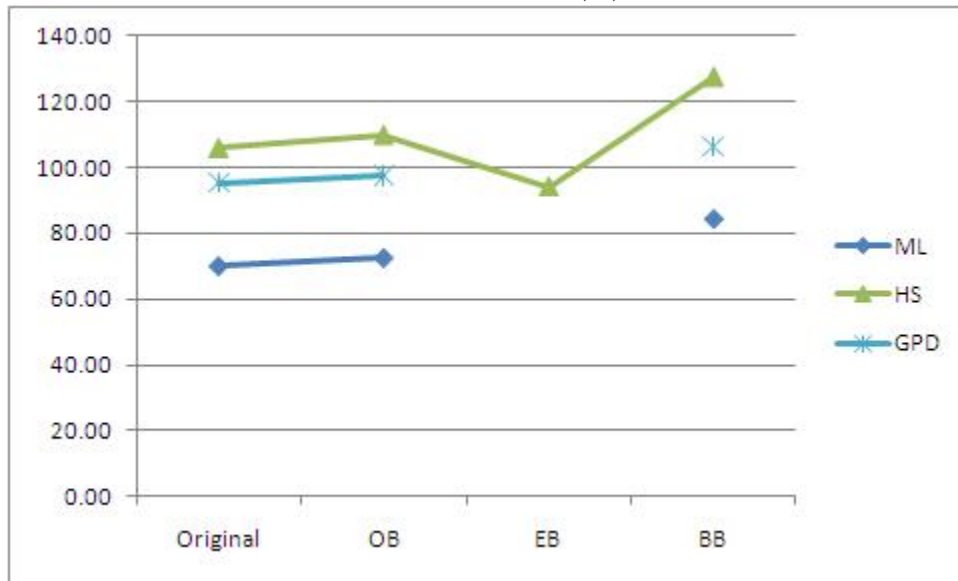
Let us proceed to examine the results for sample size of 500. The result for the original ES estimates in Table 3.11 shows that for GARCH-N and GARCH-t, GPD is likely to outperform

Table 3.12: Relative Accuracy Measures of Simple ES Estimates (Sample Size = 500)

1. Relative Bias (%)

%	Original	OB	EB	BB
ML	15.41 (2.60)	15.99 (2.64)		17.08 (2.85)
HS	34.41 (3.07)	35.22 (3.12)	<u>30.24</u> (2.92)	36.00 (3.39)
GPD	29.89 (2.94)	30.55 (2.97)		31.24 (3.11)

2. Relative MSE (%)

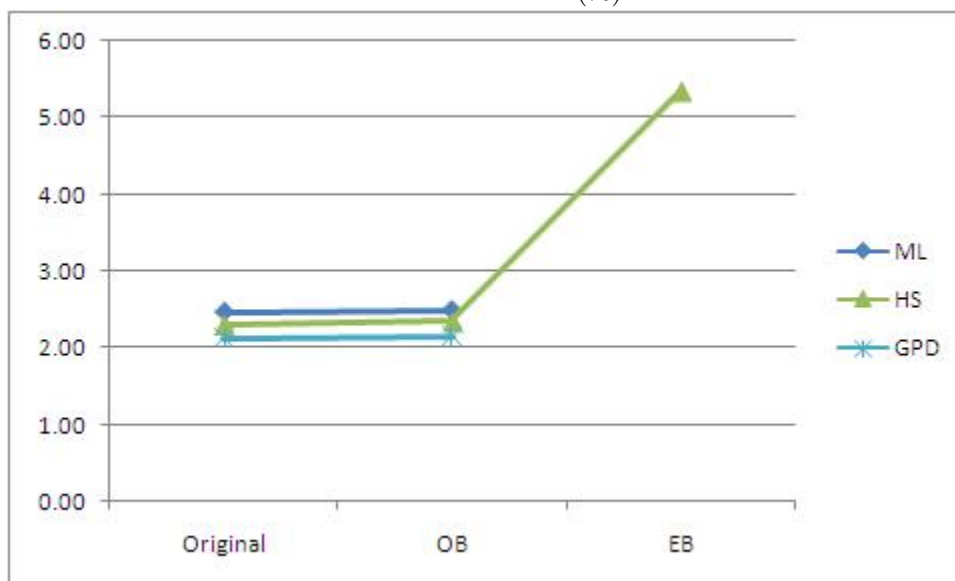


%	Original	OB	EB	BB
ML	69.96 (44.20)	72.40 (45.82)		84.17 (55.60)
HS	105.79 (60.57)	109.77 (63.51)	<u>94.11</u> (53.28)	127.58 (80.07)
GPD	95.40 (54.98)	97.64 (56.51)		106.47 (64.06)

Table 3.13: Relative Accuracy Measures of ES Estimates for GARCH-N (Sample Size = 500)

		Original		OB		EB	
1. Relative Bias (%)	ML	-9.34	(0.40)	<u>-9.04</u>	(0.41)	-4.83	(0.72)
	HS	2.34	(0.47)	2.87	(0.48)		
	GPD	-1.71	(0.46)	<u>-1.28</u>	(0.46)		

2. Relative MSE (%)

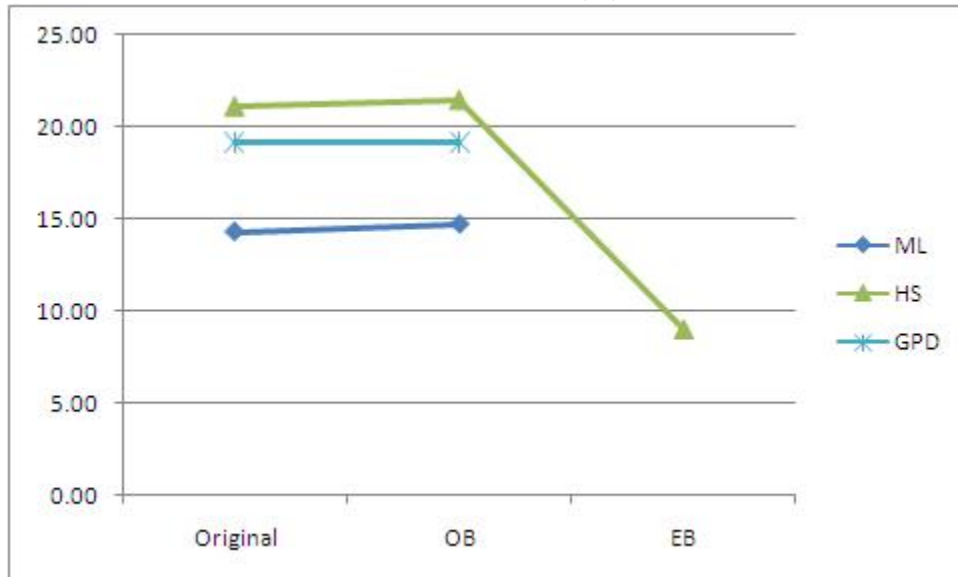


	Original		OB		EB	
ML	2.46	(0.27)	2.49	(0.30)	5.34	(2.29)
HS	2.30	(0.23)	2.34	(0.23)		
GPD	2.13	(0.19)	2.15	(0.19)		

Table 3.14: Relative Accuracy Measures of ES Estimates for EWMA (Sample Size = 500)

		Original		OB		EB	
1. Relative Bias (%)	ML	0.18	(1.20)	0.62	(1.21)		
	HS	14.13	(1.38)	14.74	(1.39)	-15.04	(0.82)
	GPD	10.05	(1.35)	10.46	(1.35)		

2. Relative MSE (%)

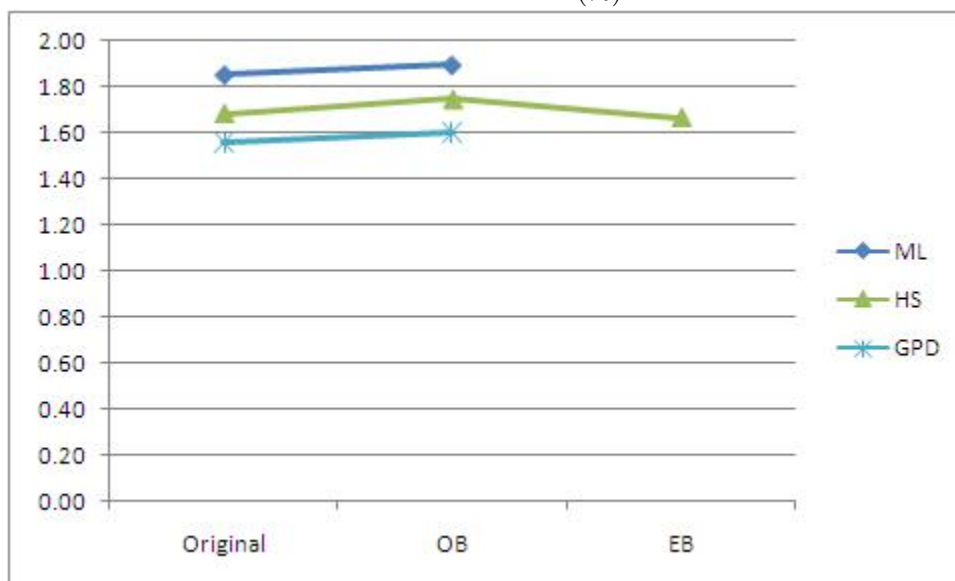


		Original		OB		EB	
ML		14.30	(3.74)	14.72	(3.82)		
HS		21.10	(5.35)	21.45	(5.38)	<u>9.03</u>	(0.82)
GPD		19.15	(5.23)	19.16	(5.05)		

Table 3.15: Relative Accuracy Measures of ES Estimates for GARCH-t (Sample Size = 500)

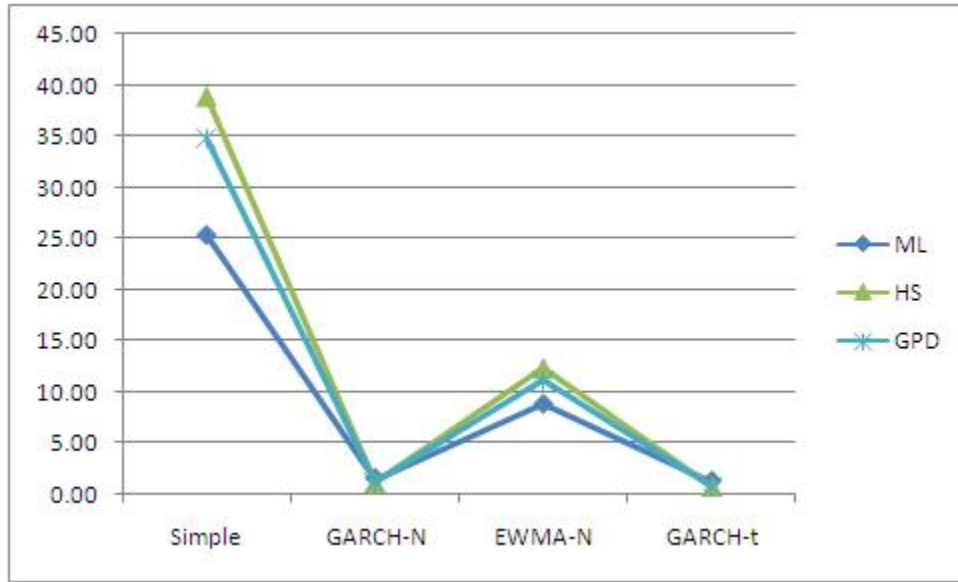
		Original		OB		EB	
1. Relative Bias (%)	ML	-9.08	(0.32)	<u>-8.76</u>	(0.34)	-2.31	(0.40)
	HS	2.25	(0.40)	2.79	(0.41)		
	GPD	-1.96	(0.39)	<u>-1.52</u>	(0.40)		

2. Relative MSE (%)



	Original		OB		EB	
ML	1.85	(0.15)	1.89	(0.19)	<u>1.67</u>	(0.10)
HS	1.68	(0.12)	1.75	(0.13)		
GPD	1.56	(0.10)	1.60	(0.11)		

Table 3.16: Relative MSE of Original ES Estimates (Sample Size = 1000)



%	ML	HS	GPD
Simple	25.35 - (2.65)	38.77 - (3.07)	34.67 - (2.80)
GARCH-N	1.59 - (0.07)	1.14 - (0.08)	1.12 - (0.07)
EWMA-N	8.83 - (0.98)	12.29 - (1.26)	11.13 - (1.16)
GARCH-t	1.30 - (0.04)	0.78 - (0.04)	0.80 - (0.03)

HS, and HS is likely to outperform ML, and that for Simple and EWMA-N, ML may be the best and HS may be the worst. From Table 3.12 for Simple, it shows that EB reduces ReBias for HS and also reduces ReMSE simultaneously, but the reduction is insignificant, according to the standard errors. From Table 3.13 for GARCH-N, although OB reduce ReBias for ML and GPD insignificantly, according to the standard errors, and it does not reduce ReMSE for any of them. From Table 3.14 for EWMA-N, bias adjustment fails to reduce ReBias but surprisingly EB reduces a quite large amount of ReMSE for HS significantly. From Table 3.15 for GARCH-t, although OB reduces ReBias for ML and GPD insignificantly, according to the standard errors, and it does not reduce ReMSE for any of them. However, it is surprising that EB reduces ReMSE for EWMA even though it fails to reduce ReBias.

Results for Sample of Size 1000

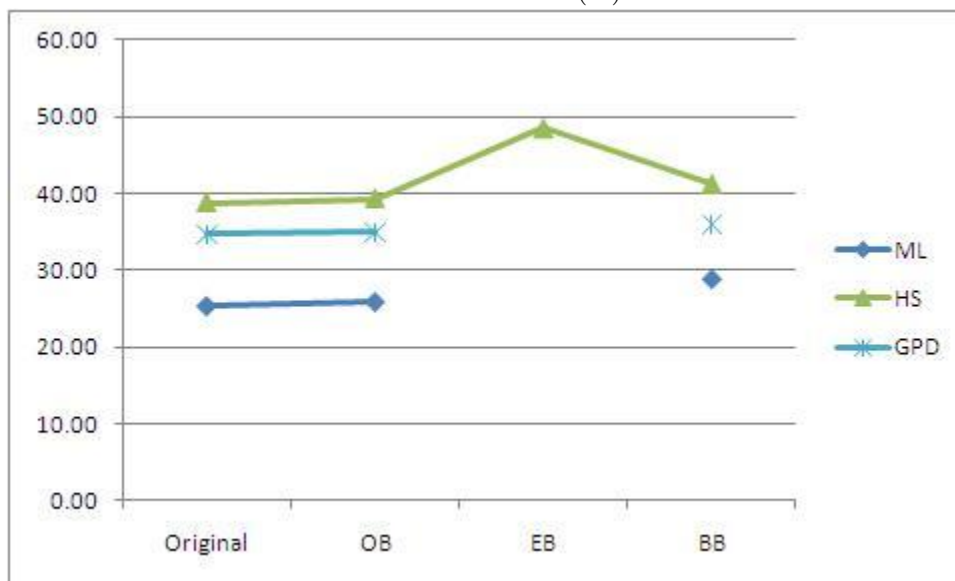
From the result for the original ES estimates in Table 3.16, it shows that ML is likely to be the best for Simple and EWMA-N while GPD is likely to be in the second place. In addition, GPD is

Table 3.17: Relative Accuracy Measures of Simple ES Estimates (Sample Size = 1000)

1. Relative Bias (%)

%	Original		OB		EB		BB	
ML	17.53	(1.49)	17.90	(1.51)			19.08	(1.59)
HS	33.86	(1.65)	34.26	(1.66)	40.28	(1.80)	34.99	(1.70)
GPD	29.98	(1.60)	30.27	(1.61)			30.78	(1.63)

2. Relative MSE (%)

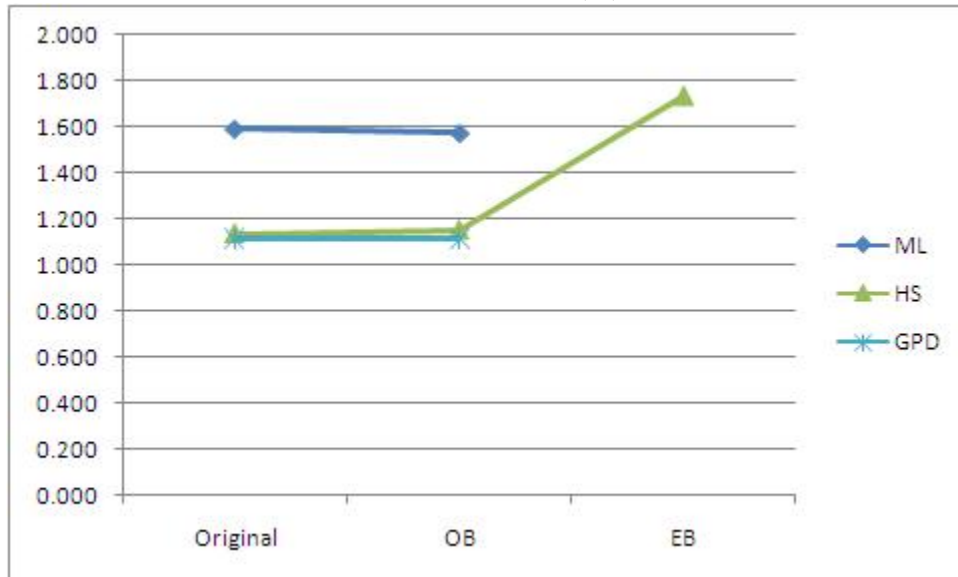


%	Original		OB		EB		BB	
ML	25.35	(2.65)	25.88	(2.72)			28.81	(3.40)
HS	38.77	(3.07)	39.32	(3.12)	48.50	(3.88)	41.21	(3.49)
GPD	34.67	(2.80)	34.96	(2.81)			35.89	(2.98)

Table 3.18: Relative Accuracy Measures of ES Estimates for GARCH-N (Sample Size = 1000)

	Original	OB	EB	
1. Relative Bias (%)	ML	-9.44 (0.27)	<u>-9.29</u> (0.27)	1.44 (0.41)
	HS	0.93 (0.34)	1.20 (0.34)	
	GPD	-2.45 (0.33)	<u>-2.26</u> (0.33)	

2. Relative MSE (%)

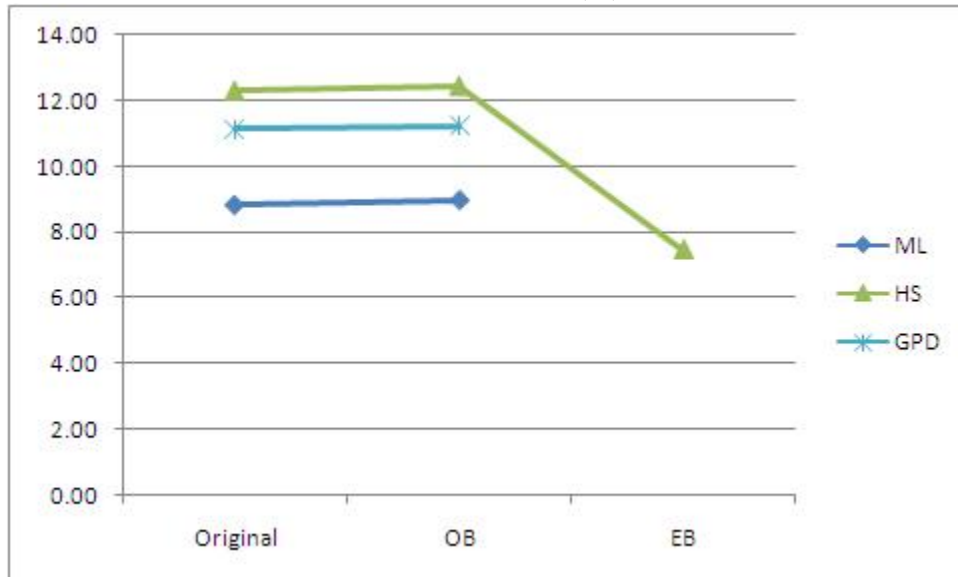


	Original	OB	EB
ML	1.59 (0.07)	<u>1.57</u> (0.07)	1.74 (0.18)
HS	1.14 (0.08)	1.15 (0.09)	
GPD	1.12 (0.07)	1.12 (0.07)	

Table 3.19: Relative Accuracy Measures of ES Estimates for EWMA (Sample Size=1000)

		Original		OB		EB	
1. Relative Bias (%)	ML	-1.25	(0.94)	<u>-1.02</u>	(0.95)		
	HS	11.15	(1.05)	11.48	(1.06)	<u>-8.04</u>	(0.83)
	GPD	7.96	(1.03)	8.18	(1.03)		

2. Relative MSE (%)

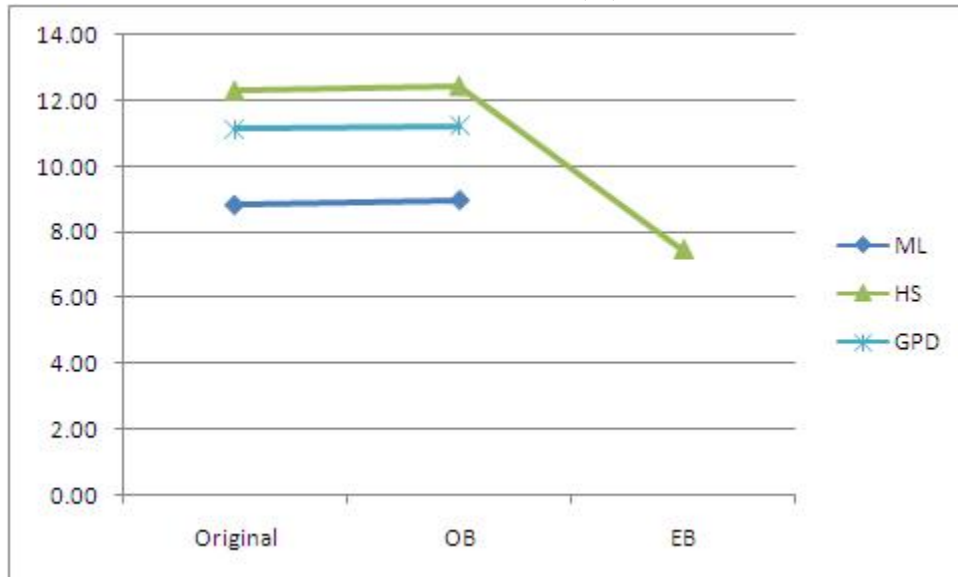


		Original		OB		EB	
ML	8.83	(0.98)	8.98	(1.02)			
HS	12.29	(1.26)	12.44	(1.27)	<u>7.46</u>	(0.55)	
GPD	11.13	(1.16)	11.24	(1.17)			

Table 3.20: Relative Accuracy Measures of ES Estimates for GARCH-t (Sample Size = 1000)

	Original	OB	EB	
1. Relative Bias (%)	ML	-9.34 (0.21)	<u>-9.14</u> (0.21)	3.30 (0.32)
	HS	0.74 (0.28)	0.99 (0.28)	
	GPD	-2.74 (0.27)	<u>-2.56</u> (0.27)	

2. Relative MSE (%)



	Original	OB	EB
ML	1.30 (0.04)	<u>1.28</u> (0.04)	1.13 (0.06)
HS	0.78 (0.04)	0.79 (0.04)	
GPD	0.80 (0.03)	0.80 (0.04)	

likely to outperform HS for GARCH-N while HS is likely to outperform ML surprisingly. Lastly, HS is likely to dominate GPD for GARCH-t while GPD is likely to dominate ML surprisingly. Table 3.17 shows that for Simple, bias adjustment fails to reduce either ReBias or ReMSE. From Table 3.18 for GARCH-N, it shows that OB reduces both ReBias and ReMSE for ML and GPD insignificantly, according to their standard errors. From Table 3.19 for EWMA-N, it shows that EB reduces both ReBias and ReMSE for HS significantly, according to their standard errors, while OB reduce ReBias for ML insignificantly, according to their standard errors, and fails to reduce ReMSE. From Table 3.19 for EWMA-N, it shows that OB slightly reduces both ReBias and ReMSE for ML and GPD, but reduces ReMSE only for ML slightly; however, each reduction is not significant, according to the standard errors.

Conclusion

Bias adjustment by the exact bootstrap for two-stage analysis, EWMA with HS, works pretty well for sample size 1,000, according to the largely reduced relative bias and relative MSE. For the method of EWMA with HS, the relative MSEs are largely reduced by the exact bootstrap for sample sizes of 250 and 500 surprisingly, although their relative biases are not reduced. Therefore, it is still worthwhile to do bias adjustment for the method of EWMA with HS, because most data from financial field have high frequency and EWMA has been widely used.

3.2 Data-based Criteria (Backtesting)

The purpose of backtesting a risk measure in practice is to test if a given procedure of predicting the risk measure deserves the label “good” based on the available data. Thus those tests for backtesting ES lie in the category of data-based criteria to evaluate the performance of a prediction of ES in the past based on historical data from the real world. The underlying heuristic is that any estimator which performed well in the past is likely to perform well in the future. Suppose that the underlying time series is a GARCH(p,q) process of the form $X_t = \sigma_t \varepsilon_t$. Let

$$I_{t+1} = I_{\{X_{t+1} > \text{VaR}_\alpha(X_{t+1}|\mathcal{F}_t)\}} \quad (3.12)$$

and

$$Y_{t+1} = (X_{t+1} - \text{ES}_\alpha(X_{t+1}|\mathcal{F}_t))I_{t+1}. \quad (3.13)$$

Define the violation indicators

$$\widehat{I}_{t+1} := I_{\{X_{t+1} > \widehat{\text{VaR}}_\alpha(X_{t+1}|\mathcal{F}_t)\}}$$

Then for an arbitrary process $\{X_t\}_{t \in \mathbb{Z}}$ the process $\{Y_t\}_{t \in \mathbb{Z}}$ forms a martingale difference series satisfying $E(Y_{t+1}|\mathcal{F}_t) = 0$. Furthermore, if $\{X_t\}_{t \in \mathbb{Z}}$ is an ARMA-GARCH process as defined in Definition 2.20, then it is easy to show that

$$Y_{t+1} = \sigma_{t+1}(\varepsilon_{t+1} - \text{ES}_\alpha(\varepsilon_{t+1}))I_{\{\varepsilon_{t+1} > \text{VaR}_\alpha(\varepsilon_t)\}}.$$

In practice, this suggests that when ES and volatility are estimated, the *violation residuals* of the following form could be formed

$$\widehat{R}_{t+1} = \widehat{Y}_{t+1}/\widehat{\sigma}_{t+1}$$

where

$$\widehat{Y}_{t+1} = [X_{t+1} - \widehat{\text{ES}}_\alpha(X_{t+1}|\mathcal{F}_t)]\widehat{I}_{t+1}.$$

These violation residuals are expected to behave like realizations of i.i.d. random variables from a distribution with mean zero and an atom of probability mass of size α at zero. The above preliminary of backtesting ES for GARCH models can be found in McNeil et al. (2005).

Accordingly, to backtest ES, the sequence of violation residuals should be examined whether or not it is i.i.d. distributed with mean zero and an atom of probability mass of size α at zero. By observing the sequence of violation residuals $\{\widehat{R}_t\}$ where $t \in \{1, 2, \dots, T\}$, it can be recorded by the sequence of violation indicators along with non-zero violation residuals. For example, a sequence of violation residuals $(0, 0, 0, x, 0, 0, 0, 0, 0, 0, y, 0, 0, 0, 0)$ can be recorded by the sequence of violation indicators $(0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0)$ and the extraction of non-zero violation residuals, (x, y) . The sequence of violation indicators is usually called *hit sequence* in related literature of backtesting VaR since it provides the information of whether a single data value exceeds the VaR or not. For convenience, let $p_0 = 1 - \alpha$ be the unconditional coverage rate, ES_α^t denote $\text{ES}_\alpha(X_{t+1}|\mathcal{F}_t)$, which is the one-step-forward predicted ES at time t , and VaR_α^t represent the one-step-forward predicted VaR at time t , namely $\text{VaR}_\alpha(X_{t+1}|\mathcal{F}_t)$. The following sections will then discuss how to perform appropriate tests for backtesting estimates of ES under the setting, that T , n_1 and $n_0 = T - n_1$ denote the total number of violation residuals, the number of one's, and the number of zero's, respectively.

3.2.1 Tests of Unconditional Coverage

Since $P(R_{t+1} = 0) = \alpha = P(I_{t+1} = 0) = \alpha$ and $P(I_{t+1} = 1) = 1 - \alpha$, testing the null hypothesis that the hit sequence is a series of i.i.d. random variables from Bernoulli with probability of $(1 - \alpha)$ at one, maybe also be applicable to testing for the null hypothesis that the sequence of violation residuals is i.i.d. from the common distribution with an atom of probability mass of size α at zero. Hence it can be easily implemented by a Binomial test performed on the hit

sequence since $\sum_{t=1}^T I_t \sim \text{Binomial}(T, p_0)$ under the null hypothesis

$$H_0 : p = p_0.$$

In fact, this testing procedure has been widely used for the well-known unconditional coverage test for VaR, where $(1 - \alpha)$ represents the unconditional coverage rate.

Moreover, the likelihood ratio (LR) test by first-order Markov chain approach can be found in Christofferson (1998) as an alternative of unconditional coverage test for VaR and hence is applicable here for our purpose as well. The notion of this test is described as follows. Let the hit sequence be denoted by \mathbf{I} hereafter; that is, $\mathbf{I} = (I_1, I_2, \dots, I_T)'$. The likelihood function of \mathbf{I} is built by

$$L(\pi|\mathbf{I}) = \pi^{n_1}(1 - \pi)^{n_0}. \quad (3.14)$$

Let $\hat{\pi} = n_1/(n_0 + n_1)$ be the MLE of π . Then the test statistic of the form

$$Q_{uc}^{LR} = -2 \ln[L(p_0|\mathbf{I})/L(\hat{\pi}|\mathbf{I})] \quad (3.15)$$

is approximately distributed as a chi squared distribution with degree of freedom one under the null hypothesis

$$H_0 : \pi = p_0,$$

when T is sufficiently large.

Both of the unconditional coverage tests above do not take account of the alternative hypothesis that zeros and ones come clustered together in a time-dependent fashion. In the unconditional coverage tests, the order of zeros and ones in the hit sequence does not matter but only the total number of ones does. To make up for this deficiency, the tests of independence and joint tests of independence and correct coverage will be discussed in turn as the following two sections discuss.

3.2.2 Tests of Independence

Testing for no-clustering is equivalent to testing for independence on the sequence of violation residuals and can then be simplified to testing on the hit sequence. Those tests of independence for backtesting VaR, developed for testing for no-clustering on the hit sequence, accordingly can be used as a part of backtesting ES as well. In the literature of backtesting VaR, the first-order Markov tests (Christofferson, 1998) and the duration-based tests (Christofferson and Pelletier, 2004) are suggested to test the null hypothesis of no violation clustering upon the idea that clustered violations indicates a signal of risk model misspecification as violation clustering implies catastrophic capital losses which could result in bankruptcy. These tests

examine whether or not the likelihood of a violation occurred on day t depends on a violation on the previous day $t - 1$. If the underlying risk model accurately reflects the portfolio risk then the chance of day violation should be independent of whether or not yesterday's violation for daily data. Tests of independence will then be introduced along with duration-based tests as follows.

First-order Markov Test of Independence

The first-order Markov test can be found in Christofferson (1998) and Christofferson and Pelletier (2004). The first-order Markov test of independence tests the independence against an explicit alternative that the hit sequence follows a binary first-order Markov chain with transition probability matrix

$$\Pi = \begin{bmatrix} 1 - \pi_{01} & \pi_{01} \\ 1 - \pi_{11} & \pi_{11} \end{bmatrix}, \quad (3.16)$$

where $\pi_{ij} = \Pr(I_t = j | I_{t-1} = i)$ represents the probability of an i on day $t - 1$ being followed by a j on day t . The approximate likelihood function for the process \mathbf{I} is then given by

$$L(\Pi | \mathbf{I}) = (1 - \pi_{01})^{n_{00}} \pi_{01}^{n_{01}} (1 - \pi_{11})^{n_{10}} \pi_{11}^{n_{11}}, \quad (3.17)$$

where n_{ij} denotes the number of observations with the value i followed by j . It follows that the LR test statistic is given by

$$\begin{aligned} Q_{\text{ind}}^{\text{Markov}} &= -2 \ln [L(\hat{\Pi}_0 | \mathbf{I}) / L(\hat{\Pi} | \mathbf{I})] \\ &\xrightarrow{d} \chi_{(1)}^2 \quad \text{as } T \rightarrow \infty, \end{aligned} \quad (3.18)$$

under the null hypothesis

$$H_0 : \pi_{01} = \pi_{11},$$

where $\hat{\Pi}_0$ is the MLE for

$$\Pi_0 = \begin{bmatrix} 1 - \pi_0 & \pi_0 \\ 1 - \pi_0 & \pi_0 \end{bmatrix}$$

with $\hat{\pi}_0 = (n_{01} + n_{11}) / (n_{00} + n_{10} + n_{01} + n_{11})$ and $\hat{\Pi}$, the MLE for Π in the form of

$$\hat{\Pi} = \begin{bmatrix} \frac{n_{00}}{n_{00} + n_{01}} & \frac{n_{01}}{n_{00} + n_{01}} \\ \frac{n_{10}}{n_{10} + n_{11}} & \frac{n_{11}}{n_{10} + n_{11}} \end{bmatrix}.$$

Table 3.21: Contingency Table for Pearson's χ^2 Test of Independence

	$I_t = 0$	$I_t = 1$	
$I_{t-1} = 0$	n_{00}	n_{01}	$R_0 = n_{00} + n_{01}$
$I_{t-1} = 1$	n_{10}	n_{11}	$R_1 = n_{10} + n_{11}$
	$C_0 = n_{00} + n_{10}$	$C_1 = n_{01} + n_{11}$	N

Pearson's χ^2 Test of Independence

Unlike the first-order Markov test which assumes the hit sequence is of first-order Markov structure, an alternate test without additional assumptions for the alternative hypothesis will be first proposed as follows. Suppose that $p_{ij} = \Pr(I_{t-1} = i \cap I_t = j)$ for $i, j \in 0, 1$, which is a different setting from the first-order Markov test in the previous section. It is a Pearson's χ^2 test based on a 2×2 contingency table that records violations on adjacent days as in Table 3.21. Thus testing for independence is equivalent to testing the null hypothesis

$$H_0 : p_{0j} = p_{1j}; j = 1, 2$$

or simply

$$H_0 : p_{01} = p_{11}$$

as well.

Proposition 3.2. *The Pearson's Chi-squared statistic is given by*

$$\begin{aligned} Q_{ind}^{\chi^2} &= \sum_{i=0}^1 \sum_{j=0}^1 \frac{(n_{ij} - R_i C_j / N)^2}{R_i C_j / N} \\ &= \frac{N(n_{00}n_{11} - n_{01}n_{10})^2}{R_0 R_1 C_0 C_1} \\ &\xrightarrow{d} \chi_{(1)}^2 \quad \text{as } N \rightarrow \infty. \end{aligned} \tag{3.19}$$

The asymptotic distribution of $Q_{ind}^{\chi^2}$ is shown in Appendix G.

Duration-based Test of Independence

The first-order Markov test above would do a reasonable job of detecting the temporal dependence in the hit sequence which follows a simple first-order Markov chain. Now consider the daily data without loss of generality, then the following will introduce the duration-based test (Christofferson and Pelletier, 2004) with power against more general forms of dependence

but simple estimation of only a few parameters. Corresponding to violation clustering, dependence in the hit sequence may present an excessive number of relatively short no-hit durations (volatile periods) and relatively long no-hit durations (calm periods). The duration-based test is motivated by this intuition by considering the duration (no-hit duration) of time (in days) between two adjacent violations as

$$D_i = t_i - t_{i-1} \quad (3.20)$$

where t_i represents the day of violation number i . Under the null hypothesis of no clustering, the duration D is distributed as Geometric(p_0) with a mean duration of $1/p_0$ days and no-hit duration have no memory (memoryless). A duration distribution can be characterized by its hazard function defined as the probability of getting a violation at day D after $D - 1$ days have been gone without a violation. It implies a flat discrete hazard function for D as shown below

$$\kappa(x) := \frac{\Pr(D = x)}{1 - \sum_{j=1}^{x-1} \Pr(D = j)} = p_0. \quad (3.21)$$

Hence the test of independence could be transformed into testing

$$H_0 : a = 1 \quad \text{vs.} \quad H_1 : a \neq 1$$

by specifying the distribution of D as Weibull(a, b) being continuous time limit of the above discrete time process. In other words,

$$f(D = x|a, b) = abx^{a-1} \exp(-bx^a) \quad (3.22)$$

with a closed form of hazard function, namely

$$\kappa(x) = \frac{f(x)}{1 - F(x)} = abx^{a-1}. \quad (3.23)$$

Under the null hypothesis of $a = 1$, the Weibull distribution becomes the exponential distribution with a flat hazard function, while the Weibull will have a decreasing hazard function as $a < 1$, which corresponds to an excess of relatively short and an excess of relatively long no-hit durations. Accordingly, when $a < 1$, it could be viewed as evidence of misspecified volatility dynamics in the risk model.

Let $\bar{F}(x) = 1 - F(x)$ be the survival function and $N(T)$ be the total number of durations. Taking account of the fact that the first and the last observations may be censored, the log-

likelihood is then developed below

$$\begin{aligned}
l(\boldsymbol{\theta}|\mathbf{D}) &:= \ln L(\boldsymbol{\theta}|\mathbf{D}) \\
&= C_1 \ln \bar{F}(D_1) + (1 - C_1) \ln f(D_1) + \sum_{i=2}^{N(T)-1} \ln f(D_i) \\
&\quad + C_{N(T)} \ln \bar{F}(D_{N(T)}) + (1 - C_{N(T)}) \ln f(D_{N(T)}),
\end{aligned} \tag{3.24}$$

where $C_1 = 0$ if the first observation of the hit sequence starts with a 1 or $C_1 = 1$ otherwise, and likewise $C_{N(T)} = 0$ if the last observation of the hit sequence is a 1 or $C_1 = 1$ otherwise. Hence the likelihood-ratio (LR) test statistic is given by

$$\begin{aligned}
Q_{\text{ind}}^D &= -2 \cdot \log \left(\frac{\sup_{\boldsymbol{\theta} \in \Theta_0} L(\boldsymbol{\theta}|\mathbf{D})}{\sup_{\boldsymbol{\theta} \in \Theta} L(\boldsymbol{\theta}|\mathbf{D})} \right) \\
&= 2[\sup_{\boldsymbol{\theta} \in \Theta} l(\boldsymbol{\theta}|\mathbf{D}) - \sup_{\boldsymbol{\theta} \in \Theta_0} l(\boldsymbol{\theta}|\mathbf{D})],
\end{aligned} \tag{3.25}$$

where $\Theta_0 = \{\boldsymbol{\theta} : \boldsymbol{\theta} = (a, b)'; a = 1, b > 0\}$ and $\Theta = \{\boldsymbol{\theta} : \boldsymbol{\theta} = (a, b)'; a > 0, b > 0\}$. It may be worth mentioning here that

$$Q_{\text{ind}}^D \xrightarrow{d} \chi_{(1)}^2 \quad \text{as } N(T) \rightarrow \infty$$

although (Christofferson and Pelletier, 2004) did not point out this fact. Since the MLEs are no longer available in closed form, Q_{ind}^D has to be computed by numerically optimizing the log-likelihood $l(\boldsymbol{\theta}|\mathbf{D})$ for $\boldsymbol{\theta} \in \Theta_0$ and $\boldsymbol{\theta} \in \Theta$, separately. Furthermore, the optimization can be done by maximizing $l(\boldsymbol{\theta}|\mathbf{D})$ over only one parameter since for a given value of a , an explicit solution to b is of the form

$$\hat{b} = \left(\frac{N(T) - C_1 - C_{N(T)}}{\sum_{i=1}^{N(T)} D_i^a} \right)^{1/a}, \tag{3.26}$$

which attains the unique maximum of the log-likelihood for a fixed a by solving the equation

$$\frac{\partial}{\partial b} l(\boldsymbol{\theta}|\mathbf{D}) = 0$$

for b since its second derivative is negative for all $b > 0$.

Although the sample size of the hit sequence can be reasonably large, the scarcity of violations renders the effective sample size of the durations small. Christofferson and Pelletier (2004) suggested using Monte Carlo testing technique to overcome it for test of independence; however, we argued that it is suitable for the joint test of coverage and independence because of the null hypothesis. We therefore suggest the permutation testing technique, which is very useful to overcome it for the test of independence and will be introduced as follows. It com-

compares the observed test statistic $Q_0 = Q_{\text{ind}}^{\text{D}}$ to N independent realizations of the test statistic, $Q_i^*; i = 1, \dots, N$, each one of which is calculated from a shuffling of the hit sequence. The shuffling of the hit sequence is actually the resampling without replacement. According to Davison and Hinkley (1997), the p-value of the permutation test for discrete data is therefore obtained by

$$\hat{p}(Q_0) = \frac{k+1}{N+1} \quad (3.27)$$

where

$$k = \sum_{i=1}^N I_{\{Q_i^* \geq Q_0\}} \quad (3.28)$$

due to the duration sequence is counted in discrete times. Since the continuous distribution is used for the duration sequence which is counted in discrete times, to account for the discreteness bias, Christofferson and Pelletier (2004) also suggested the value k differently as

$$k = 1 - \sum_{i=1}^N I_{\{Q_i^* < Q_0\}} + \sum_{i=1}^N I_{\{Q_i^* = Q_0\}} I_{\{U_i^* \geq U_0\}} \quad (3.29)$$

where $U_i; 1, 2, \dots, N$ are drawn independently from a Uniform distribution on the interval $[0, 1]$.

3.2.3 Joint Tests of Coverage and Independence

A joint test of coverage and independence is suggested as a complete test of correct conditional coverage, which combines two tests, one for unconditional coverage and the other for independence. The tests of independence discussed in the foregoing could be modified into joint tests of conditional coverage and independence by forming the composite null hypothesis into constraints $h(\boldsymbol{\theta}) = \mathbf{0}$ of $\boldsymbol{\theta}$, vector of all parameters of the postulated distribution for durations. The following two sections will describe the details from this point of view and point out the fact that the test statistics are all asymptotically distributed as a χ^2 distribution with degree of freedom two.

First-order Markov Test of Coverage and Independence

The first-order Markov joint test of coverage and independence has been already suggested in (Christofferson, 1998) and Christofferson and Pelletier (2004). It becomes much simpler to start from the essential notion that the null hypothesis could be formed as a constraint of the parameters of the first-order Markov structure. That is, with $\boldsymbol{\theta} = (\pi_{01}, \pi_{11})'$, the null hypothesis

$$H_0 : \pi_{01} = \pi_{11} = p_0$$

is equivalent to the constraint

$$h(\boldsymbol{\theta}) = \begin{pmatrix} \pi_{01} - p_0 \\ \pi_{11} - p_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Recall that $p_0 = 1 - \alpha$ with the confidence level α and

$$\widehat{\Pi} = \begin{bmatrix} \frac{n_{00}}{n_{00}+n_{01}} & \frac{n_{01}}{n_{00}+n_{01}} \\ \frac{n_{10}}{n_{10}+n_{11}} & \frac{n_{11}}{n_{10}+n_{11}} \end{bmatrix},$$

It follows that the LR test statistic for the Markov joint test is of the form

$$\begin{aligned} Q_{\text{joint}}^{\text{Markov}} &= -2 \ln[L(\Pi_0^*|\mathbf{I})/L(\widehat{\Pi}|\mathbf{I})] \\ &\stackrel{d}{\rightarrow} \chi_{(2)}^2 \quad \text{as } T \rightarrow \infty, \end{aligned} \tag{3.30}$$

where

$$\Pi_0^* = \begin{bmatrix} 1 - p_0 & p_0 \\ 1 - p_0 & p_0 \end{bmatrix}.$$

Pearson's χ^2 Test of Coverage and Independence

For testing the null hypothesis

$$H_0 : p_{01} = p_{11} = p_0 = 1 - \alpha$$

or equivalently

$$H_0 : h(\boldsymbol{\theta}) = \begin{pmatrix} p_{01} - p_0 \\ p_{11} - p_0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

it is easy to modify the χ^2 test of independence into a joint χ^2 test of coverage and independence.

Proposition 3.3. *The Pearson's χ^2 test statistic for the joint test of coverage and independence is given by*

$$\begin{aligned} Q_{\text{joint}}^{\chi^2} &= \sum_{i=0}^1 \sum_{j=0}^1 \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \\ &\stackrel{d}{\rightarrow} \chi_{(2)}^2 \quad \text{as } N \rightarrow \infty \end{aligned} \tag{3.31}$$

where $e_{ij} = R_i \cdot p_j$ with $p_j = \alpha^{1-j}(1 - \alpha)^j$ for $j \in \{0, 1\}$ according to the contingency table, Table 3.21.

The asymptotic distribution of $Q_{\text{joint}}^{\chi^2}$ is shown in Appendix H.

Duration-based Test of Coverage and Independence

Parallel to the first-order Markov joint test above, the joint test of unconditional coverage and independence extended from the duration-based approach in Section 3.2.2 will be first proposed as follows. Since under $a = 1$, the flat hazard function has the value b from the Weibull distribution and the fixed value p_0 theoretically from the setting of the backtesting problem, it follows that $b = p_0$. Hence the null hypothesis becomes $H_0 : a = 1, b = p_0$. Let $\theta = (a, b)'$ and then the null hypothesis, $H_0 : \theta = (1, p_0)'$, can be thought of as a constraint

$$h(\boldsymbol{\theta}) := \boldsymbol{\theta} - \boldsymbol{\theta}_0^* = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

where

$$\boldsymbol{\theta}_0^* = (1, p_0)'.$$

Accordingly, the implementing procedure of the duration-based test in Section 3.2.2 remains the same overall except that Q_{ind}^D in (3.18) has to be replaced by

$$\begin{aligned} Q_{\text{joint}}^D &= 2[\sup_{\boldsymbol{\theta} \in \Theta} l(\boldsymbol{\theta}|\mathbf{D}) - l(\boldsymbol{\theta}_0^*|\mathbf{D})] \\ &\stackrel{d}{\rightarrow} \chi_{(2)}^2 \quad \text{as } N(T) \rightarrow \infty, \end{aligned} \tag{3.32}$$

and that the permutation testing technique has to be replaced by Monte Carlo testing technique such that it compares the observed test statistic $Q_0 = Q_{\text{joint}}^D$ to N independent realizations of the test statistic, Q_i^* ; $i = 1, \dots, N$, which are calculated from i.i.d. Bernoulli(p_0) samples.

3.2.4 The Bootstrap Test of Zero-mean Behaviour

To test for the zero-mean behaviour of violation residuals, a bootstrap test (Efron and Tibshirani, 1993, p. 224, Section 16.4) is appropriate to be performed on the sequence of violation residuals since it makes no assumption about the distribution. Let \bar{R} be the average of violation residuals and $\hat{\sigma}$ the standard error of them. Recall that T denotes the total number of violation residuals, and calculate the statistic t_{obs} by

$$t_{obs} := t(\mathbf{R}) = \frac{\bar{R} - \mu}{\hat{\sigma}/\sqrt{T}}$$

where \mathbf{R} represents the sequence of violation residuals and $\mu = 0$ under null hypothesis. To proceed the bootstrap test, generate B bootstrap replications of violation residuals $\mathbf{R}_1^*, \mathbf{R}_2^*, \dots, \mathbf{R}_B^*$.

For each bootstrap replication \mathbf{R}_b^* , calculate the statistic

$$t(\mathbf{R}_b^*) = \frac{\bar{R}^* - \bar{R}}{\hat{\sigma}^*/\sqrt{T}}$$

where \bar{R}^* is the mean of \mathbf{R}_b^* and $\hat{\sigma}^*$ the standard error of \mathbf{R}_b^* . Then compute the approximate *achieved significance level* (ASL) of the test for the two-sided alternatives by

$$\widehat{ASL} = \#\{t^2(\mathbf{R}_b^*) \geq t_{obs}^2\}/B,$$

where \widehat{ASL} is the approximation of the probability of observing at least that large a value when the null hypothesis is true. The smaller the value of \widehat{ASL} , the stronger the evidence against the null hypothesis.

3.2.5 Simulation Study of Backtesting

The simulation study has been carried out to compare those tests proposed in this chapter through examining their power functions. The total number of simulations is 4,000, for each comparison study, and there are 1,000 rolling windows of length 500 for each simulation. For each simulation, a stationary GARCH(1,1) process of length 1,500 is simulated in the form of

$$X_t = \sigma_t \varepsilon_t \quad \text{and} \quad \sigma_t^2 = c_0 + aX_{t-1}^2 + b\sigma_{t-1}^2$$

where $\varepsilon_t \sim N(0, 1)$, $b = 0.75$, $a = 4(1-b)/5$ and $c_0 = v(1-a-b)$ with $v = 0.00004$, which is the long-term variance such that the long-term volatility is square root of it, 0.63%. The confidence level of ES is set to be $\alpha = 0.95$ for estimating ES and ESC, such that the unconditional coverage rate is $1 - \alpha = 0.05$, and the significance level of a test to be 0.05, which is also known as the size of the test or Type I error. Since for a GARCH model, the estimate of ES based on a simple Gaussian model (GM) is known for violating independence property while the pseudo estimate of ES (the true value of ES) is supposed to satisfy the independence property and has the correct coverage rate, which equals the confidence level α of ES, these two types of ES estimates are considered here to examine the power functions of those available tests. The simulation results are reported in the following; that is the values of the power function was displayed, along with their standard errors in brackets.

Comparison of Power Functions for Tests of Unconditional Coverage

The null hypothesis for the test of unconditional coverage is

$$H_0 : \text{coverage rate} = 0.05$$

The first column in Table 3.22 and Table 3.23 displays the size of the test under the above null hypothesis, since in this case, both the pseudo and GM estimates of VaR are estimated at the same coverage rate 0.05, the one in the null hypothesis. The rest of the columns in Table 3.22 and Table 3.23 reports the power of the test, since in those cases, the pseudo and GM estimates of VaR are estimated at different coverage rates other than 0.05, the one assumed in the null hypothesis. Figure 3.1 and 3.2 are plots of Table 3.22 and Table 3.23, respectively. They all show that those powers increase towards one, as the coverage rate goes up. The result suggests that both approaches detect the wrong coverage rate pretty well.

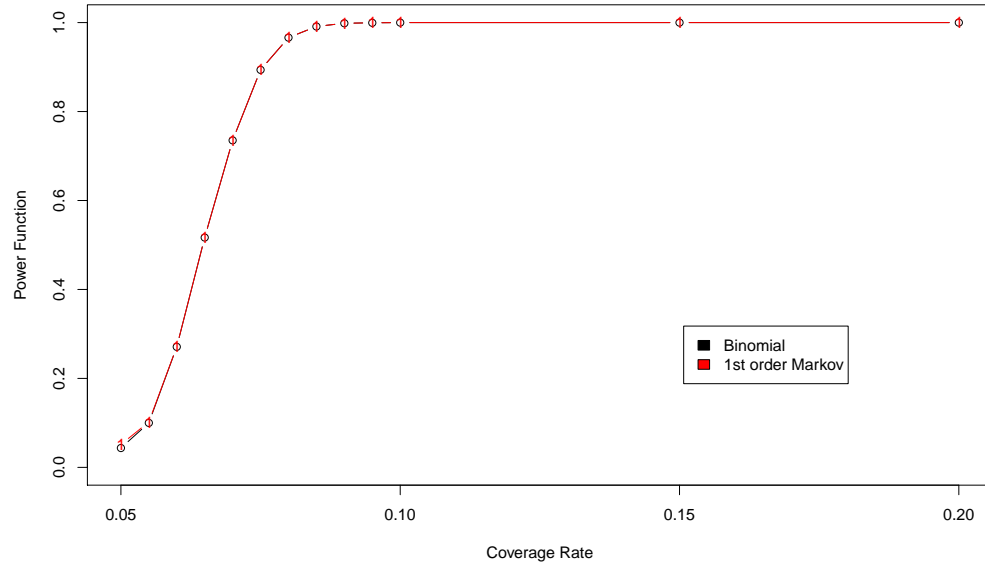


Figure 3.1: Sizes and Powers for Tests of Unconditional Coverage (Pseudo Estimate of VaR)

Table 3.22: Sizes and Powers for Tests of Unconditional Coverage (Pseudo Estimate of VaR)

Coverage Rate	0.05	0.055	0.06	0.065	0.07	0.075	0.08	0.085	0.09	0.095	0.1	0.15	0.2
Binomial	0.044 (0.003)	0.1 (0.005)	0.271 (0.007)	0.517 (0.008)	0.735 (0.007)	0.894 (0.005)	0.966 (0.003)	0.991 (0.001)	0.998 (0.001)	1 (0.000)	1 (0.000)	1 (0.000)	1 (0.000)
1st order Markov	0.052 (0.003)	0.102 (0.005)	0.271 (0.007)	0.517 (0.008)	0.735 (0.007)	0.894 (0.005)	0.966 (0.003)	0.991 (0.001)	0.998 (0.001)	1 (0.000)	1 (0.000)	1 (0.000)	1 (0.000)

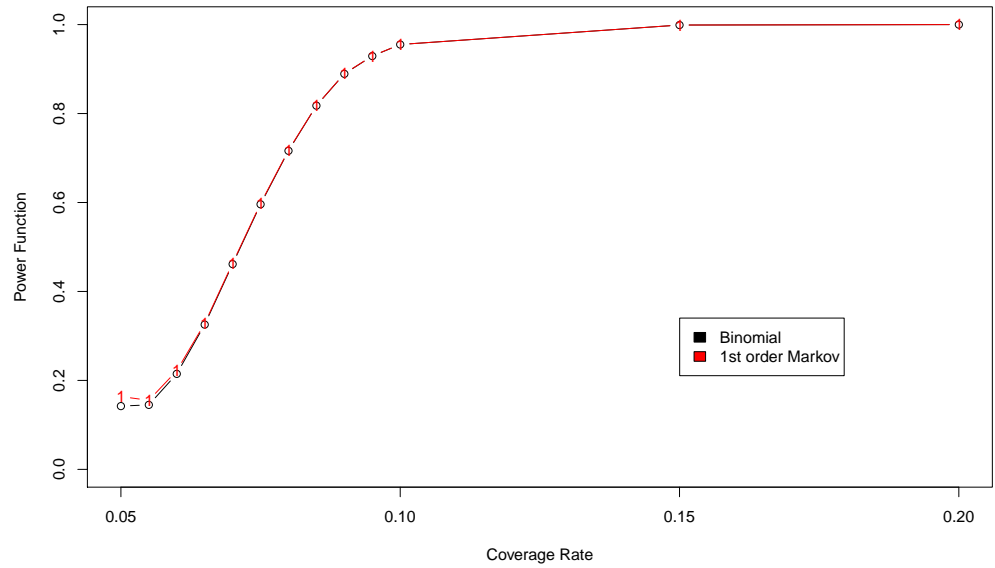


Figure 3.2: Sizes and Powers for Tests of Unconditional Coverage (GM Estimate of VaR)

Table 3.23: Sizes and Powers for Tests of Unconditional Coverage (GM Estimate of VaR)

Coverage Rate	0.05	0.055	0.06	0.065	0.07	0.075	0.08	0.085	0.09	0.095	0.1	0.15	0.2
Binomial	0.142 (0.006)	0.145 (0.006)	0.215 (0.006)	0.326 (0.007)	0.462 (0.008)	0.596 (0.008)	0.716 (0.007)	0.818 (0.006)	0.889 (0.005)	0.929 (0.004)	0.955 (0.003)	0.999 (0.000)	1 (0.000)
1st order Markov	0.164 (0.006)	0.155 (0.006)	0.222 (0.007)	0.328 (0.007)	0.464 (0.008)	0.598 (0.008)	0.718 (0.007)	0.818 (0.006)	0.89 (0.005)	0.93 (0.004)	0.955 (0.003)	0.999 (0.000)	1 (0.000)

Comparisons of Power Functions for Tests of Independence

The null hypothesis for the test of independence is that the sequence of the violation residuals or the hit sequence has no clustering. Figure 3.3 and Table 3.24 displays sizes of the test, according to the fact that pseudo estimates of VaR along the time horizon are independent. Those dash lines draw confident bands for the sizes of the test and the blue dash line draws the true size. Those sizes are all around the significance level of the test 0.05, but several confidence intervals of them exclude the true size, 0.05, which may not be good, although most of them are reasonable overall. Figure 3.4 and Table 3.25 presents the power of the test, according to the fact that the hits of GM estimates of VaR tends to have clustering under the GARCH model.

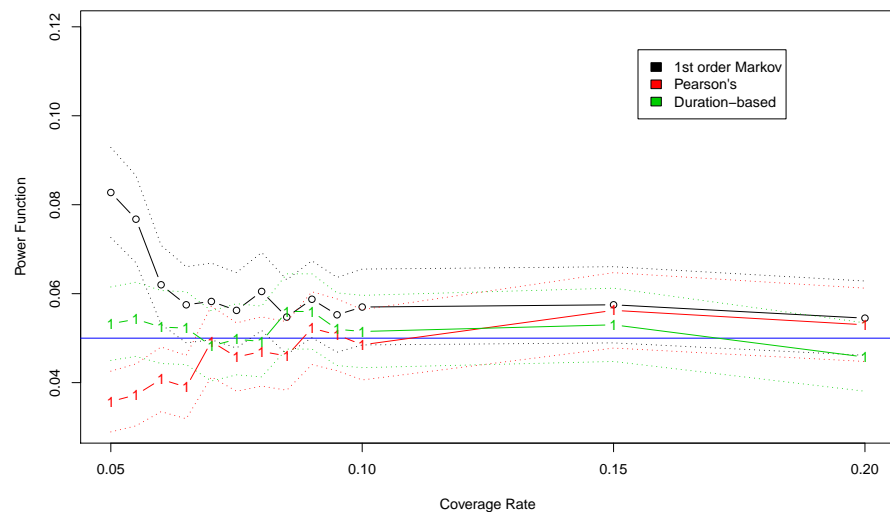


Figure 3.3: Sizes for Tests of Independence (Pseudo Estimate of VaR)

Table 3.24: Sizes for Tests of Independence (Pseudo Estimate of VaR)

Coverage Rate	0.05	0.055	0.06	0.065	0.07	0.075	0.08	0.085	0.09	0.095	0.1	0.15	0.2
1st order Markov	0.083 (0.004)	0.077 (0.004)	0.062 (0.004)	0.058 (0.004)	0.058 (0.004)	0.056 (0.004)	0.06 (0.004)	0.055 (0.004)	0.059 (0.004)	0.055 (0.004)	0.057 (0.004)	0.058 (0.004)	0.054 (0.004)
Pearson's	0.036 (0.003)	0.037 (0.003)	0.041 (0.003)	0.039 (0.003)	0.049 (0.003)	0.046 (0.003)	0.047 (0.003)	0.046 (0.003)	0.052 (0.004)	0.051 (0.003)	0.048 (0.003)	0.056 (0.004)	0.053 (0.004)
Duration-based	0.053 (0.004)	0.054 (0.004)	0.052 (0.004)	0.052 (0.004)	0.048 (0.003)	0.05 (0.003)	0.049 (0.003)	0.056 (0.004)	0.056 (0.004)	0.052 (0.004)	0.052 (0.003)	0.053 (0.004)	0.046 (0.003)

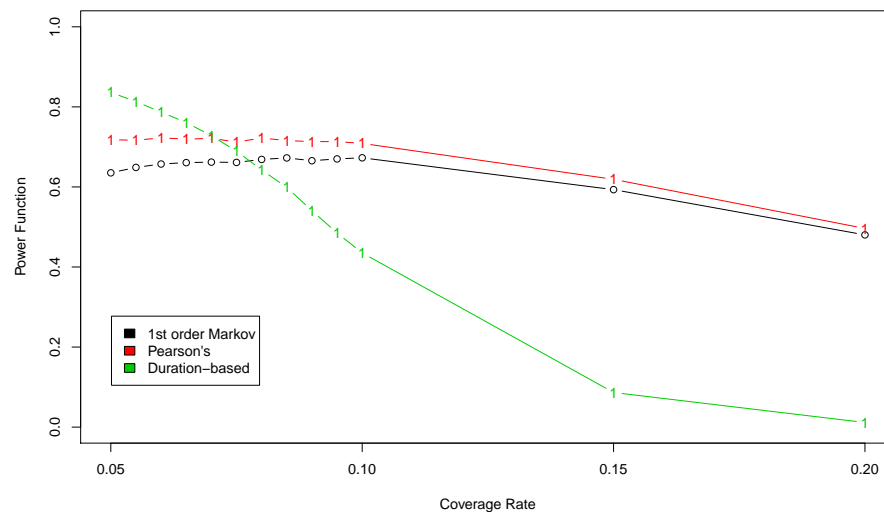


Figure 3.4: Powers for Tests of Independence (GM Estimate of VaR)

Table 3.25: Powers for Tests of Independence (GM Estimate of VaR)

Coverage Rate	0.05	0.055	0.06	0.065	0.07	0.075	0.08	0.085	0.09	0.095	0.1	0.15	0.2
1st order Markov	0.635 (0.008)	0.649 (0.008)	0.657 (0.008)	0.661 (0.007)	0.662 (0.007)	0.661 (0.007)	0.669 (0.007)	0.672 (0.007)	0.666 (0.007)	0.67 (0.007)	0.673 (0.007)	0.593 (0.008)	0.48 (0.008)
Pearson's	0.718 (0.007)	0.716 (0.007)	0.723 (0.007)	0.719 (0.007)	0.722 (0.007)	0.712 (0.007)	0.722 (0.007)	0.716 (0.007)	0.713 (0.007)	0.713 (0.007)	0.709 (0.007)	0.619 (0.008)	0.496 (0.008)
Duration-based	0.836 (0.006)	0.813 (0.006)	0.787 (0.006)	0.761 (0.007)	0.728 (0.007)	0.69 (0.007)	0.642 (0.008)	0.599 (0.008)	0.541 (0.008)	0.485 (0.008)	0.435 (0.008)	0.086 (0.004)	0.012 (0.002)

Comparisons of Power Functions for Joint Tests of Coverage and Independence

The null hypothesis is that the coverage rate is 0.05 and that the hit sequence has no clustering. The first column in Table 3.26 presents the size of the test, since the pseudo estimate of VaR is computed at the correct coverage rate 0.05 and its hit sequence is independent. Otherwise, their power all increase dramatically as the coverage rate increases between 0.05 and 0.10. Figure 3.6 and Table 3.27 displays the power of the test, according to the fact that GM estimates of VaR are not independent. As the coverage rate goes up, those powers increase as well towards one.

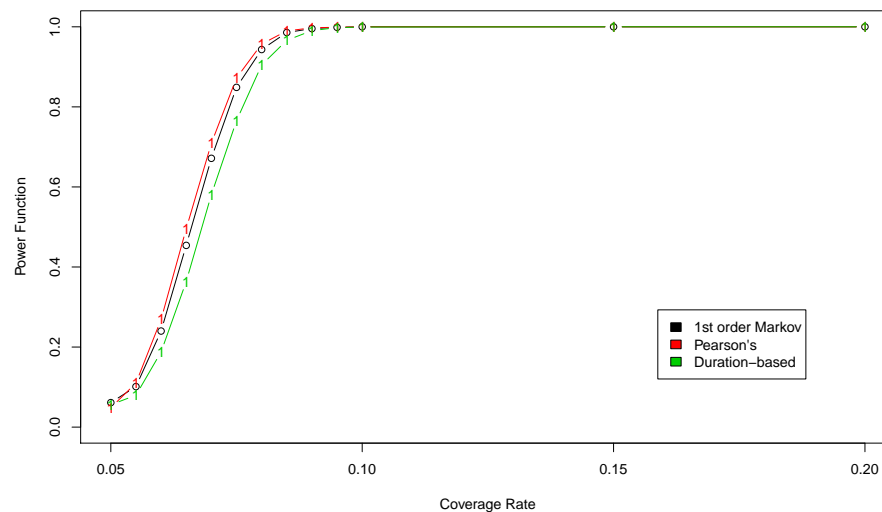


Figure 3.5: Sizes and Powers for Joint Tests of Coverage and Independence (Pseudo Estimate of VaR)

Table 3.26: Sizes and Powers for Joint Tests of Coverage and Independence (Pseudo Estimate of VaR)

Coverage Rate	0.05	0.055	0.06	0.065	0.07	0.075	0.08	0.085	0.09	0.095	0.1	0.15	0.2
1st order Markov	0.061 (0.004)	0.101 (0.005)	0.24 (0.007)	0.454 (0.008)	0.672 (0.007)	0.849 (0.006)	0.943 (0.004)	0.986 (0.002)	0.996 (0.001)	0.999 (0.001)	1 (0.000)	1 (0.000)	1 (0.000)
Pearson's	0.048 (0.003)	0.11 (0.005)	0.269 (0.007)	0.496 (0.008)	0.709 (0.007)	0.872 (0.005)	0.957 (0.003)	0.99 (0.002)	0.998 (0.001)	0.999 (0.000)	1 (0.000)	1 (0.000)	1 (0.000)
Duration-based	0.056 (0.004)	0.08 (0.004)	0.188 (0.006)	0.364 (0.008)	0.581 (0.008)	0.766 (0.007)	0.905 (0.005)	0.966 (0.003)	0.99 (0.002)	0.998 (0.001)	1 (0.000)	1 (0.000)	1 (0.000)

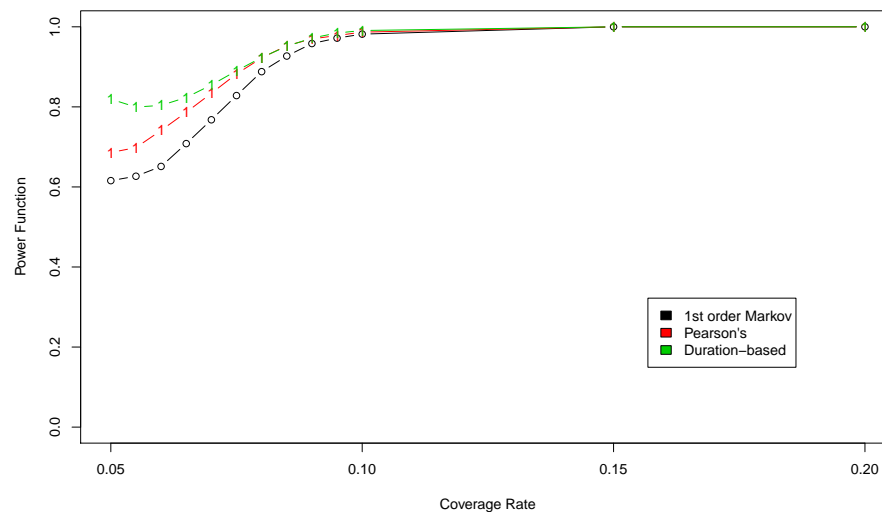


Figure 3.6: Powers for Joint Tests of Coverage and Independence (GM Estimate of VaR)

Table 3.27: Powers for Joint Tests of Coverage and Independence (GM Estimate of VaR)

Coverage Rate	0.05	0.055	0.06	0.065	0.07	0.075	0.08	0.085	0.09	0.095	0.1	0.15	0.2
1st order Markov	0.616 (0.008)	0.627 (0.008)	0.651 (0.008)	0.708 (0.007)	0.768 (0.007)	0.828 (0.006)	0.888 (0.005)	0.927 (0.004)	0.958 (0.003)	0.972 (0.003)	0.982 (0.002)	1 (0.000)	1 (0.000)
Pearson's	0.686 (0.007)	0.698 (0.007)	0.741 (0.007)	0.786 (0.006)	0.834 (0.006)	0.882 (0.005)	0.922 (0.004)	0.953 (0.003)	0.97 (0.003)	0.978 (0.002)	0.987 (0.002)	1 (0.000)	1 (0.000)
Duration-based	0.819 (0.006)	0.8 (0.006)	0.804 (0.006)	0.824 (0.006)	0.855 (0.006)	0.89 (0.005)	0.924 (0.004)	0.951 (0.003)	0.972 (0.003)	0.984 (0.002)	0.991 (0.002)	1 (0.000)	1 (0.000)

Conclusion

From the results of the simulation study, both proposed approaches, the Pearson's χ^2 test and the duration-based test for testing the coverage rate and independence jointly perform as well as those for testing independence. Hence the Pearson's χ^2 test could be a good alternative to the duration-based test. As a practical guideline, we recommend that it may be convenient to perform the joint tests first. If the null hypothesis of the joint tests is rejected, then it would be worthwhile to proceed to perform the test of the unconditional coverage and the test of independence separately to see if both null hypotheses or only one of them are rejected. Otherwise, it may not be necessary to perform all of them separately if the null hypothesis of the joint test is not rejected.

3.3 Summary of Contributions

The major contributions in this chapter are summarized below.

- Model-based criteria: (for evaluating the performance of bias adjustment)
 - Application of relative accuracy measures
 - Simulation study
- Data-based criteria (backtesting):
 - Pearson's χ^2 test for both test of independence and joint test of coverage and independence
 - Duration-based approach for both test of independence and joint test of coverage and independence
 - Simulation study for evaluating backtesting procedures.

Chapter 4

Multivariate Framework for Estimating ES and ESC

4.1 Introduction

Caillault and Guégan (2004) suggested several non-parametric and parametric models to compute VaR and ES: the RiskMetrics methodology (EWMA), the multivariate GARCH models (indeed, Vector GARCH models in McNeil et al. (2005)), the multivariate Markov-switching models, the empirical histogram and dynamic copulas. The ES and ESC formulas for multivariate Gaussian distributions have been derived in Landsman and Valdez (2003) and will be enhanced, exploited, and extended to multivariate changing volatility models. This chapter will comprehensively introduce the multivariate framework to estimate individual ES, aggregate ES, and ESC for multivariate elliptical distributions along with the conditional version of them for multivariate changing volatility models, multivariate GARCH (MGARCH) and EWMA. Section 4.2 will discuss the multivariate framework without any volatility filter and then Section 4.3 will cover multivariate changing volatility models. The last section will introduce the relative excess function of ESC parallel to that of ES for backtesting issue.

4.2 Methods Without a Volatility Filter

4.2.1 Multivariate Version of Historical Simulation

Recall that the formula of risk contributions derived from ES, ESC, is given by

$$\text{ESC}_\alpha(X_k) = E[X_k | S \geq \text{VaR}_\alpha(S)] = E[X_k | S \geq q_\alpha(S)] \quad ; k = 1, \dots, d,$$

where $F^{-1}(u) = q_u(S) = \inf\{s \in R : F(s) \geq u\} = \text{VaR}_u(S)$ is the lower u -th quantile (VaR at confidence level u) of S as an inverse function of F . Suppose that the random sample $\mathbf{S} = (S_1, \dots, S_n)'$, such that $S_i = \sum_{j=1}^d X_{i,j}$, $i = 1, 2, \dots, n$, is given with $\mathbf{S}_{:n} = (S_{(1)}, \dots, S_{(n)})'$ as its ordered sample. Parallel to the plug-in version of empirical estimator for ES in (2.12), the plug-in principle would also yield an empirical estimator for ESC as well. And when $\lfloor n\alpha \rfloor = n\alpha$, the empirical estimator is simply of the form in (4.1) as introduced in Proposition 4.1.

Proposition 4.1. (HS estimates of ESC). *For a specific business unit, say k , the empirical estimate of the expected shortfall contribution (ESC) of it is proposed herein as:*

$$\widehat{\text{ESC}}_\alpha(X_k|S) = \frac{1}{n(1-\alpha)} \sum_{i=\lfloor n\alpha \rfloor + 1}^n X_{\nu_i, k} \quad (4.1)$$

where ν_1, \dots, ν_n is a permutation such that $S_{(i)} = S_{\nu_i} = \sum_{j=1}^d X_{\nu_i, j}$. In other words, ν_i is the position of $S_{(i)}$ in the original sample $(S_1, \dots, S_n)'$.

The formula in (4.1) is in accordance with the ES formula in (2.13) when $\lfloor n\alpha \rfloor = n\alpha$. Let $\mathbf{c} = (n(1-\alpha))^{-1}(0, \dots, 0, 1, \dots, 1)'$ with zeros for the first $\lfloor n\alpha \rfloor$ elements. The HS estimator of ESC at confidence level α can be re-written in matrix form by

$$\widehat{\text{ESC}}_\alpha(X_k|S) = \mathbf{c}' \mathbf{X}_k, \quad (4.2)$$

where ν_1, \dots, ν_n are permutations such that $S_{(i)} = S_{\nu_i} = \sum_{j=1}^d X_{\nu_i, j}$ and

$$\mathbf{X}_k = (X_{\nu_1, k}, X_{\nu_2, k}, \dots, X_{\nu_n, k})'$$

is the vector of k -th components of $\mathbf{S}_{:n}$. In other words, ν_i is the position of $S_{(i)}$ in the original sample $(S_1, \dots, S_n)'$.

4.2.2 Multivariate Elliptical Distributions

The following theorem is originated from formulas of aggregate ES and ESC given in Landsman and Valdez (2003, Theorem 2 and Theorem 3) and extended from Theorem 2.18 for individual ES.

Theorem 4.2. (Formulas of ES and ESC for multivariate elliptical distributions). *Suppose that $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, g_d)$ where $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)'$ and $(\Sigma)_{ij} = \sigma_{ij}$. Denote the k -th component of \mathbf{X}_t by $X_{k,t} = \mathbf{e}'_k \mathbf{X}_t$, and the sum of all components by $S_t = \sum_{j=1}^d X_{j,t} = \mathbf{1}' \mathbf{X}_t$, where \mathbf{e}_k is the vector of zeros but one at the k -th location and $\mathbf{1} = (1, \dots, 1)'$ the vector of ones with dimension d . Let $\lambda_{g_1, \alpha}$ be the λ -factor, defined in Definition 2.17. Then under condition (2.37) which guarantees the existence of the mean vector such that $E(\mathbf{X}) = \boldsymbol{\mu}$:*

1. The individual ES of X_k for $k = 1, \dots, d$ is given by

$$\text{ES}_\alpha(X_k) = \mu_k + \lambda_{g_1, \alpha} \cdot \sigma_k^*, \quad (4.3)$$

where $\sigma_k^* = \sqrt{\sigma_{kk}}$.

2. Let $\tilde{\sigma}_{k,S} = \text{Cov}(X_k, S) = \sum_{j=1}^d \sigma_{kj}$. The aggregate ES of S can be represented by

$$\text{ES}_\alpha(S) = \mu_S + \lambda_{g_1, \alpha} \cdot \sigma_S^*, \quad (4.4)$$

where $\mu_S = \sum_{j=1}^d X_j$ and $\sigma_S^* = \sqrt{\sum_{i=1}^d \sum_{j=1}^d \sigma_{ij}} = \sqrt{\sum_{k=1}^d \tilde{\sigma}_{k,S}}$.

3. The risk contribution of X_k to the total ES, $\text{ESC}_\alpha(X_k)$, for $k = 1, \dots, d$ can be represented by

$$\begin{aligned} \text{ESC}_\alpha(X_k) &= \mu_k + \lambda_{g_1, \alpha} \cdot \sigma_k^* \cdot \rho_{k,S} \\ &= \mu_k + \lambda_{g_1, \alpha} \cdot \tilde{\sigma}_{k,S} / \sigma_S^*, \end{aligned} \quad (4.5)$$

where $\rho_{k,S} = \frac{\tilde{\sigma}_{k,S}}{\sigma_k^* \sigma_S^*}$.

The proof of Theorem 4.2 is provided in Appendix D. Under this development, ESC is definitely smaller than ES for each component and hence it obviously leads to satisfaction of the full allocation and no-undercut axioms of coherent and fair allocation principles, as the coherency and fairness of ESC has been proved. Obviously the formulas given above all depend on the same λ_α and hence the estimating problem of conditional ESC, individual ES, and aggregate ES is pulled back to be the ordinary problem of estimating model parameters for multivariate elliptical distributions. It means that estimated parameters will immediately yield estimates of ESC, individual ES, and aggregate ES through Equations (4.5), (4.3), and (4.4), respectively, by substituting corresponding estimates for parameters. The simplified forms of λ_α for Gaussian and Student's t distributions are the same as those in (2.54) and (2.55), respectively, discussed in Section 2.3.3.

As mentioned in Section 1.2.2, Panjer (2002) has shown that ESC, which is the Euler allocation derived from ES, coincides with the allocation formula of CAPM, which is literally a covariance-based allocation, in case of multivariate Gaussian distributions. By taking advantage of Theorem 4.2 and following the same argument in Panjer (2002), it is easy to show in Proposition 4.3 that ESC coincides to the allocation formula of CAPM for all elliptical distributions as well.

Proposition 4.3. (Identity of internal beta and beta of CAPM). *For all elliptical distributions, the internal beta, which is the ratio of ESC to total ES, is identical to the beta*

of CAPM, which represents the proportion of the total risk capital allocated back to the specific component.

4.3 Conditional ES and ESC for MGARCH Processes

The multivariate GARCH (MGARCH) process is defined as follows.

Definition 4.4. (Multivariate GARCH (MGARCH) processes). *Suppose that $\{\mathbf{Z}_t\}_{t \in \mathbb{Z}}$ is a vector of strict white noises, with mean $\mathbf{0}$ and variance-covariance matrix I_d . The process $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ is said to be a multivariate GARCH (MGARCH) process if it is strictly stationary and satisfies equations of the form*

$$\mathbf{X}_t = \Sigma_t^{1/2} \mathbf{Z}_t, \quad t \in \mathbb{Z}, \quad (4.6)$$

where $\Sigma_t^{1/2} \in \mathbb{R}^{d \times d}$ is the Cholesky factor of a positive-definite matrix Σ_t which is measurable with respect to $\mathcal{F}_{t-1} = \sigma\{\mathbf{X}_s : s \leq t-1\}$, the history of the process up to time $t-1$.

Any distribution with mean zero and variance-covariance matrix I_d is permissible for \mathbf{Z}_t . In practice the innovations are generally taken to be from either a multivariate Gaussian distribution or, more realistically for daily returns, an appropriate scaled spherical multivariate t distribution. In other words, \mathbf{Z}_t is either from $N_d(\mathbf{0}, I_d)$ or $t_d(\mathbf{0}, (r-2)I_d/r; r)$.

Now consider that the innovations of a MGARCH processes are from a multivariate elliptical distribution with mean $\mathbf{0}$ and variance-covariance matrix I_d such as multivariate Gaussian or multivariate t . Precisely, $\mathbf{Z}_t \sim E_d(\mathbf{0}, W_d, g_d)$ such that $\text{Cov}(\mathbf{Z}_t) = I_d$, providing that the density generator g_d exists. It follows that the distribution of \mathbf{X}_t conditional on \mathcal{F}_{t-1} turns out to be $E_d(\mathbf{0}, V_t, g_d)$ with

$$\begin{aligned} V_t &= \Sigma_t^{1/2} W_d (\Sigma_t^{1/2})' \\ &= \Sigma_t^{1/2} W_d \Sigma_t^{1/2}, \end{aligned}$$

providing that $\Sigma_t^{1/2}$ is symmetric. Note that a sequence of the form $\{\mathbf{v}' \cdot \mathbf{X}_t\}$ may not be a univariate GARCH process and that the distribution of this sequence depends on the structure of Σ_t and W_d . Let $v_{t,ij}$ denote the ij -th element of V_t , and $v_{t,k}^* = \sqrt{v_{t,kk}}$. Denote the k -th component of \mathbf{X}_t by $X_{k,t} = \mathbf{e}_k' \mathbf{X}_t$, and the sum of all components $S_t = \sum_{j=1}^d X_{j,t} = \mathbf{1}' \mathbf{X}_t$, where \mathbf{e}_k and $\mathbf{1}$ are defined in Section 4.2.2. The following results are thus obtained by exploiting Theorem 4.2.

Theorem 4.5. (Conditional ES and ESC for MGARCH processes with elliptical innovations). *Suppose that $\mathbf{Z}_t \sim E_d(\mathbf{0}, W_d, g_d)$ such that $\text{Cov}(\mathbf{Z}_t) = I_d$, providing that the*

density generator g_d exists. Assume that the condition (2.37), which guarantees the existence of the mean vector such that $E(\mathbf{Z}) = \mathbf{0}$, holds. Define $\lambda_{g,\alpha} = \frac{\overline{G}(\frac{1}{2}z_\alpha^2)}{1-\alpha}$ where $G(\cdot)$ is the cumulative generator defined in Definition 2.16 and $z_\alpha = \text{VaR}_\alpha(Z)$ with $Z \sim E(0, 1, g_1)$.

Then for a general MGARCH process $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ of the form $\mathbf{X}_t = \Sigma_t^{1/2} \mathbf{Z}_t$, we have the following results.

1. For $k = 1, 2, \dots, d$, the individual ES of $X_{k,t+1}$ conditional on the history \mathcal{F}_t is given by

$$\text{ES}_\alpha(X_{k,t+1} | \mathcal{F}_t) = \lambda_{g_1, \alpha} v_{t+1, k}^*, \quad (4.7)$$

where $v_{t+1, k}^* = \sqrt{v_{t+1, kk}}$.

2. Let $\tilde{v}_{t+1, kS} = \text{Cov}(X_{t+1, k}, S_{t+1}) = \sum_{j=1}^d v_{t+1, kj}$. The aggregate ES is of the form

$$\text{ES}_\alpha(S_{t+1} | \mathcal{F}_t) = \lambda_{g_1, \alpha} \tilde{v}_{S_{t+1}}^*, \quad (4.8)$$

where $\tilde{v}_{S_{t+1}}^* = \sqrt{\sum_{i=1}^d \sum_{j=1}^d v_{t+1, ij}} = \sqrt{\sum_{k=1}^d \tilde{v}_{t+1, kS}}$.

3. ESC for the k -th component of \mathbf{X}_{t+1} is of the form

$$\text{ESC}_\alpha(X_{k,t+1} | \mathcal{F}_t) = \lambda_{g_1, \alpha} v_{t+1, k}^* \cdot \rho_{t+1, k} \quad (4.9)$$

where $\rho_{t+1, k} = \frac{\tilde{v}_{t+1, kS}}{v_{t+1, k}^* \tilde{v}_{S_{t+1}}^*}$ or equivalently $\rho_{t+1, k} = \frac{\tilde{v}_{t+1, kS}}{v_{t+1, k}^* \sqrt{\sum_{i=1}^d \tilde{v}_{t+1, kS}}}$.

The proof of Theorem 4.5 is provided in Appendix F. Note that $\lambda_{g_1, \alpha}$ remains the same as given in (2.50) for a specified subclass of elliptical distributions with the same density generator g_1 .

4.3.1 Conditional ES and ESC for MGARCH-Gaussian Process

Suppose that $\mathbf{Z}_t \sim N_d(\mathbf{0}, I_d)$ such that $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ is a MGARCH-Gaussian process, and hence $\mathbf{X}_{t+1} | \mathcal{F}_t \sim N_d(\mathbf{0}, \Sigma_t)$. Denote $\sigma_{t, k}^* = \sqrt{\sigma_{t, kk}}$. Then applying Theorem 4.5 by replacing W_d with I_d and V_t with Σ_t such that $v_{t, ij} = \sigma_{t, ij}$, it gives the conditional ES and ESC for MGARCH-Gaussian processes as stated in Example 4.6 below.

Example 4.6. (Conditional ES and ESC for MGARCH-Gaussian processes). Let λ_α be the λ -factor given by (2.54) for Gaussian distributions. Then for the MGARCH-Gaussian process $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$, we have the following results.

1. For $k = 1, 2, \dots, d$, the individual ES of $X_{k,t+1}$ conditional on the history \mathcal{F}_t is given by

$$\text{ES}_\alpha(X_{k,t+1} | \mathcal{F}_T) = \sigma_{t+1, k}^* \cdot \lambda_\alpha. \quad (4.10)$$

2. Let $\tilde{\sigma}_{t+1,kS} = \text{Cov}(X_{k,t+1}, S_{t+1}) = \sum_{i=1}^d \sigma_{t+1,ki}$. The aggregate ES is of the form

$$\text{ES}_\alpha(S_{t+1}|\mathcal{F}_t) = \sigma_{S_{t+1}}^* \cdot \lambda_\alpha \quad (4.11)$$

where $\sigma_{S_{t+1}}^* = \sqrt{\sum_{i=1}^d \sum_{j=1}^d \sigma_{t+1,kj}} = \sqrt{\sum_{k=1}^d \tilde{\sigma}_{t+1,kS}}$.

3. ESC for the k -th component of \mathbf{X}_{t+1} is of the form

$$\begin{aligned} \text{ESC}_\alpha(X_{k,t+1}|\mathcal{F}_t) &= \lambda_\alpha \sigma_{t+1,k}^* \cdot \rho_{t+1,k} \\ &= \lambda_\alpha \tilde{\sigma}_{t+1,kS} / \sigma_{S_{t+1}}^*, \end{aligned} \quad (4.12)$$

where $\rho_{t+1,k} = \frac{\tilde{\sigma}_{t+1,kS}}{\sigma_{t+1,k}^* \sigma_{S_{t+1}}^*}$.

Note that λ_α remains the same as given in (2.54) for Gaussian distributions.

4.3.2 Conditional ES and ESC for MGARCH- t Processes

Now suppose that $\mathbf{Z}_t \sim t_d(\mathbf{0}, W_d)$ such that $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ is a MGARCH- t process where $W_d = \frac{r-2}{r} I_d$ such that $\text{Cov}(\mathbf{Z}_t) = I_d$. It follows that $\mathbf{X}_{t+1}|\mathcal{F}_t \sim t_d(\mathbf{0}, V_t)$ where $V_t = \frac{r-2}{r} \Sigma_t$. Then by Theorem 4.5 again by replacing W_d with $\frac{r-2}{r} I_d$ and V_t with $\frac{r-2}{r} \Sigma_t$ such that $v_{t,ij} = \frac{r-2}{r} \sigma_{t,ij}$, it gives the conditional ES and ESC for MGARCH- t processes as stated in Example 4.7 below.

Example 4.7. (Conditional ES and ESC for MGARCH- t processes). Let λ_α be the λ -factor given by (2.55) for t distributions. Then for a MGARCH- t process $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$, we have the following results.

1. For $k = 1, 2, \dots, d$, the individual ES of $X_{k,t+1}$ conditional on the history \mathcal{F}_t is given by

$$\text{ES}_\alpha(X_{k,t+1}|\mathcal{F}_T) = \eta_{r,\alpha} \sigma_{t+1,k}. \quad (4.13)$$

2. Let $\tilde{\sigma}_{t+1,kS} = ((r-2)/r) \text{Cov}(X_{k,t+1}, S_{t+1}) = \sum_{i=1}^d \sigma_{t+1,ki}$. The aggregate ES is of the form

$$\text{ES}_\alpha(S_{t+1}|\mathcal{F}_t) = \eta_{r,\alpha} \sigma_{S_{t+1}} \quad (4.14)$$

where $\sigma_{S_{t+1}} = \sqrt{\sum_{i=1}^d \sum_{j=1}^d \sigma_{t+1,kj}} = \sqrt{\sum_{k=1}^d \tilde{\sigma}_{t+1,kS}}$.

3. ESC for the k -th component of \mathbf{X}_{t+1} is to be of the form

$$\begin{aligned} \text{ESC}_\alpha(X_{k,t+1}|\mathcal{F}_t) &= \eta_{r,\alpha} \sigma_{t+1,k} \cdot \rho_{t+1,k} \\ &= \eta_{r,\alpha} \tilde{\sigma}_{t+1,kS} / \sigma_{S_{t+1}}, \end{aligned} \quad (4.15)$$

where $\rho_{t+1,k} = \frac{\tilde{\sigma}_{t+1,kS}}{\sigma_{t+1,k} \sigma_{S_{t+1}}}$.

Note also that λ_α remains the same as given in (2.55) for t distributions.

4.3.3 Estimation Strategies

Without any volatility filter, MLE of ES from a Gaussian model, and HS estimates of ESC are easy to compute from available data $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ as discussed in Section 4.2.1, while HS estimates of individual ES and aggregate ES are able to be obtained as discussed in Section 2.3.1. Note that HS method is directly applied to the original data rather than model residuals for MGARCH processes.

With a volatility filter, the estimation of conditional ESC, conditionally individual ES, and conditionally aggregate ES can be implemented by the general scheme discussed in Section 4.2.2, since the conditional distributions of MGARCH processes are assumed to be multivariate elliptical distributions. The first step is to fit a MGARCH model to data and then make one-step ahead prediction of the conditional variance-covariance matrix (Σ_{t+1}). The Baba-Engle-Kraft-Kroner (BEKK-GARCH) model (Engle and Kroner, 1995) will be considered to fit a MGARCH model for the empirical study in next chapter, because it not only can easily guarantee the positive definiteness of the conditional variance-covariance matrix but also has moderate number of parameters to be estimated. The conditional variance-covariance matrix of a d -dimensional BEKK-GARCH(p, q) is defined as

$$\Sigma_t = A_0 A_0' + \sum_{i=1}^p A_i \Sigma_{t-1} A_i' + \sum_{j=1}^q B_j Z_{t-j} Z_{t-j}' B_j',$$

where A_0, A_i and B_j are $d \times d$ matrices and A_0 is triangular for identification purposes. Nonetheless, the EWMA or exponential smoothing procedure could be used as a simple alternative to GARCH volatility prediction as seen in Section 2.4.2. It can then be extended to predict the the conditional variance-covariance matrix (Σ_{t+1}) one-step ahead. The multivariate analogue of the univariate EWMA updating scheme is given by

$$\widehat{\Sigma}_{t+1} = (1 - \lambda) \sum_{i=0}^{n-1} \lambda^i \mathbf{X}_{t-i} \mathbf{X}_{t-i}',$$

as is consistent with the idea of estimating Σ_{t+1} by a weighted sum of past values of the matrices $\mathbf{X}_t \mathbf{X}_t'$, where the weights decay exponentially. Then the multivariate analogue of the univariate EWMA recursive scheme (2.77) is given by

$$\widehat{\Sigma}_{t+1} = (1 - \lambda) \mathbf{X}_t \mathbf{X}_t' + \lambda \widehat{\Sigma}_t, \tag{4.16}$$

where λ is typically taken to be 0.94 by RiskMetrics. The second step is to substitute corre-

sponding estimates for model parameters and predictions for conditional variance-covariance matrix in Equations (4.9), (4.7) and (4.8), and hence obtain estimates of conditional ESC, conditionally individual ES, and conditionally aggregate ES, respectively.

4.4 Backtesting of ESC

Because backtesting is all about making sure that things that are supposed to happen in theory actually do happen in practice, it is necessary to implement backtesting of ESC. Given an arbitrary bivariate process $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ where $\mathbf{X}_t = (X_{1,t}, X_{2,t})'$, define

$$Y_{i,t} = [X_{i,t} - \text{ESC}_\alpha(X_{i,t}|\mathcal{F}_{t-1})] I_{\{S_t > \text{VaR}_\alpha(S_t|\mathcal{F}_{t-1})\}} \quad \text{for } i = 1, 2 \quad (4.17)$$

where $S_t = X_{1,t} + X_{2,t}$. Then the processes $\{Y_{i,t}\}_{t \in \mathbb{Z}}$; $i = 1, 2$ form two martingale difference series satisfying

$$E(Y_{i,t}|\mathcal{F}_{t-1}) = 0; \quad \forall i = 1, 2.$$

Let $\widehat{I}_t = I_{\{S_t > \widehat{\text{VaR}}_\alpha(S_t|\mathcal{F}_{t-1})\}}$. Then in practice, this follows an argument, similar to that of backtesting ES, that when ESC and volatility are estimated, then the sequences, given by

$$\widehat{Y}_{i,t} = [X_{i,t} - \widehat{\text{ESC}}_\alpha(X_{i,t}|\mathcal{F}_{t-1})]\widehat{I}_t; \quad \forall i = 1, 2 \quad (4.18)$$

are expected to behave like realizations of i.i.d. random variables from a distribution with mean zero and an atom of probability mass of size α at zero. Therefore, for each component $X_{i,t}$, the bootstrap test is applicable to be implemented on $\widehat{Y}_{i,t}$ in order to test for the zero-mean behavior of $\widehat{Y}_{i,t}$.

4.5 Summary of Contributions

- HS formulas of ESC (Landsman and Valdez, 2003)
- ESC formula for elliptical distributions (Landsman and Valdez, 2003) through λ -factor
- ES and ESC formulas for multivariate GARCH (MGARCH) processes: MGARCH-Gaussian and MGARCH- t
- Backtesting of ESC

Chapter 5

Empirical Study

As mentioned in the foregoing, it can be foreseen that the performance of parametric estimators of ES for elliptical distributions will be mostly driven by the performance of estimated parameters from model fitting, since ES of elliptical distributions can be explicitly expressed by a linear function of parameters of the underlying distribution. It is also true for conditional ES and ESC estimates of MGARCH processes with innovations of elliptical distributions, since the underlying distribution conditional on the past is to be elliptical as well. Besides, conditional ES could be estimated by a function of past realizations and estimated model parameters by Theorem 4.2, and so could conditional ESC.

The structure of the empirical study is shown in Figure 5.1. The empirical study investigates two Russell Indices, Russell 1000 Index and Russell 2000 Index, using their “Value Without Dividend” from June 1st, 1995 to December 31st, 2008 in a bivariate framework. Data have to be transformed into losses which are the negative of daily returns first for the purpose of interest. More precisely, the 3,427 losses are dated from June 2nd, 1995 to December 31st, 2008 of the form

$$X_t = -\log(Y_t/Y_{t-1})$$

where Y_t 's are the original data. The length of rolling window is chosen to be 1904 and the total number of windows is 1523 such that the last window will predict ES for December 31st, 2008 for the purpose of backtesting. The confidence level α of ES and ESC is set to be 0.95. The first window consisting of data from day 1 to day 1904 is used to predict ES for 1905th day as the first prediction. The second predicted ES is estimated by moving to the second window from day 2 to day 1905. Then keep moving the window one day forward and predict ES until reaching the last window. Hence the sequences of violation residuals and hit sequences can be easily calculated for backtesting.

Now, let us take a look at two graphics to have the first insight of the original data and the calculated losses. Figure 5.2 is the plot of the daily values for these two indexes. Figure 5.3 is

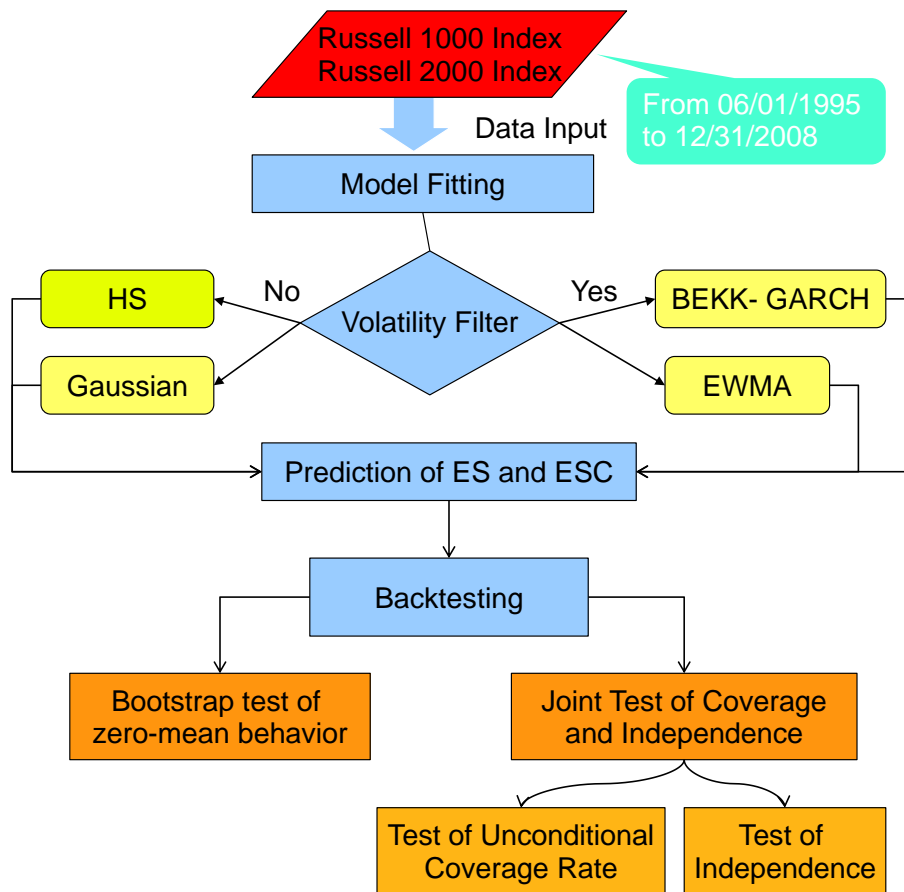


Figure 5.1: Structure of The Empirical Study

the plot of the daily losses for them and shows that to fit a model using a volatility filter such as GARCH is applicable.

Hereafter, the notation “HS”, “GM”, “BEKK” and “EWMA” stand for “Historical Simulation method”, “Gaussian model”, “BEKK-GARCH model”, “EWMA method”, respectively. Besides, “Russell 1000” and “Russell 2000” are short for “Russell 1000 Index” and “Russell 2000 Index”, respectively. “Aggregate Index” represents the sum of these two single indexes. The significance level is set to be 0.01 for all tests except as specified.

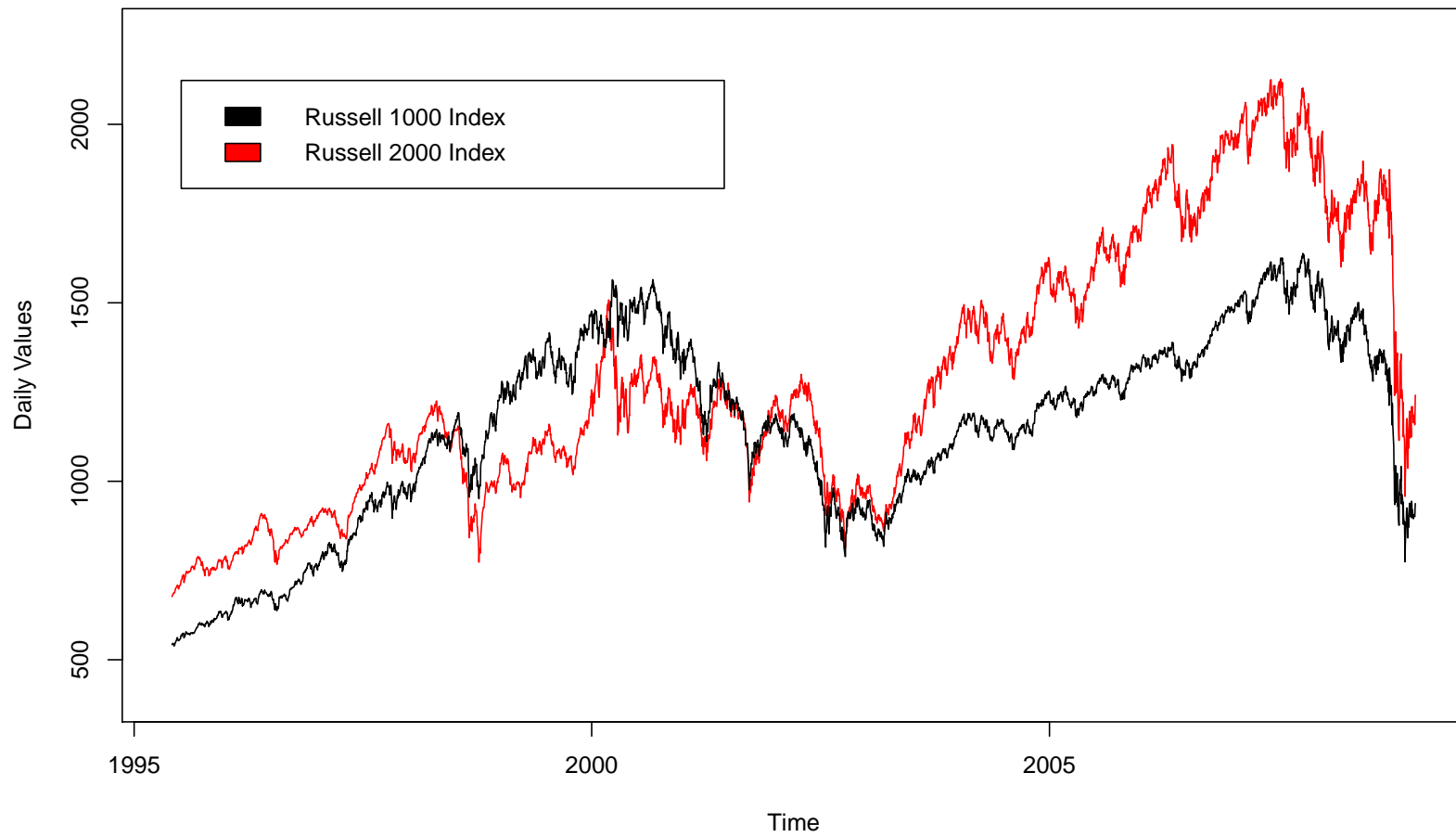
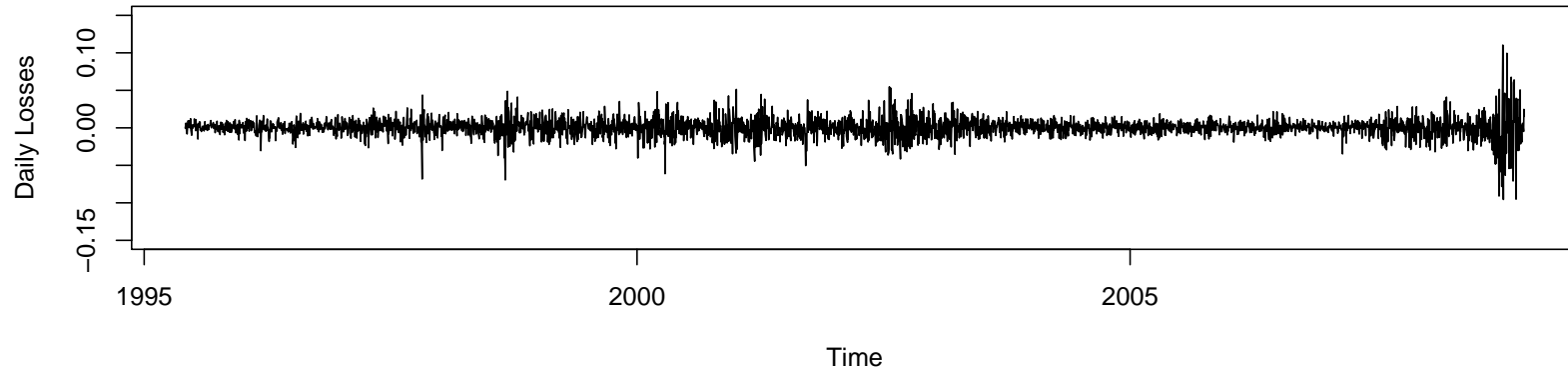


Figure 5.2: Plot of Daily Values

Russell 1000 Index



Russell 2000 Index

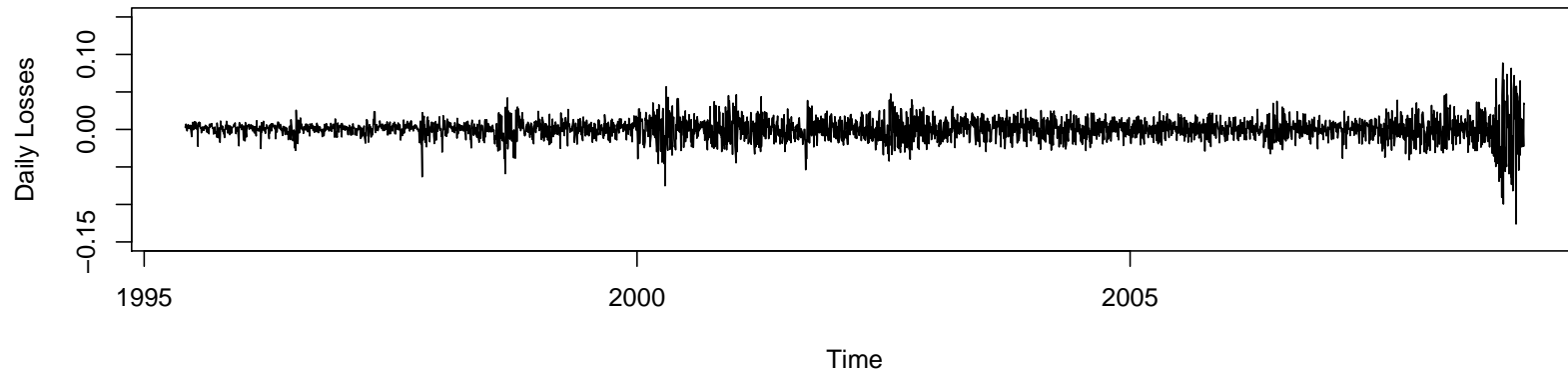


Figure 5.3: Plot of Daily Losses

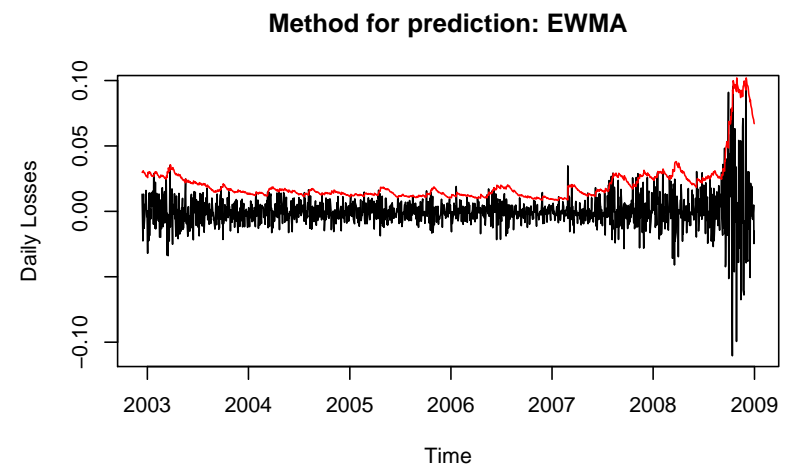
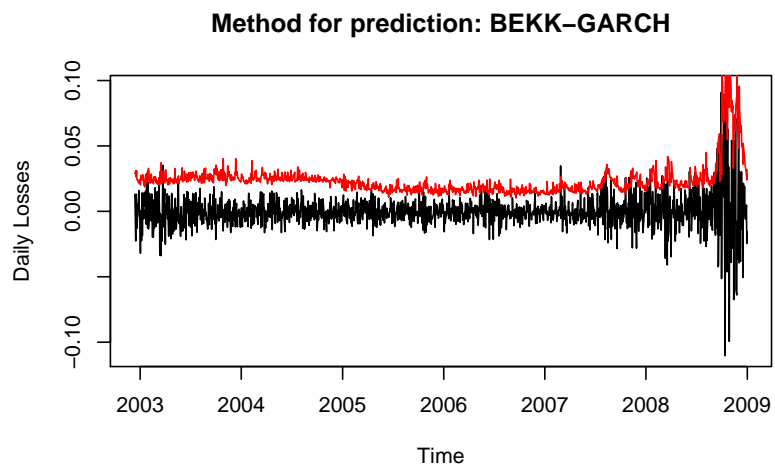
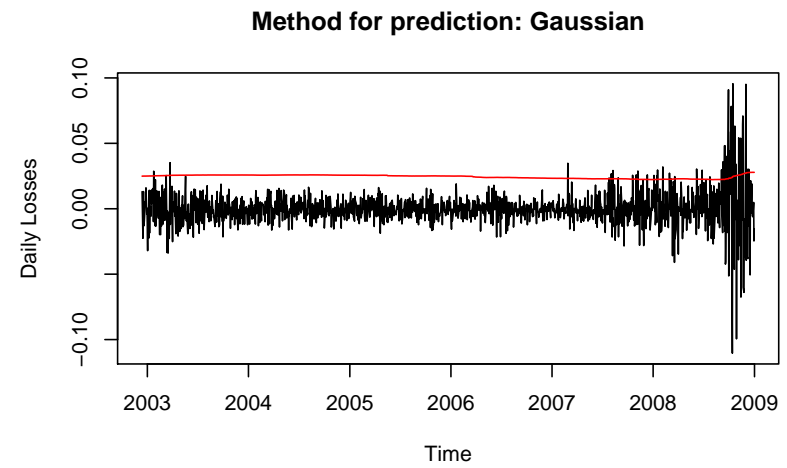
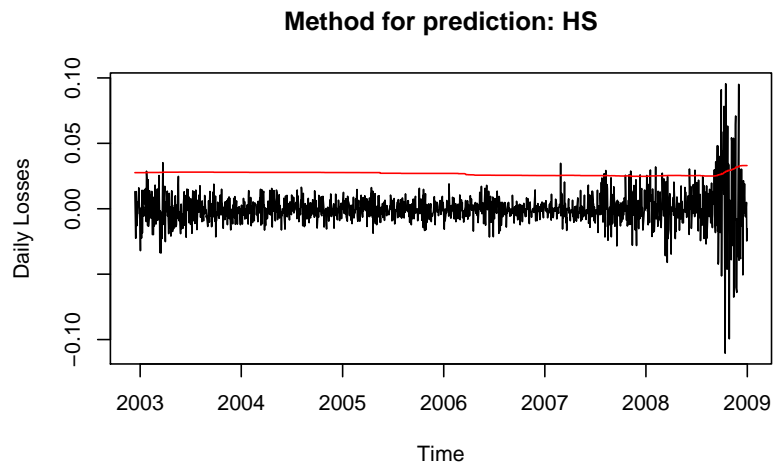


Figure 5.4: Plot of Russell 1000 Index and the Predicted ES

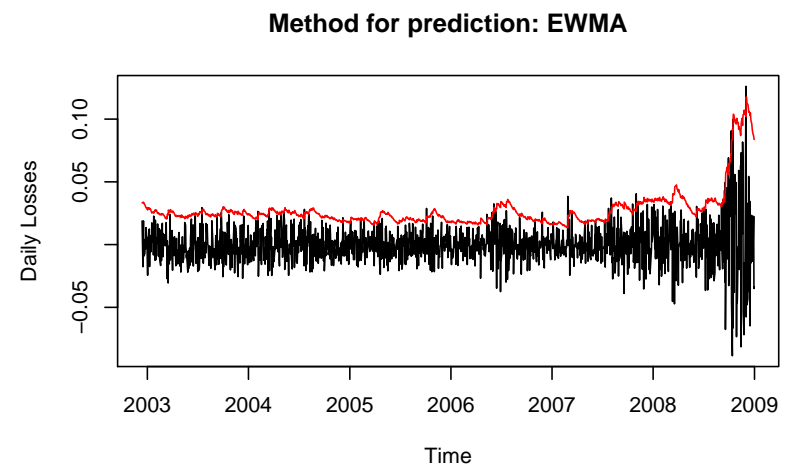
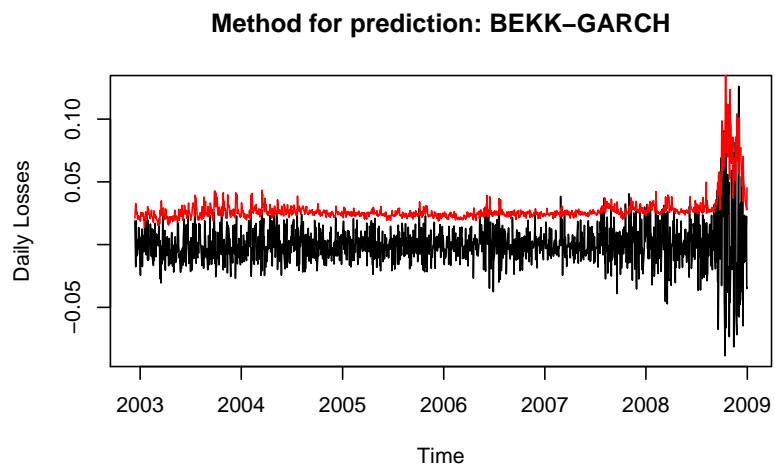
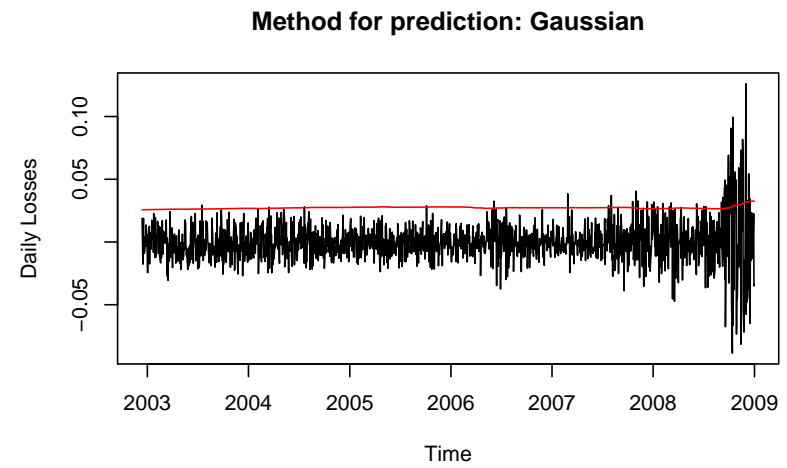
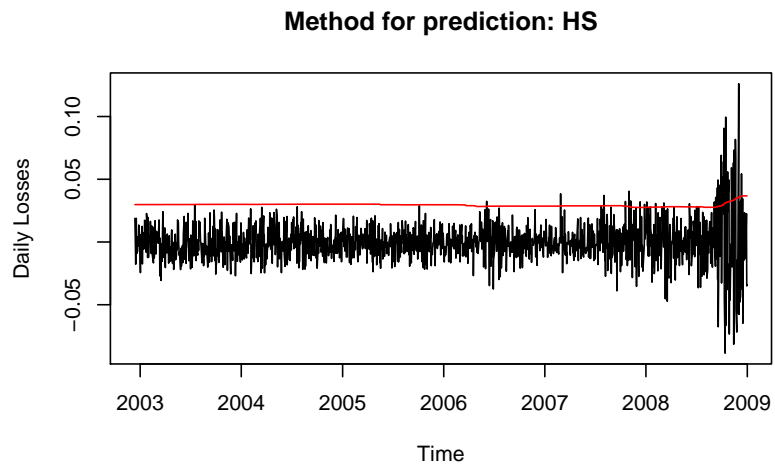


Figure 5.5: Plot of Russell 2000 Index and the Predicted ES

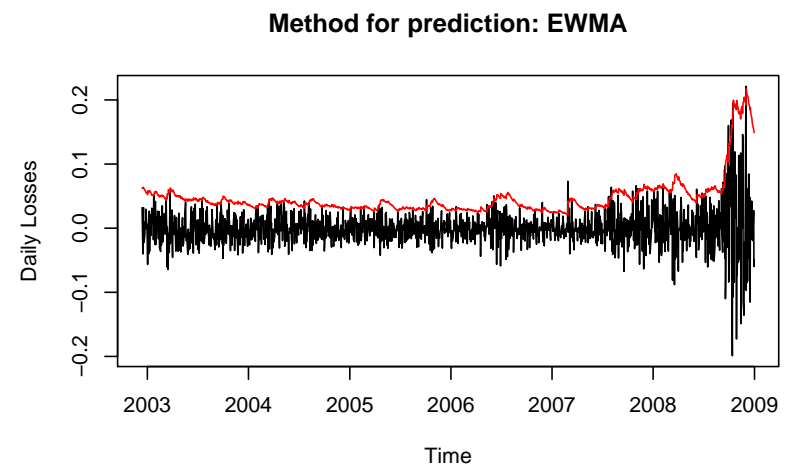
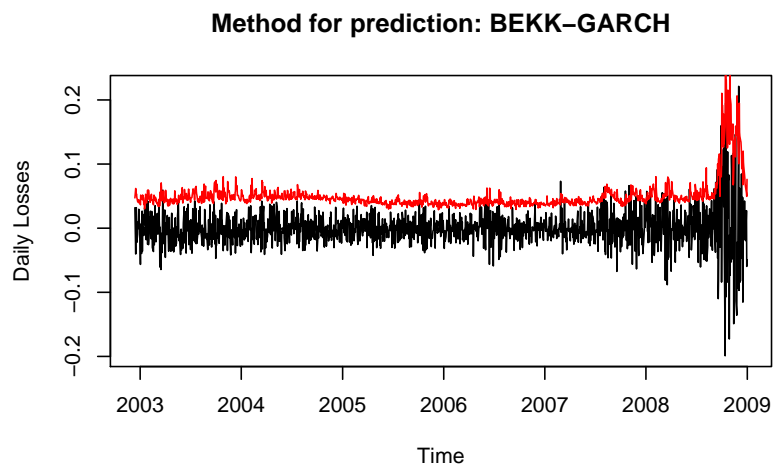
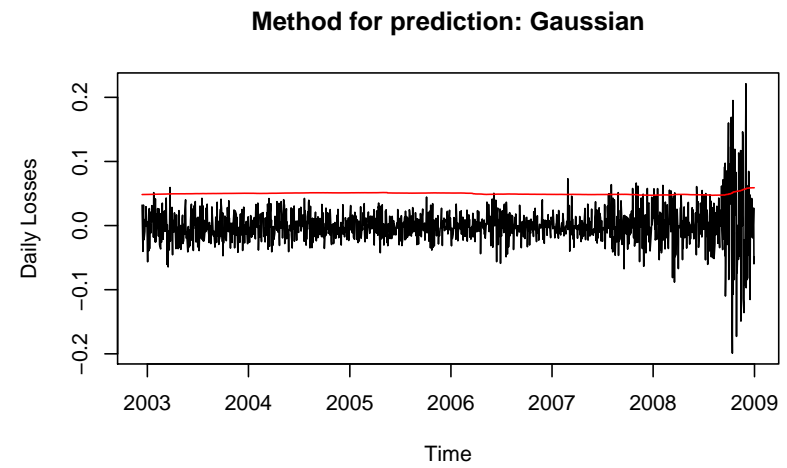
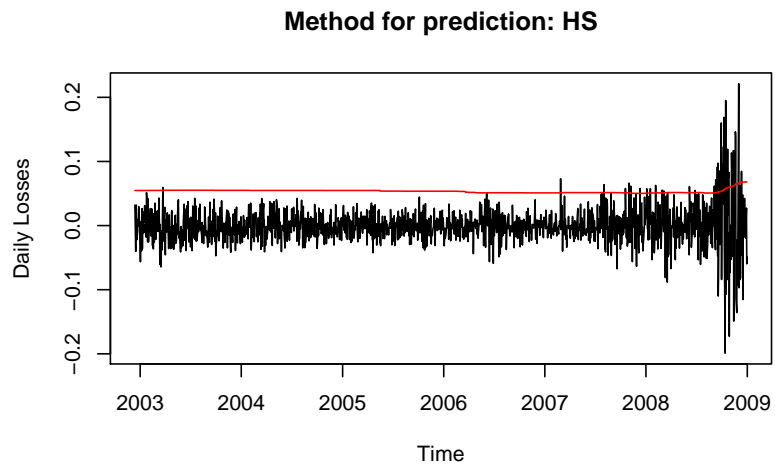


Figure 5.6: Plot of Aggregate Index and the Predicted ES

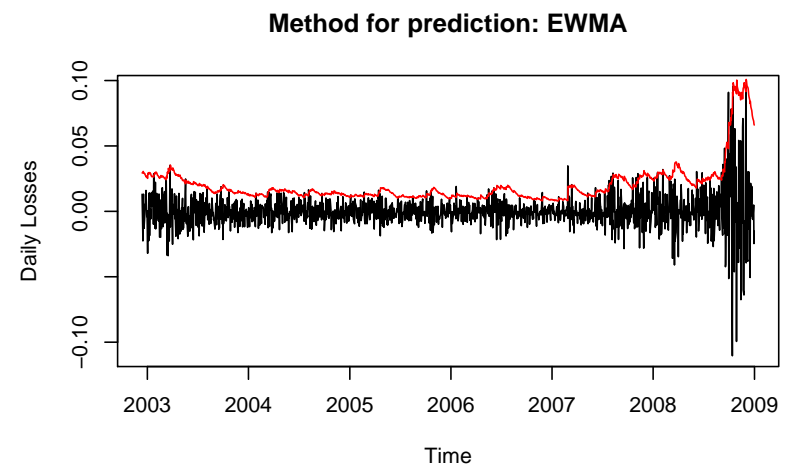
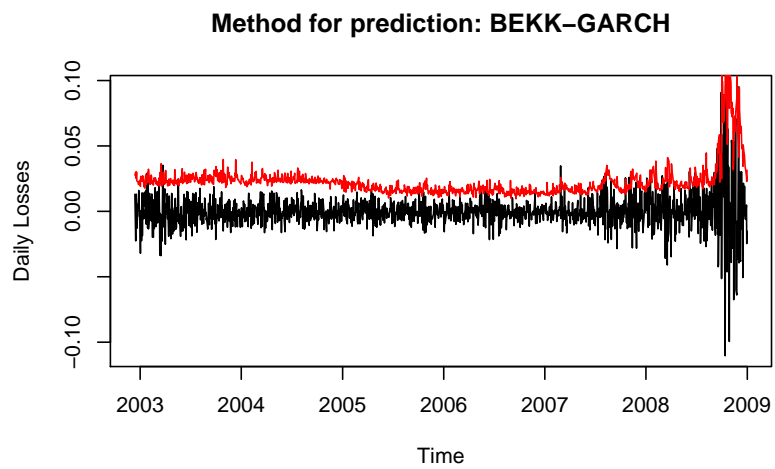
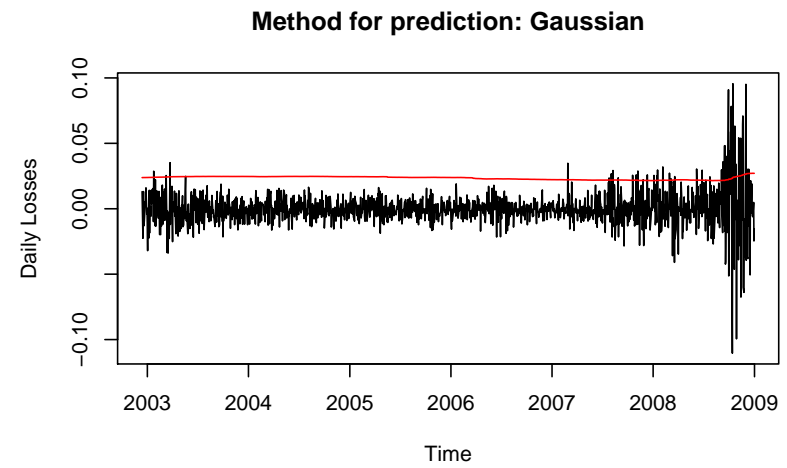
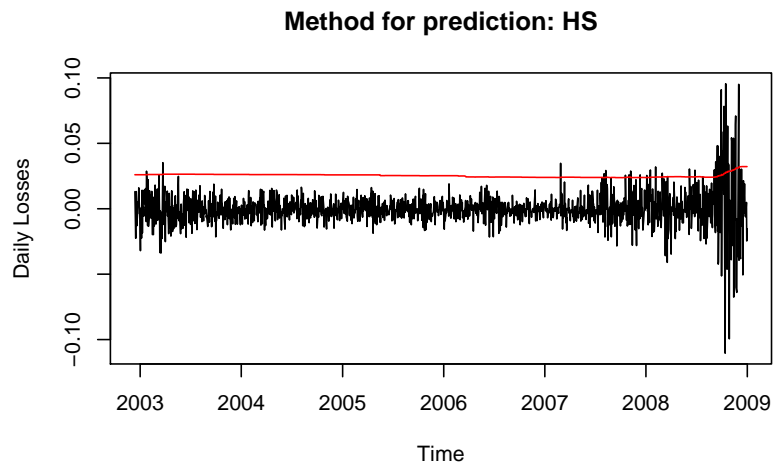


Figure 5.7: Plot of Russell 1000 Index and the Predicted ESC

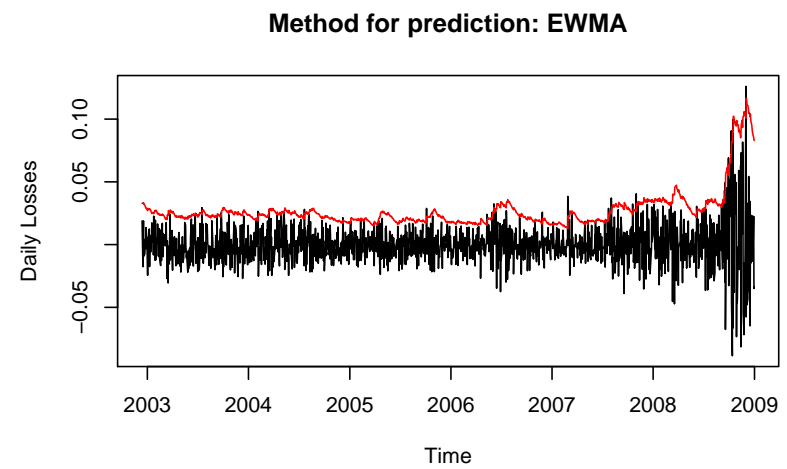
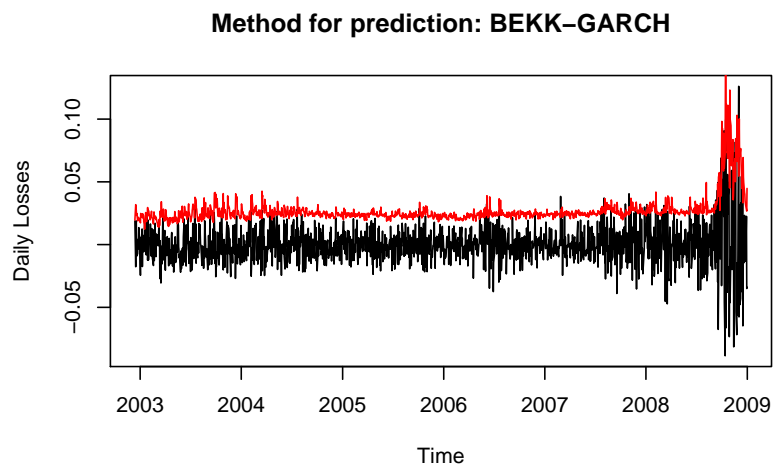
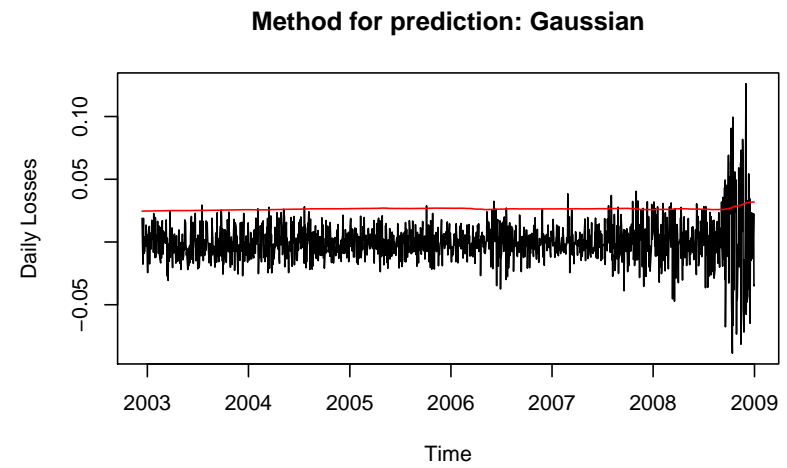
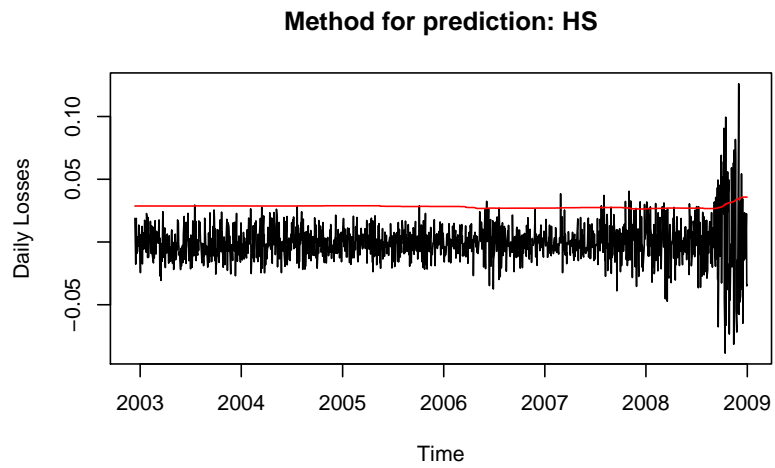


Figure 5.8: Plot of Russell 2000 Index and the Predicted ESC

Table 5.1: P -values of the Bootstrap Test of Zero-mean Behaviour for ES

	Russell 1000	Russell 2000	Aggregate Index
HS	0.063	<u>0.015</u>	<u>0.044</u>
Gaussian	<u>0.017</u>	<u>0.001</u>	<u>0.006</u>
BEKK	<u>0.022</u>	0.100	0.148
EWMA	0.857	0.320	0.644

Table 5.2: P -values of the Duration-based Test of Coverage and Independence

	Russell 1000	Russell 2000	Aggregate Index
HS	<u>0.002</u>	<u>0.002</u>	<u>0.002</u>
Gaussian	<u>0.002</u>	<u>0.002</u>	<u>0.002</u>
BEKK	<u>0.002</u>	0.998	0.515
EWMA	0.060	<u>0.004</u>	<u>0.006</u>

5.1 Backtesting VaR and ES

Before proceeding to backtesting, plots of losses together with predicted ES or ESC for each index is to give a visual sense of how well those predictions are. From Figures 5.4, 5.5, 5.6, 5.7 and 5.8, they all show that the prediction by HS and a Gaussian model are most likely flat, while the prediction of ES by BEKK-GARCH and EWMA vary with volatilities and stay close to the true losses but not exceed too many times. Therefore, the prediction either by BEKK-GARCH or by EWMA is expected to outperform that either by HS or by a GM. The following sections reports the p -values of each test for backtesting ES.

5.1.1 The Bootstrap Test of Zero-mean Behaviour

Table 5.1 reports p -values of the test, and shows that only EWMA passes the bootstrap test of zero-mean behaviour for all indexes, while BEKK passes it for Russell 2000 Index and the aggregate index; neither HS or GM passes the test for any index.

5.1.2 Joint Tests of Coverage and Independence

Table 5.2 reports p -values of the joint test by duration-based approach, and shows that the null hypothesis tends to be rejected for both HS and Gaussian by the duration-based test of independence. Then from p -values of joint test by Pearson's approach in Table 5.3, the result is similar to the corresponding results from Table 5.2.

Table 5.3: P -values of Pearson's χ^2 Test of Coverage and Independence

	Russell 1000	Russell 2000	Aggregate Index
HS	<u>0.009</u>	<u>0.005</u>	<u>0.038</u>
Gaussian	<u>0.004</u>	<u>0.002</u>	<u>0.038</u>
BEKK	<u>0.008</u>	0.305	0.155
EWMA	0.321	0.077	0.100

Table 5.4: P -values of the Duration-based Test of Independence

	Russell 1000	Russell 2000	Aggregate Index
HS	<u>0.002</u>	<u>0.002</u>	<u>0.002</u>
Gaussian	<u>0.002</u>	<u>0.002</u>	<u>0.002</u>
BEKK	<u>0.012</u>	0.986	0.359
EWMA	0.024	<u>0.002</u>	<u>0.002</u>

5.1.3 Test of Independence

Table 5.4 and Table 5.5 report the p -values of the duration-based and Pearson's χ^2 tests for testing independence. It shows that the null hypothesis tends to be rejected for both HS and Gaussian while the null hypothesis is also rejected for EWMA for Russell 2000 Index and the aggregate index.

5.1.4 The Unconditional Coverage Test

Table 5.6 reports p -values of the unconditional coverage test by Binomial approach, and shows that almost all predicting methods pass the unconditional coverage test.

Table 5.5: P -values of Pearson's χ^2 Test of Independence

	Russell 1000	Russell 2000	Aggregate Index
HS	0.092	<u>0.001</u>	<u>0.026</u>
Gaussian	0.063	<u>0.0003</u>	<u>0.026</u>
BEKK	0.619	0.126	0.062
EWMA	0.371	<u>0.030</u>	<u>0.035</u>

Table 5.6: P -values of the Unconditional Coverage Test by Binomial Approach

	Russell 1000	Russell 2000	Aggregate Index
HS	<u>0.007</u>	0.549	0.123
Gaussian	<u>0.003</u>	0.805	0.123
BEKK	<u>0.002</u>	0.916	0.473
EWMA	0.244	0.646	0.823

Table 5.7: P -values of the Bootstrap Test of Zero-mean Behaviour for ESC

	Russell 1000	Russell 2000
HS	0.113	<u>0.013</u>
Gaussian	<u>0.029</u>	<u>0.001</u>
BEKK	0.459	0.104
EWMA	0.637	0.554

5.2 Backtesting ESC

Table 5.7 reports p -values of the bootstrap test of zero-mean behaviour for ESC, and shows that the null hypothesis is rejected for Gaussian for both indices, while it is rejected for HS only for Russell 1000. Both BEKK and EWMA pass the test as expected.

5.3 Conclusions

Combining all the evidence, we tend to reject most of the null hypotheses for both HS and GM and hence these two risk models may be misspecified. Both BEKK method and EWMA tend to pass most tests and hence these two risk models deserve a good label for predicting VaR, ES, and ESC. For Russell Indices, multivariate changing volatility models are better risk models than simple models without a volatility filter, as expected.

Chapter 6

Conclusion and Future Topics

6.1 Conclusion

1. **From the simulation study of bias adjustment:** Bias adjustment by the exact bootstrap for two-stage analysis, EWMA with HS, works pretty well for sample size 1,000, according to the largely reduced relative bias and relative MSE. For the method of EWMA with HS, the relative MSEs are largely reduced by the exact bootstrap for sample sizes of 250 and 500 surprisingly, although their relative biases are not reduced. Therefore, it is still worthwhile to do bias adjustment for the method of EWMA with HS, because most data from financial field have high frequency and EWMA has been widely used.
2. **From the simulation study of backtesting:** From the results of the simulation study, both proposed approaches, the Pearson's χ^2 test and the duration-based test for testing the coverage rate and independence jointly perform as well as those for testing independence. Hence the Pearson's χ^2 test could be a good alternative to the duration-based test. As a practical guideline, we recommend that it may be convenient to perform the joint tests first. If the null hypothesis of the joint tests is rejected, then it would be worthwhile to proceed to perform the test of the unconditional coverage and the test of independence separately to see if both null hypotheses or only one of them are rejected. Otherwise, it may not be necessary to perform all of them separately if the null hypothesis of the joint test is not rejected.
3. **From the empirical study for Russell Indices:** Overall, we tend to reject most of the null hypotheses for both HS and GM and hence these two risk models may be misspecified. Both BEKK method and EWMA tend to pass most tests and hence these two risk models deserve a good label for predicting VaR, ES, and ESC. For Russell Indices, multivariate

changing volatility models are better risk models than simple models without a volatility filter, as expected.

6.2 Future Topics

6.2.1 Sequential Estimation of ESC for Higher Dimensional MGARCH

The consistency axiom of fair allocation (Definition 1.14) facilitates iterative calculation of expected shortfall (risk) contributions (ESC). Therefore the prediction of ESC for a multivariate GARCH (MGARCH) model with dimension $d > 2$ may be achieved by repeating allocation of a bivariate GARCH (BGARCH) model $Y_t = (S_t(j+1), X_{j,t})'$ for $j = 1, 2, \dots, (d-1)$ where $S_t(k) = \sum_{i=k}^d X_{i,t}$, because it is always easier to fit a BGARCH model to data than to fit the full model of MGARCH. Besides, the prediction by the above scheme can be compared with that by sequential-BEKK multivariate GARCH models proposed by Xu and Lin (2008).

6.2.2 Improvement of GPD Method With Optimal Threshold

Since the choice of the threshold in GPD method may cause severe impact on the tail estimation, it has been an important issue in EVT. Jalal and Rockinger (2006, Section 2.3) discussed available methods of threshold selection. An optimal choice of the threshold may also help obtain a better estimator of ES derived from GPD approach.

6.2.3 Other Estimation Methods for GARCH Models

Fan and Yao (2003, section 4.2.3, pp.156) addressed three types of estimators for parameters in a GARCH model: the conditional maximum likelihood estimator (MLE), Whittle's estimator (Giraitis and Robinson, 2001), and the least absolute deviation estimator (Peng and Yao, 2002). The first one is the same ML method as discussed in McNeil et al. (2005) and has been widely used in the banking industry as a benchmark estimator. The last one is more appealing for models with heavy-tailed innovations. Therefore it may also give better estimator of ES.

REFERENCES

- C. Acerbi. Spectral measures of risk: a coherent representation of subjective risk aversion. *Journal of Banking and Finance*, 26:1505–1518, 2002.
- C. Acerbi. Coherent representations of subjective risk-aversion. In G. Szego, editor, *Risk Measures for the 21th Century*, pages 147–207. Wiley: New York, 2004.
- C. Acerbi and D. Tasche. Coherent allocation of risk capital. Working paper, Zentrum Mathematik (SCA), TU Munchen, Germany, 1999.
- C. Acerbi and D. Tasche. On the coherence of expected shortfall. *Journal of Banking and Finance*, 26:1487–1503, 2002.
- C. Acerbi, C. Nordio, and C. Sirtori. Expected shortfall as a tool for financial risk management. Technical report.
- P. Artzner, F. Delbaen, J. M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9(3):203–228, 1999.
- A. A. Balkema and L. de Haan. Residual life time at great age. *Annals of Probability*, 2: 792–804, 1974.
- R. E. Baysal and J. Staum. Empirical likelihood for value at risk and expected shortfall. IEMS working paper 07-01, Northwestern University, 2007. under revision.
- BCBS. Amendment to the capital accord to incorporate market risks. Technical report, Basel Committee on Bank Supervision, 1996.
- E.K. Berndt, B.H. Hall, Hall R. E., and J. A. Hausman. Estimation and inference in nonlinear structural models. *Annals of Economic and Social Measurement*, pages 653–665, 1974.
- D. Bertsimas, G.J. Lauprete, and A. Samarov. Shortfall as a risk measure: properties, optimization and applications. Working paper, Sloan School of Management, MIT, Cambridge, 2000.
- P. Billingsley. *Probability and Measure*. A Wiley-Interscience publication, John Wiley & Sons, third edition, 1995.
- J. Cai and H. Li. Conditional tail expectations for multivariate phase-type distributions. *Journal Risk and Insurance*, 42(3):810–825, 2005.
- J. Cai and K. S. Tan. Optimal retention for a stop-loss re-insurance under the VaR and CTE risk measures, 2006.
- Z. Cai and X. Wang. Nonparametric methods for estimating conditional var and expected shortfall. Wise working paper series WISEWP0604, 2006.
- C. Caillault and D. Guégan. Forecasting VaR and expected shortfall using dynamic systems: a risk management strategy. Working paper No 03-2004, IDHE-MORA, 2004.

- S. Cambanis, S. Huang, and G. Simons. On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis*, 11:368–385, 1981.
- E. Carlstein. The use of subseries values for estimating the variance of a general statistic from a stationary sequence. *Annals of Statistics*, 14:1171–1194, 1986.
- S. X. Chen. Nonparametric estimation of expected shortfall. *Journal of Financial Econometrics*, 6(1):87–107, 2008.
- P. F. Christofferson. Evaluating interval forecasts. *International Economic Review*, 39:841–862, 1998.
- P. F. Christofferson and D. Pelletier. Backtesting value-at risk: A duration-based approach. *Journal of Empirical Finance*, 2:84–108, 2004.
- J. Cotter and K. Dowd. Estimating financial risk measures for futures positions: a parametric approach. MPRA paper no. 3503, November 2007.
- A. DasGupta. *Asymptotic Theory of Statistics and Probability*. Springer Texts in Statistics. Springer, 2008.
- H. A. David and H. N. Nagaraja. *Order Statistics*. Wiley series in probability and statistics. John Wiley & Sons, Inc., Hoboken, New Jersey, third edition edition, 2003.
- A. C. Davison and D. V. Hinkley. *Bootstrap methods and their application*. Cambridge University Press, New York, 1997.
- F. Delbaen. Coherent risk measures on general probability spaces. In K. Sandmann and Philipp J. Schönbucher, editors, *Advances in Finance and Stochastics*, pages 1–37. Springer, Berlin, 2002.
- M. Denault. Coherent allocation of risk capital. Working paper, ETH RiskLab, Zürich, 2001.
- J. Dhaene, L. Henard, Z. Landsman, A. Vandendorpe, and S. Vanduffel. Some results on the CTE based capital allocation rule, September 2007.
- K. Dowd. Estimating risk measures. *Financial Engineering News*, 43:13, 2005a.
- K. Dowd. *Measuring Market Risk*. Wiley: Chichester and New York, second edition, 2005b.
- B. Efron. Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, 7:1–26, 1979.
- B. Efron and R. J. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall, New York, 1993.
- P. Embrechts, C. Klüppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer, 1997.
- R. Engle and F. Kroner. Multivariate simultaneous generalized arch. *Econometric Theory*, 11: 122–150, 1995.

- J. Fan and Q Yao. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer Series in Statistics. Springer, 2003.
- E. Furman and Z. Landsman. Risk capital decomposition for a multivariate dependent gamma portfolio. *Insurance: Mathematics and Economics*, 37(3):635–649, 2005.
- E. Furman and Z. Landsman. Economic capital allocations for non-negative portfolios of dependent risks. *ASTIN Bulletin*, 2008. forthcoming.
- L. Giraitis and P.M. Robinson. Whittle estimation of arch models. *Econometric Theory*, 17: 649–668, 2001.
- M. Hardy. Simulating VaR and CTE. *Financial Engineering News*, 47:17, 2006.
- M. R. Hardy. *Investment Guarantees: Modeling and Risk Management for Equity-Linked Life Insurance*. John Wiley & Sons, New York, 2003.
- B. Hill. A simple general approach to inference about the tail of a distribution. *Annals of Statistics*, 3:1163–1174, 1975.
- Glyn A. Holton. *Value-at-Risk: Theory and Practice*. Academic Press, 2003.
- Simon Hurst. The characteristic function of the student-t distribution. Financial mathematics research report no. fmrr006-95, Mathematical Sciences Institute, 1995. Statistics Research Report No. SRR044-95.
- A. D. Hutson and M. D. Ernst. The exact bootstrap mean and variance of an L -estimator. *Journal of Royal Statistical Society: Series B*, 62:89–94, 2000.
- A. Jalal and M. Rockinger. Predicting tail-related risk measures: The consequences of using garch filters for a non garch data. Working paper, National Centre of Competence in Research, Financial Valuation and Risk Management (NCCR FINRISK), April 2006. NCCR FINRISK project on “Financial Econometrics for Risk Management”.
- P. Jorion. *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill Trade, 2nd edition, 2001.
- H. T. Kim. *Estimation and allocation of insurance risk capital*. PhD thesis, Actuarial Science, University of Waterloo, 2007.
- Z. Landsman and E. A. Valdez. Tail conditional expectations for elliptical distributions. *North American Actuarial Journal*, 7(4), 2003.
- Z. Landsman and E. A. Valdez. Tail conditional expectations for exponential dispersion models. *ASTIN Bulletin*, 35(1):189–209, 2005.
- A. J. McNeil and R. Frey. Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach. *Journal of Empirical Finance*, 7(3-4):271–300, 2000.
- A. J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management*. Princeton University Press, New Jersey, 2005.

- J.P. Morgan. Riskmetrics - technical document. Technical report, Morgan Guaranty Trust Company, New York, 1995. 3rd Edition.
- L. Overbeck. *Measuring Risk in Complex Stochastic Systems*, chapter Allocation of economic capital in loan portfolios, pages 15–30. Springer, 2000.
- H. H. Panjer. Measurement of risk, solvency requirements, and allocation of capital within financial conglomerates. Research report 01-15, Institute of Insurance and Pension Research, University of Waterloo, 2002.
- S. Pattarathammas, S. Mokkavesa, and P. Nilla-Or. Value-at-Risk and expected shortfall under extreme value theory framework: An empirical study on Asian markets, 2008.
- L. Peng and Q. Yao. Least absolute deviation estimation for ARCH and GARCH models. Preprint, 2002.
- J. Pickands. Statistical inference using extreme order statistics. *Annals of Statistics*, 3:119–131, 1975.
- M. L. Puri and F. H. Ruymgaart. Asymptotic behavior of L -statistics for a large class of time series. *Annals of the Institute of Statistical Mathematics*, 45(4):687–701, 1993.
- RiskMetrics. RiskMetrics Technical Document. J.P. Morgan, 1996. 4th edn.
- O. Scaillet. Nonparametric estimation and sensitivity analysis of expected shortfall. *Mathematical Finance*, 14(1):115–129, 2004.
- O. Scaillet. Nonparametric estimation of conditional expected shortfall. *Revue Assurances et Gestion des Risques/Insurance and Risk Management Journal*, 74:639–660, 2005.
- J. Shao and D. Tu. *The Jackknife and Bootstrap*. Springer, 1995.
- X. Shi. Bootstrap estimate for m -dependent sample means. *Kexue Tongbao (Chinese Bulletin of Science)*, 31:404–407, 1986.
- R. L. Smith. Estimating tails of probability distributions. *Annals of Statistics*, 15:1174–1207, 1987.
- Tatiana Solcá. Expected risk-adjusted return for insurance based models. Master’s thesis, ETH Zürich, 2000.
- SPAN. Standard portfolio analysis of risk. Technical report, Chicago Merchantile Exchange, Chicago, 1995.
- Brajendra C. Sutradhar. On the characteristic function of multivariate student t-distribution. *The Canadian Journal of Statistics*, 14(4):329–337, 1986.
- D. Tasche. Risk contributions and performance measurement. Working paper, Zentrum Marhe-matic (SCA), TU Munchen, Germany, 1999.

- D. Tasche. Expected shortfall: a natural coherent alternative to value at risk. *Economic Notes*, 31(2):379–388, 2002a.
- D. Tasche. Expected shortfall and beyond. *Journal of Banking and Finance*, 26(7):1519–1533, 2002b.
- D. Tasche. Euler allocation: theory and practice. Working paper, Zentrum Marhematic (SCA), TU Munchen, Germany, August 2007.
- D. Tasche. Capital allocation to business units and sub-portfolios: the euler principle. Working paper, Zentrum Marhematic (SCA), TU Munchen, Germany, June 2008.
- E. A. Valdez and A. Chernih. Wang’s capital allocation formula for elliptical contoured distribution. *Insurance: Mathematics and Economics*, 33:517–532, 2003.
- S. S. Wang. A set of new methods and tools for enterprise risk capital management and portfolio optimization. Working paper, SCOR Reinsurance company, 2002.
- J. Wirch and M. Hardy. A synthesis of risk measures for capital adequacy. *Insurance: Mathematics and Economics*, 25:337–348, 1999.
- Jincai Xu and Kuan-Pin Lin. Sequential-BEKK multivariate garch models and its application in stock market. Technical report, 2008.

APPENDICES

Appendix A

Proof of Proposition 1.9

Generally, $\text{CTE}_\alpha(X) \leq \text{ES}_\alpha(X)$.

Proof. We want to show that $\text{CTE}_\alpha(X) \leq \text{ES}_\alpha(X)$. It is equivalent to show that $\text{ES}_\alpha(X) - \text{CTE}_\alpha(X) \geq 0$. Recall that

$$\text{CTE}_\alpha(X) := \mathbb{E}[-X | X \leq -\text{VaR}_\alpha(X)]. \quad (\text{A.1})$$

Denote A and \tilde{x} as $\{\omega : X(\omega) < -\text{VaR}_\alpha(X)\}$ and $-\text{VaR}_\alpha(X)$, respectively. Rewrite (A.1) as

$$\begin{aligned} \text{CTE}_\alpha(X) &= \frac{1}{P(X \leq -\text{VaR}_\alpha(X))} \cdot \mathbb{E}[-X \cdot I_{\{X \leq -\text{VaR}_\alpha(X)\}}] \\ &= \frac{1}{P(A) + P(X = \tilde{x})} \cdot \mathbb{E}[-X \cdot I_{\{X \leq \tilde{x}\}}] \\ &= \frac{1}{P(A) + P(X = \tilde{x})} \cdot \{\mathbb{E}[-X \cdot I_A] - \tilde{x} \cdot P(X = \tilde{x})\}. \end{aligned} \quad (\text{A.2})$$

Since the definition of ES in (1.2) can be rewritten as

$$\text{ES}_\alpha(X) = k \cdot \mathbb{E}_P[-X \cdot I_A] - \tilde{x} \cdot k' \cdot P(X = \tilde{x}) \quad (\text{A.3})$$

where $k := 1/(1 - \alpha)$, and k' satisfies $0 \leq k' \leq k$ and is determined by the equation

$$k \cdot P(X < \tilde{x}) + k' \cdot P(X = \tilde{x}) = 1.$$

Let $e = \mathbb{E}_P[-X \cdot I_A]$, $p = P(X = \tilde{x})$, and $a = P(A)$. Then

$$\begin{aligned}
\text{ES}_\alpha(X) - \text{CTE}_\alpha(X) &= [k \cdot e - k' \tilde{x} p] - \left\{ \frac{e - \tilde{x} \cdot p}{a + p} \right\} \\
&= \frac{1}{a + p} [\tilde{x} \cdot p - e + (a + p) \cdot (k \cdot e - k' \cdot \tilde{x} \cdot p)] \\
&= \frac{1}{a + p} [\tilde{x} \cdot p + k \cdot e \cdot (a + p) - e - k' \cdot \tilde{x} \cdot p \cdot (a + p)] \\
&= \frac{1}{a + p} \{ \tilde{x} \cdot p \cdot [1 - k' \cdot (a + p)] + e \cdot [k(a + p) - 1] \}
\end{aligned}$$

Note that $k \cdot r + k' \cdot p = 1$ and $0 \leq k' \leq k$ infers $0 \leq 1 - k(a + p) \leq 1 - k'(a + p) \leq 1$ and $k(a + p) \geq 1$. Hence

$$\begin{aligned}
\text{ES}_\alpha(X) - \text{CTE}_\alpha(X) &\geq \frac{1}{a + p} \{ \tilde{x} \cdot p \cdot [1 - k \cdot (a + p)] + e \cdot [k(a + p) - 1] \} \\
&= \frac{1}{a + p} \cdot [k(a + p) - 1] \cdot [e - \tilde{x} \cdot p] \\
&\geq \frac{1}{a + p} \cdot (1 - 1) \cdot [e - \tilde{x} \cdot p] = 0.
\end{aligned}$$

□

Appendix B

Proof of Proposition 1.10

$$\text{CTE}_\alpha(X) = \text{ES}_\alpha(X) \tag{B.1}$$

holds if and only if one of the following conditions is true.

1.

$$\Pr[X \leq -\text{VaR}_\alpha(X)] = 1 - \alpha. \tag{B.2}$$

2.

$$\Pr[X < -\text{VaR}_\alpha(X)] = 0. \tag{B.3}$$

Proof. Recall from (A.3) and (A.2) that

$$\text{ES}_\alpha(X) = k \cdot \mathbb{E}_P[-X \cdot I_A] - \tilde{x} \cdot k' \cdot P(X = \tilde{x})$$

and

$$\text{CTE}_\alpha(X) = \frac{1}{P(A) + P(X = \tilde{x})} \cdot \{\mathbb{E}[-X \cdot I_A] - \tilde{x} \cdot P(X = \tilde{x})\}$$

where $\tilde{x} := -\text{VaR}_\alpha(X)$, $A := \{\omega : X(\omega) < \tilde{x}\}$, $k := 1/(1 - \alpha)$, and k' is determined by the equation

$$k \cdot P(X < \tilde{x}) + k' \cdot P(X = \tilde{x}) = 1.$$

B.1 To show (B.1) holds if either (B.2) or (B.3) is true

Note first that obviously both $\text{ES}_\alpha(X)$ in (A.3) and $\text{CTE}_\alpha(X)$ in (A.2) are reduced to $k \cdot \mathbb{E}_P[-X \cdot I_A]$ if $P[X = -\text{VaR}_\alpha(X)] = 0$. However, it is just a sufficient but not necessary condition to make (B.1) holds.

First, we want to show that $\text{CTE}_\alpha(X) = \text{ES}_\alpha(X)$ holds if (B.2) is true. i.e. suppose the

following is true.

$$P[X \leq -\text{VaR}_\alpha(X)] = 1 - \alpha. \quad (\text{B.4})$$

And it is known that

$$P[X \leq \tilde{x}] = 1 - \alpha := 1/k \quad (\text{B.5})$$

where $\tilde{x} := -\text{VaR}_\alpha(X)$. Besides, we know that

$$k \cdot P(X < \tilde{x}) + k' \cdot P(X = \tilde{x}) = 1.$$

Then

$$\begin{aligned} 1 &= k \cdot [P(X \leq \tilde{x}) - P(X = \tilde{x})] + k' \cdot P(X = \tilde{x}) \\ &= k \cdot \frac{1}{k} + (k - k') \cdot P(X = \tilde{x}) \\ &= 1 - (k - k') \cdot P(X = \tilde{x}) \end{aligned}$$

and hence we get

$$(k - k') \cdot P(X = \tilde{x}) = 0.$$

i.e.

$$k - k' = 0 \text{ or } P(X = \tilde{x}) = 0.$$

When $P(X = \tilde{x}) = 0$, it has been shown previously that $\text{CTE}_\alpha(X) = \text{ES}_\alpha(X)$ holds. When $k - k' = 0$ or equivalently $k = k'$, then both $\text{ES}_\alpha(X)$ in (A.3) and $\text{CTE}_\alpha(X)$ in (A.2) are reduced to

$$k \cdot \{\mathbb{E}_P[-X \cdot I_A] - \tilde{x} \cdot P(X = \tilde{x})\}.$$

Finally, we want to show that $\text{CTE}_\alpha(X) = \text{ES}_\alpha(X)$ holds if (B.3) is true. Suppose that

$$\Pr[X < -\text{VaR}_\alpha(X)] = 0.$$

Then both $\text{ES}_\alpha(X)$ in (A.3) and $\text{CTE}_\alpha(X)$ in (A.2) are reduced to $\text{VaR}_\alpha(X)$.

Therefore, we can conclude that the equality, $\text{CTE}_\alpha(X) = \text{ES}_\alpha(X)$, holds, if one of the following statements is true: $P[X \leq -\text{VaR}_\alpha(X)] = 1 - \alpha$ and $P[X < -\text{VaR}_\alpha(X)] = 0$.

B.2 To show either (B.2) or (B.3) is true, if (B.1) holds

Suppose that $\text{CTE}_\alpha(X) = \text{ES}_\alpha(X)$. Denote B as the set $\{X \leq \tilde{x}\}$ and rewrite (A.3) and (A.2) as

$$\text{ES}_\alpha(X) = k \cdot \{\mathbb{E}_P[-X \cdot I_B] - \tilde{x} \cdot [P(X \leq \tilde{x}) - k^{-1}]\} \quad (\text{B.6})$$

and

$$\text{CTE}_\alpha(X) = \frac{1}{P(B)} \cdot \mathbb{E}[-X \cdot I_B] \quad (\text{B.7})$$

respectively. Now set (B.6) and (B.7) equal and then get

$$k \cdot \{\mathbb{E}_P[-X \cdot I_B] - \tilde{x} \cdot [P(X \leq \tilde{x}) - (1 - \alpha)]\} = \frac{1}{P(B)} \cdot \mathbb{E}[-X \cdot I_B].$$

Let $b = P(B)$ and $e_B = \mathbb{E}[-X \cdot I_B]$. Thus we obtain

$$\begin{aligned} b \cdot k \cdot e_B + b \cdot k \cdot \tilde{x} \cdot [b - k^{-1}] &= e_B \\ \Rightarrow (k \cdot \tilde{x}) \cdot b^2 + [k \cdot e_B - k \cdot \tilde{x} \cdot k^{-1}] \cdot b - e_B &= 0 \\ \Rightarrow (k \cdot \tilde{x}) \cdot b^2 + [k \cdot e_B - \tilde{x}] \cdot b - e_B &= 0 \\ \Rightarrow (k \cdot b - 1) \cdot (\tilde{x} \cdot b + e_B) &= 0 \\ \Rightarrow b &= \begin{cases} 1/k, & \text{if } k \neq 0 \\ -e_B/\tilde{x}, & \text{if } \tilde{x} \neq 0. \end{cases} \end{aligned}$$

In terms of the original notation, we have

$$P(B) = \begin{cases} (1 - \alpha), & \text{if } \alpha \neq 1 \\ -E[-X \cdot I_B]/\tilde{x}, & \text{if } \tilde{x} \neq 0. \end{cases} \quad (\text{B.8})$$

i.e.

$$P[X \leq \text{VaR}_\alpha(X)] = 1 - \alpha \text{ if } \alpha \neq 1, \quad (\text{B.9})$$

or

$$-E[-X \cdot I_B] = \tilde{x} \cdot P(B). \quad (\text{B.10})$$

(B.10) needs more derivatoin as follows.

$$\begin{aligned}
& -E[-X \cdot I_B] = \tilde{x} \cdot P(B) \\
\Rightarrow & -E[-X \cdot I_B]/P(B) = \tilde{x} \\
\Rightarrow & -\text{CTE}_\alpha(X) = \tilde{x} = -\text{VaR}_\alpha(X) \\
\Rightarrow & \text{CTE}_\alpha(X) = \text{VaR}_\alpha(X).
\end{aligned}$$

Now we have to prove the following claim by transposition:

$$\text{CTE}_\alpha(X) = \text{VaR}_\alpha(X) \Rightarrow \Pr[X < -\text{VaR}_\alpha(X)] = 0, \quad (\text{B.11})$$

That is, we want to show its contrapositive statement is true. If $\Pr[X < -\text{VaR}_\alpha(X)] \neq 0$, then $\Pr[X < -\text{VaR}_\alpha(X)] > 0$. If $\text{CTE}_\alpha(X) = \text{VaR}_\alpha(X)$, then $\text{CTE}_\alpha(X) > \text{VaR}_\alpha(X)$. Suppose that $\Pr[X < -\text{VaR}_\alpha(X)] > 0$ and we'd like to show that $\text{CTE}_\alpha(X) > \text{VaR}_\alpha(X)$. Then

$$\begin{aligned}
E[-X \cdot I_B] &= \int_{X \leq \tilde{x}} (-X) dP \\
&> \int_{X \leq \tilde{x}} (-\tilde{x}) dP \text{ since } \Pr[X < -\text{VaR}_\alpha(X)] > 0 \\
&= (-\tilde{x}) \cdot P(B).
\end{aligned}$$

Thus we get $E[-X \cdot I_B]/P(B) > -\tilde{x}$ and hence $\text{CTE}_\alpha(X) > \text{VaR}_\alpha(X)$. It results in the conclusion that $\Pr[X < -\text{VaR}_\alpha(X)] = 0$ if $\text{CTE}_\alpha(X) = \text{VaR}_\alpha(X)$. \square

Appendix C

Proof of Theorem 2.18

Suppose that $X \sim E(\mu, \sigma, g)$. Then under condition (2.37), we have ES of X

$$\text{ES}_\alpha(X) = \mu + \sigma \cdot \lambda_{g,\alpha}$$

where $\lambda_{g,\alpha}$ is defined in Definition 2.17.

Proof. Suppose that $X \sim E(\mu, \sigma, g)$ and let $c = c_1$ and $g(\cdot) = g_1(\cdot)$ are respectively the normalizing constant and density generator. It has been shown in Dhaene et al. (2007) that

$$\text{ES}_\alpha(X) = \mu + \sigma \frac{c}{1 - \alpha} \int_{\frac{1}{2}z_\alpha^2}^{\infty} g(x) dx$$

where $z_\alpha = \frac{\text{VaR}_\alpha(X) - \mu}{\sigma}$. It is easy to show that $z_\alpha = \text{VaR}_\alpha(Z)$ where $Z \sim E(0, 1, g)$ and examples for Gaussian and t distributions can be found in McNeil et al. (2005, Example 2.14). Then by definition of $\overline{G}(x) = G(\infty) - G(x)$ where $G(\cdot)$ is the cumulative generator, it is trivial to show that $\overline{G}(\frac{1}{2}z_\alpha^2) = \int_{\frac{1}{2}z_\alpha^2}^{\infty} g(x) dx$. Hence

$$\text{ES}_\alpha(X) = \mu + \sigma \cdot \lambda_{g,\alpha}$$

where

$$\lambda_{g,\alpha} = \frac{\overline{G}(\frac{1}{2}z_\alpha^2)}{1 - \alpha}$$

with $z_\alpha = \text{VaR}_\alpha(Z)$ for $Z \sim E(0, 1, g)$.

Alternatively, the proof is done by changing λ to $\lambda_{g_1,\alpha}$ with corresponding corrections that $\sigma^2 \lambda = \sigma \lambda_{g_1,\alpha}$ □

Appendix D

Proof of Theorem 4.2

Suppose that $\mathbf{X} \sim \text{E}_d(\boldsymbol{\mu}, \Sigma, g_d)$ where $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)'$ and $(\Sigma)_{ij} = \sigma_{ij}$. Denote the k -th component of \mathbf{X}_t by $X_{k,t} = \mathbf{e}'_k \mathbf{X}_t$, and the sum of all components by $S_t = \sum_{j=1}^d X_{j,t} = \mathbf{1}' \mathbf{X}_t$, where \mathbf{e}_k is the vector of zeros but one at the k -th location, and $\mathbf{1} = (1, \dots, 1)'$ the vector of ones with dimension d . Let $\lambda_{g_1, \alpha}$ be the λ -factor, defined in Definition 2.17. Then under condition (2.37) which guarantees the existence of the mean vector such that $E(\mathbf{X}) = \boldsymbol{\mu}$:

1. The individual ES of X_k for $k = 1, \dots, d$ is given by

$$\text{ES}_\alpha(X_k) = \mu_k + \lambda_{g_1, \alpha} \cdot \sigma_k^*,$$

where $\sigma_k^* = \sqrt{\sigma_{kk}}$.

2. Let $\tilde{\sigma}_{k,S} = \text{Cov}(X_k, S) = \sum_{j=1}^d \sigma_{kj}$. The aggregate ES of S can be represented by

$$\text{ES}_\alpha(S) = \mu_S + \lambda_{g_1, \alpha} \cdot \sigma_S^*,$$

where $\mu_S = \sum_{j=1}^d \mu_j$ and $\sigma_S^* = \sqrt{\sum_{i=1}^d \sum_{j=1}^d \sigma_{ij}} = \sqrt{\sum_{k=1}^d \tilde{\sigma}_{k,S}}$.

3. The risk contribution of X_k to the total ES, $\text{ESC}_\alpha(X_k)$, for $k = 1, \dots, d$ can be represented by

$$\begin{aligned} \text{ESC}_\alpha(X_k) &= \mu_k + \lambda_{g_1, \alpha} \cdot \sigma_k^* \cdot \rho_{k,S} \\ &= \mu_k + \lambda_{g_1, \alpha} \cdot \tilde{\sigma}_{k,S} / \sigma_S^*, \end{aligned}$$

where $\rho_{k,S} = \frac{\tilde{\sigma}_{k,S}}{\sigma_k^* \sigma_S^*}$.

Proof. Suppose that $\mathbf{X} \sim \text{E}_d(\boldsymbol{\mu}, \Sigma, g_d)$. Since Theorem 4.2 is originated from that in Landsman and Valdez (2003, Theorem 2 and Theorem 3), the proof of aggregate ES and ESC is done by changing λ to $\lambda_{g_1, \alpha}$ with corresponding corrections that $\sigma^2 \lambda = \sigma \lambda_{g_1, \alpha}$.

Let $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)'$, $(\Sigma)_{ij} = \sigma_{ij}$ and $\sigma_k^* = \sqrt{\sigma_{kk}}$. Then it is easy to show by Theorem 2.18 that the individual ES of X_k for $k = 1, \dots, d$ is given by

$$\text{ES}_\alpha(X_k) = \mu_k + \lambda_\alpha \cdot \sigma_k^*$$

since $X_k \sim E(\mu_k, \sigma_k^{*2}, g_1)$. □

Appendix E

Proof of Proposition 4.3

For all elliptical distributions, the internal beta, which is the ratio of ESC to total ES, is identical to the beta of CAPM, which represents the proportion of the total risk capital allocated back to the specific component.

Proof. Suppose that $\mathbf{X} \sim E_d(\boldsymbol{\mu}, \Sigma, g_d)$ where $\boldsymbol{\mu} = (\mu_1, \dots, \mu_d)'$ and $(\Sigma)_{ij} = \sigma_{ij}$. Denote the k -th component of \mathbf{X}_t by $X_{k,t} = \mathbf{e}'_k \mathbf{X}_t$, and the sum of all components by $S_t = \sum_{j=1}^d X_{j,t} = \mathbf{1}' \mathbf{X}$, where \mathbf{e}_k is the vector of zeros but one at the k -th location and $\mathbf{1} = (1, \dots, 1)'$ the vector of ones with dimension d . Define the λ -factor as

$$\lambda_\alpha = \frac{\overline{G}(\frac{1}{2}z_\alpha^2)}{1 - \alpha} \quad (\text{E.1})$$

where $z_\alpha = \text{VaR}_\alpha(Z)$ with $Z \sim E(0, 1, g_1)$ and $\overline{G}(x) = G(\infty) - G(x)$ is the cumulative generator defined in Definition 2.16. Let $\tilde{\sigma}_{k,S} = \text{Cov}(X_k, S) = \sum_{j=1}^d \sigma_{kj}$.

Then by Theorem 4.2 under condition (2.37), for $k = 1, \dots, d$, we have The aggregate ES of S can be represented by

$$\text{ES}_\alpha(S) = \mu_S + \lambda_\alpha \cdot \sigma_S^*, \quad (\text{E.2})$$

where $\mu_S = \sum_{j=1}^d \mu_j$ and

$$\begin{aligned} \sigma_S^* &= \sum_{i=1}^d \sum_{j=1}^d \sigma_{ij} \\ &= \sum_{k=1}^d \tilde{\sigma}_{k,S}. \end{aligned}$$

and

$$\begin{aligned}\text{ESC}_\alpha(X_k) &= \mu_k + \lambda_\alpha \cdot \sigma_k^* \cdot \rho_{k,S} \\ &= \mu_k + \lambda_\alpha \cdot \tilde{\sigma}_{k,S} / \sigma_S^*,\end{aligned}\tag{E.3}$$

where

$$\rho_{k,S} = \frac{\tilde{\sigma}_{k,S}}{\sigma_k^* \sigma_S^*}.$$

Define the *internal beta* by

$$\begin{aligned}\beta_k &= \frac{\text{ESC}_\alpha(X_k) - \mu_k}{\text{ES}_\alpha(S) - \mu_S} \\ &= \frac{\tilde{\sigma}_{k,S}}{(\sigma_S^*)^2}.\end{aligned}\tag{E.4}$$

Then the internal beta is exactly the same as the beta in the CAPM. □

Appendix F

Proof of Theorem 4.5

Suppose that $\mathbf{Z}_t \sim E_d(\mathbf{0}, W_d, g_d)$ such that $\text{Cov}(\mathbf{Z}_t) = I_d$ is a function of W_d , providing that the density generator g_d exists. Assume that the condition (2.37), which guarantees the existence of the mean vector such that $E(\mathbf{Z}) = \mathbf{0}$, holds. Define $\lambda_{g,\alpha} = \frac{\overline{G}(\frac{1}{2}z_\alpha^2)}{1-\alpha}$ where $G(\cdot)$ is the cumulative generator defined in Definition 2.16 and $z_\alpha = \text{VaR}_\alpha(Z)$ with $Z \sim E(0, 1, g_1)$.

Then for a general MGARCH process $\{\mathbf{X}_t\}_{t \in \mathbb{Z}}$ of the form $\mathbf{X}_t = \Sigma_t^{1/2} \mathbf{Z}_t$, we have the following results.

1. For $k = 1, 2, \dots, d$, the individual ES of $X_{k,t+1}$ conditional on the history \mathcal{F}_t is given by

$$\text{ES}_\alpha(X_{k,t+1}|\mathcal{F}_t) = \lambda_{g_1,\alpha} v_{t+1,k}^*,$$

where $v_{t+1,k}^* = \sqrt{v_{t+1,kk}}$.

2. Let $\tilde{v}_{t+1,kS} = \text{Cov}(X_{t+1,k}, S_{t+1}) = \sum_{j=1}^d v_{t+1,kj}$. The aggregate ES is of the form

$$\text{ES}_\alpha(S_{t+1}|\mathcal{F}_t) = \lambda_{g_1,\alpha} v_{S_{t+1}}^*,$$

where $v_{S_{t+1}}^* = \sqrt{\sum_{i=1}^d \sum_{j=1}^d v_{t+1,kj}} = \sqrt{\sum_{k=1}^d \tilde{v}_{t+1,kS}}$.

3. ESC for the k -th component of \mathbf{X}_{t+1} is of the form

$$\text{ESC}_\alpha(X_{k,t+1}|\mathcal{F}_t) = \lambda_{g_1,\alpha} v_{t+1,k}^* \cdot \rho_{t+1,k}$$

where $\rho_{t+1,k} = \frac{\tilde{v}_{t+1,kS}}{v_{t+1,k}^* v_{S_{t+1}}^*}$ or equivalently $\rho_{t+1,k} = \frac{\tilde{v}_{t+1,kS}}{v_{t+1,k}^* \sqrt{\sum_{i=1}^d \tilde{v}_{t+1,kS}}}$.

Proof. consider that the innovations of a MGARCH processes are from a multivariate elliptical distribution with mean and variance-covariance matrix I_d such as multivariate Gaussian or multivariate t . Precisely, $\mathbf{Z}_t \sim E_d(\mathbf{0}, W_d, g_d)$ providing that the density generator g_d exists

such that $\text{Cov}(\mathbf{Z}_t) = I_d$ is a function of W_d . It follows that the distribution of \mathbf{X}_t conditional on \mathcal{F}_{t-1} turns out to be $E_d(\mathbf{0}, V_t, g_d)$ with

$$\begin{aligned} V_t &= \Sigma_t^{1/2} W_d (\Sigma_t^{1/2})' \\ &= \Sigma_t^{1/2} W_d \Sigma_t^{1/2}. \end{aligned}$$

Take notice of that a sequence of the form $\{\mathbf{v}' \cdot \mathbf{X}_t\}$ may not be a univariate GARCH process and that it depends on the structure of Σ_t and W_d . Let $v_{t,ij}$ denote the ij -th element of V_t , and $v_{t,k}^* = \sqrt{v_{t,kk}}$. The following results are thus obtained by applying Theorem 4.2 to conditional distributions. Define

$$\lambda_{g,\alpha} = \frac{\overline{G}(\frac{1}{2}z_\alpha^2)}{1-\alpha}$$

where $G(\cdot)$ is the cumulative generator defined in Section 2.3.3 and $z_\alpha = \text{VaR}_\alpha(Z)$ with $Z \sim E(0, 1, g_1)$.

1. For $k = 1, 2, \dots, d$, the k -th component of $X_{k,t+1} = \mathbf{e}'_k \mathbf{X}_t$ conditional on the history \mathcal{F}_t is $E(\mathbf{e}'_k \mathbf{0} \mathbf{e}_k, \mathbf{e}'_k V_{t+1} \mathbf{e}_k) = E(0, v_{t,k}^*, g_1)$. By Theorem 4.2, for $k = 1, 2, \dots, d$, the individual ES of $X_{k,t+1}$ conditional on the history \mathcal{F}_t is given by

$$\text{ES}_\alpha(X_{k,t+1} | \mathcal{F}_t) = v_{t+1,k}^* \cdot \lambda_{g_1, \alpha}. \quad (\text{F.1})$$

2. Let $S_t = \mathbf{1}' \cdot \mathbf{X}_t = \sum_{i=1}^d X_{i,t}$ denote the sum of components of \mathbf{X}_t . Then the conditional distribution of S_{t+1} is to be $E(\mathbf{1}' \mathbf{0} \mathbf{1}, \mathbf{1}' V_{t+1} \mathbf{1})$ and hence $E(0, v_{S_{t+1}}^*)$ where

$$\begin{aligned} v_{S_{t+1}}^* &= \sqrt{\sum_{i=1}^d \sum_{j=1}^d v_{t+1,ij}} \\ &= \sqrt{\sum_{k=1}^d \tilde{v}_{t+1,kS}}. \end{aligned}$$

Thus by Theorem 4.2, the aggregate ES is of the form

$$\text{ES}_\alpha(S_{t+1} | \mathcal{F}_t) = v_{S_{t+1}}^* \cdot \lambda_{g_1, \alpha}. \quad (\text{F.2})$$

3. By Theorem 4.2, the ESC for the k -th component of \mathbf{X}_{t+1} is to be

$$\text{ESC}_\alpha(X_{k,t+1} | \mathcal{F}_t) = v_{t+1,k}^* \cdot \rho_{t+1,k} \cdot \lambda_{g_1, \alpha} \quad (\text{F.3})$$

where

$$\rho_{t+1,k} = \frac{\tilde{v}_{t+1,kS}}{v_{t+1,k}^* v_{S_{t+1}}^*}$$

or equivalently

$$\rho_{t+1,k} = \frac{\tilde{v}_{t+1,kS}}{v_{t+1,k}^* \sqrt{\sum_{i=1}^d \tilde{v}_{t+1,kS}}}$$

□

Appendix G

Proof of Proposition 3.2

We want to show that

$$\begin{aligned}
 \chi_{\text{ind}}^2 &= \sum_{i=0}^1 \sum_{j=0}^1 \frac{(n_{ij} - R_i C_j / N)^2}{R_i C_j / N} \\
 &= \frac{N(n_{00}n_{11} - n_{01}n_{10})^2}{R_0 R_1 C_0 C_1} \\
 &\xrightarrow{d} \chi_{(1)}^2 \quad \text{as } N \rightarrow \infty
 \end{aligned} \tag{G.1}$$

according to the contingency table, Table 3.21.

Proof. Denote hits (violation indicators) by $X_t \sim \text{i.i.d. Bernoulli}(p)$; $t > 0$. Define new indicators by $Y_{i,j,t} = \mathbf{1}_{X_t=i \cap X_{t+1}=j}$; $i, j \in \{0, 1\}, t > 0$. Let $p_i = p^i(1-p)^{1-i}$; $i \in \{0, 1\}$. Obviously, $p_0 + p_1 = 1$. Then we have moments in the following:

$$\begin{aligned}
 \mathbb{E}(Y_{i,j,t}) &= \mathbb{P}(X_t = i \cap X_{t+1} = j) = p^i(1-p)^{1-i}p^j(1-p)^{1-j} = p_i p_j; \\
 \mathbb{E}(Y_{i,j,t} Y_{i',j',t}) &= \delta_{i,i'} \delta_{j,j'} p_i p_j; \\
 \mathbb{E}(Y_{i,j,t} Y_{i',j',t+1}) &= \delta_{i',j} \mathbb{P}(X_t = i \cap X_{t+1} = j \cap X_{t+2} = j') = \delta_{i',j} p_i p_j p_{j'}; \\
 \mathbb{E}(Y_{i,j,t} Y_{i',j',t-1}) &= \mathbb{E}(Y_{i,j,t+1} Y_{i',j',t}) = \delta_{i,j'} p_i p_j p_{i'}; \\
 \mathbb{E}(Y_{i,j,t} Y_{i',j',t'}) &= p_i p_j p_{i'} p_{j'}, \quad |t - t'| > 1.
 \end{aligned}$$

Therefore,

$$\begin{aligned}
\chi_{\text{ind}}^2 &= \frac{N(n_{00}n_{11} - n_{01}n_{10})^2}{R_0 R_1 C_0 C_1} \\
&\approx \frac{N(\Delta n_{00} N p_1^2 + \Delta n_{11} N p_0^2 - \Delta n_{01} N p_0 p_1 - \Delta n_{10} N p_0 p_1)^2}{N^4 p_0^2 p_1^2} \quad (\text{by Slutsky's theorem}) \\
&= \frac{(\Delta n_{00} p_1^2 + \Delta n_{11} p_0^2 - \Delta n_{01} p_0 p_1 - \Delta n_{10} p_0 p_1)^2}{N p_0^2 p_1^2} \\
&= \left(\frac{\Delta n_{00} p_1^2 + \Delta n_{11} p_0^2 - \Delta n_{01} p_0 p_1 - \Delta n_{10} p_0 p_1}{\sqrt{N} p_0 p_1} \right)^2,
\end{aligned}$$

where

$$\Delta n_{ij} = n_{ij} - N p_i p_j.$$

Write

$$Z_{\text{ind}} = \frac{\Delta n_{00} p_1^2 + \Delta n_{11} p_0^2 - \Delta n_{01} p_0 p_1 - \Delta n_{10} p_0 p_1}{\sqrt{N} p_0 p_1}.$$

Now we argue that if Z_{ind} is asymptotically Normal, then it will follow that Z_{ind}^2 is asymptotically χ_1^2 by the Continuous Mapping Theorem (CMT) (DasGupta, 2008, p.8, Theorem 1.14). Because the indicators $Y_{i,j,t}$ are not independent, the standard Central Limit Theorem (CLT) is not applicable to assert asymptotic normality of Z_{ind} ; however, they are m -dependent¹ (specifically, 1-dependent), so instead a CLT for α -mixing² sequences (Billingsley, 1995, p.364, Theorem 27.4) can be used, since an m -dependent sequence is α -mixing with $\alpha_n = 0$ for $n > m$. To verify the asymptotically normality, re-write Z_{ind} as

$$Z_{\text{ind}} = S_N / (\sqrt{N} p_0 p_1)$$

where

$$\begin{aligned}
V_t &= p_1^2 (Y_{0,0,t} - p_0^2) + p_0^2 (Y_{1,1,t} - p_1^2) - p_0 p_1 (Y_{0,1,t} - p_0 p_1) - p_1 p_0 (Y_{1,0,t} - p_1 p_0) \\
&= p_1^2 Y_{0,0,t} + p_0^2 Y_{1,1,t} - p_0 p_1 Y_{0,1,t} - p_1 p_0 Y_{1,0,t}
\end{aligned}$$

such that $S_N = \sum_{t=1}^N V_t$. Besides 1-dependency, V_t is stationary because the distribution of the random vector $(V_n, V_{n+1}, \dots, V_{n+j})$ does not depend on n . It is easy to verify that $E(V_t) = 0$,

¹A sequence of random variables, $\{X_t\}$, is m -dependent if (X_1, \dots, X_k) and $(X_{k+n}, \dots, X_{k+n+l})$ whenever $n > m$ for a given fixed m .

²A sequence of random variables, $\{X_n\}$, is said to be α -mixing, if

$$\Pr(A \cap B) - \Pr(A) \Pr(B) \leq \alpha_n \rightarrow 0$$

for a number α_n , $A \in \sigma(X_1, \dots, X_k)$, $B \in \sigma(X_{k+n}, X_{k+n+1}, \dots)$, and $k \geq 1$, $n \geq 1$.

and hence $\text{Var}(V_t) = \mathbb{E}(V_t^2)$, $\mathbb{E}(Z_{\text{ind}}) = 0$. Since the Y s are indicators of a partition, we have $Y_{i,j,t}^2 = Y_{i,j,t}$ and $Y_{i,j,t}Y_{i',j',t} = \delta_{i,i'}\delta_{j,j'}$. It then follows that

$$V_t^2 = p_1^4 Y_{0,0,t} + p_0^4 Y_{1,1,t} + (p_0 p_1)^2 Y_{0,1,t} + (p_1 p_0)^2 Y_{1,0,t},$$

and

$$\begin{aligned} \mathbb{E}(V_t^2) &= p_1^4 p_0^2 + p_0^4 p_1^2 + (p_0 p_1)^2 p_0 p_1 + (p_1 p_0)^2 p_1 p_0 \\ &= p_0^2 p_1^2 (p_1^2 + p_0^2 + 2p_0 p_1) \\ &= p_0^2 p_1^2. \end{aligned}$$

Similarly, $Y_{i,j,t}Y_{i',j',t+1} = 0$ if $i' \neq j$, so

$$\begin{aligned} V_t V_{t+1} &= p_1^2 Y_{0,0,t} p_1^2 Y_{0,0,t+1} - p_1^2 Y_{0,0,t} p_0 p_1 Y_{0,1,t+1} \\ &\quad + p_0^2 Y_{1,1,t} p_0^2 Y_{1,1,t+1} - p_0^2 Y_{1,1,t} p_1 p_0 Y_{1,0,t+1} \\ &\quad - p_0 p_1 Y_{0,1,t} p_0^2 Y_{1,1,t+1} + p_0 p_1 Y_{0,1,t} p_1 p_0 Y_{1,0,t+1} \\ &\quad - p_1 p_0 Y_{1,0,t} p_1^2 Y_{0,0,t+1} + p_1 p_0 Y_{1,0,t} p_0 p_1 Y_{0,1,t+1} \end{aligned}$$

and

$$\mathbb{E}(V_t V_{t+1}) = 0.$$

Hence we have $\text{Var}(Z_{\text{ind}}) = 1$. Now let $\sigma^2 = \mathbb{E}(V_1^2) + 2 \sum_{t=1}^{\infty} \mathbb{E}(V_1 V_{1+k})$ and then we have

$$\sigma = \sqrt{\mathbb{E}(V_1^2) + 2\mathbb{E}(V_1 V_2)} = p_0 p_1 > 0.$$

Then by the CLT for α -mixing sequences, $Z_{\text{ind}} = S_N/(\sigma\sqrt{N})$ is asymptotically $N(0, 1)$ and hence Z_{ind}^2 is asymptotically χ_1^2 by CMT. Therefore, χ_{ind}^2 is asymptotically χ_1^2 . \square

Appendix H

Proof of Proposition 3.3

We want to show that

$$\begin{aligned}\chi_{\text{joint}}^2 &= \sum_{i=0}^1 \sum_{j=0}^1 \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \\ &\xrightarrow{d} \chi_{(2)}^2 \quad \text{as } N \rightarrow \infty\end{aligned}\tag{H.1}$$

where $e_{ij} = R_i \cdot p_j$ with $p_j = \alpha^{1-j}(1 - \alpha)^j$ for $j \in \{0, 1\}$ according to the contingency table, Table 3.21.

Proof. Recall definitions of n_{ij} , Δn_{ij} , $Y_{i,j,t}$, and the moments of $Y_{i,j,t}$ from Appendix G. Rewrite χ_{joint}^2 as follows.

$$\begin{aligned}\chi_{\text{joint}}^2 &= \sum_{i=0}^1 \sum_{j=0}^1 \frac{(n_{ij} - R_i p_j)^2}{R_i p_j} \\ &= \sum_{i=0}^1 \frac{(n_{i0} p_1 - n_{i1} p_0)^2}{R_i p_0 p_1} \\ &= \sum_{i=0}^1 \frac{(\Delta n_{i0} p_1 - \Delta n_{i1} p_0)^2}{R_i p_0 p_1} \\ &= \sum_{i=0}^1 \frac{(\Delta n_{i0} p_1 - \Delta n_{i1} p_0)^2}{N p_i p_0 p_1} / \left(\frac{R_i/N}{p_i} \right) \\ &\approx \sum_{i=0}^1 \frac{(\Delta n_{i0} p_1 - \Delta n_{i1} p_0)^2}{N p_i p_0 p_1}\end{aligned}$$

(since $\frac{R_i/N}{p_i} \rightarrow 1$ in law as $N \rightarrow \infty$)

$$= \sum_{i=0}^1 \left(\frac{\Delta n_{i0} p_1 - \Delta n_{i1} p_0}{\sqrt{N p_i p_0 p_1}} \right)^2.$$

Let

$$Z_{\text{joint},i} = \frac{\Delta n_{i0} p_1 - \Delta n_{i1} p_0}{\sqrt{N p_i p_0 p_1}}$$

and

$$\mathbf{Z}_{\text{joint}} = \begin{pmatrix} Z_{\text{joint},0} \\ Z_{\text{joint},1} \end{pmatrix}.$$

Clearly $E(\mathbf{Z}_{\text{joint}}) = \mathbf{0}$, and it is easy to show that

$$\text{Var}(\mathbf{Z}_{\text{joint}}) = \mathbf{I}_2.$$

Write $W_i = Z_{\text{joint},i} / \sqrt{\left(\frac{R_i/N}{p_i}\right)}$ for $i \in \{0,1\}$. Now we argue that if $\mathbf{Z}_{\text{joint}}$ is asymptotically $N(\mathbf{0}, \mathbf{I}_2)$, then

$$\mathbf{W} = \begin{pmatrix} W_0 \\ W_1 \end{pmatrix}$$

is also asymptotically $N(\mathbf{0}, \mathbf{I}_2)$ by Slutsky's Theorem, because $\frac{R_i/N}{p_i} \rightarrow 1$ in probability as $N \rightarrow \infty$ for all i , and hence $\sqrt{\left(\frac{R_i/N}{p_i}\right)} \rightarrow 1$ in probability as well by the Continuous Mapping Theorem (CMT). Furthermore, if \mathbf{W} is asymptotically $N(\mathbf{0}, \mathbf{I}_2)$, then $\chi_{\text{joint}}^2 = \mathbf{W}'\mathbf{W}$ is asymptotically χ_2^2 by CMT. Thus the remaining is to verify the asymptotic normality of $\mathbf{Z}_{\text{joint}}$. We then argue that if $\mathbf{c}'\mathbf{Z}_{\text{joint}}$ is asymptotically $N(\mathbf{c}'\mathbf{0}, \mathbf{c}'\mathbf{I}_2\mathbf{c}) = N(0, (a^2 + b^2))$ for any given $\mathbf{c} = (a, b)' \in \mathbb{R}^2$, then the characteristic function of $\mathbf{Z}_{\text{joint}}$ converges pointwise to that of $N(\mathbf{0}, \mathbf{I}_2)$, and it implies that $\mathbf{Z}_{\text{joint}}$ is asymptotically $N(\mathbf{0}, \mathbf{I}_2)$ by Levy's continuity theorem.

Hence we still need to show that $\mathbf{c}'\mathbf{Z}_{\text{joint}}$ is asymptotically $N(0, (a^2 + b^2))$ for any given

$\mathbf{c} = (a, b)' \in \mathbb{R}^2$. For any given $\mathbf{c} = (a, b)' \in \mathbb{R}^2$,

$$\begin{aligned}
\mathbf{c}'\mathbf{Z}_{\text{joint}} &= (a, b)\mathbf{Z}_{\text{joint}} \\
&= \frac{1}{\sqrt{N}} \sum_{t=1}^N (a, b) \begin{pmatrix} V_{0,t}/\sigma_0 \\ V_{1,t}/\sigma_1 \end{pmatrix} \\
&= \frac{1}{\sqrt{N}} \sum_{t=1}^N (a \cdot V_{0,t}/\sigma_0 + b \cdot V_{1,t}/\sigma_1) \\
&= \frac{1}{\sqrt{N}} \sum_{t=1}^N U_t \\
&= \frac{1}{\sqrt{N}} S_N
\end{aligned}$$

where $V_{i,t} = \Delta n_{i0}p_1 - \Delta n_{i1}p_0$ and $\sigma_i = \sqrt{p_i p_0 p_1}$ for $i \in \{0, 1\}$, and $U_t = a \cdot V_{0,t}/\sigma_0 + b \cdot V_{1,t}/\sigma_1$ and $S_N = \sum_{t=1}^N U_t$. It is easy to verify that U_t is stationary and 1-dependent with $E(U_t) = 0$. Furthermore, $E(U_1^2) = a^2 + b^2$ and $E(U_1 U_2) = 0$. Let $\sigma^2 = E(U_1^2) + 2 \sum_{t=1}^{\infty} E(U_1 U_{1+k})$ and hence we have

$$\sigma = \sqrt{E(U_1^2) + 2E(U_1 U_2)} = \sqrt{a^2 + b^2} > 0.$$

Then by the CLT for α -mixing sequences, S_N/\sqrt{N} is asymptotically $N(0, \sigma^2)$. \square