University of Venda

MODELLING VOLATILITY, EQUITY RISK AND EXTREMAL DEPENDENCE OF THE BRICS STOCK MARKETS

By

Rosinah Mphedziseni Mukhodobwane

8500656

submitted in fulfillment of the requirements for
the degree of Doctor of Philosophy in Mathematics
in the
Department of Mathematical and Computational Sciences
Faculty of Science, Engineering and Agriculture
University of Venda
Thohoyandou, South Africa

Promoter: Dr. Caston Sigauke
Co-Promoter: Dr. Wilbert Chagwiza
Co-Promoter: Prof. Winston Garira

Date submitted: January 2021

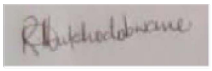
Abstract

With the use of empirical data of the BRICS (Brazil, Russia, India, China, and South Africa) stock markets, this thesis focuses on solving three main financial and investment issues involving returns volatility, risk and extremal dependence via robust statistical modelling. The first issue involves modelling financial returns volatility (when the true distribution is unknown) using the univariate GARCH model under the assumptions of seven error distributions. The findings, using two of the error distributions, show that the Chinese market has the highest volatility persistence, followed by the South African, Russian, Indian and Brazilian markets in that order. For risk modelling and analysis, the findings show that the Russian market has the highest risk level, followed by the South African, Chinese, Brazilian and Indian markets, respectively. For the extremal dependence modelling, using the bivariate point process and conditional multivariate extreme value (CMEV) models, the findings show varied levels of low extremal dependence structure whose outcomes are highly beneficial to investors, portfolio managers and other market participants who are interested in maximising their investment returns and financial gains. However, it is observed that the point process was able to model many more extreme observations or exceedances that contribute to the likelihood estimation and it gives more information than the threshold excess method of the CMEV model.

Keywords: Conditional extreme value model, Equity markets, Equity-risk, GARCH model, Point process, Risk management, Volatility.

Declaration

I, Rosinah Mphedziseni Mukhodobwane (student number: 8500656), hereby declare that the thesis titled: "Modelling volatility, equity risk and extremal dependence of the BRICS stock markets" for the Doctor of Philosophy in the Department of Mathematical and Computational Sciences at the University of Venda, hereby submitted by me, has not been submitted for any degree at this or any other university, that it is my own work in design and in execution, and that all reference material contained therein has been duly acknowledged.

Signature:  ..... Date: 07/11/2021 .....

Dedication

To my husband, Nditsheni Lawrence, children, Hangwani, Vele, Mukondeleli and Thakhani and granddaughters, Tondani and Mukundi.

Acknowledgements

I would like to express my special appreciation and thanks to my supervisors: Dr. Caston Sigauke, Dr. Wilbert Chagwiza and Prof. Winston Garira. They have been tremendous mentors to me and I would like to thank them for their valuable guidance, scholarly inputs and consistent encouragement throughout the research period. Your patience, motivation and immense knowledge helped me during the research work and while writing the thesis.

Completing this work would have been more difficult if it was not for the backing and friendly support provided by my colleagues from the department of Mathematics and Applied Mathematics at the University of Venda. I am indebted to the team of postgraduate students who contributed to this research. I am especially grateful to Richard Abayomi Samuel, Dr. Daphney Mathebula, Azwindini Maphiri, Nancy Mukwevho and Rendani Netshikweta for their extended support in a very special way. I gained a lot through their personal suggestions and scholarly interactions at various points of the research programme.

Special thanks are due to my husband, children and lovely granddaughters. I thank you all for your continuous support, understanding and encouraging words even when things were tough to handle. Above all, I give my profound thanks and gratitude to the Almighty God for granting me the wisdom, health and strength to undertake this research task and enabling me through to its completion.

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Acronyms

ACF	Autocorrelation function
AD	Anderson Darling
ADF	Augmented Dick-Fuller
AIC	Akaike Information Criterion
AR	Autoregressive
ARCH	Autoregressive Conditional Heteroscedasticity
ARCH-LM	Autoregressive Conditional Heteroscedasticity Lagrange Multiplier
ARMA	Autoregressive Moving Average
AVE	Average
BIC	Bayesian Information Criterion
BMM	Block Maxima Method
BRIC	Brazil, Russia, India and China
BRICS	Brazil, Russia, India, China and South Africa
CAC	Cotation Assistée en Continu

CEV	Conditional Extreme Value
CGARCH	Component GARCH
CMEV	Conditional Multivariate Extreme Value
CoVaR	Conditional Value-at-Risk
CUEV	Conditional Univariate Extreme Value
CRA	Contingent Reserve Arrangement
CVM	Cramer-von-Mises
CV	Critical value
DAX	Deutscher Aktien Index
DCC-FIAPARCH	Dynamic Conditional Correlation Fractional Integrated Asymptotic Power Autoregressive Conditional Heteroscedasticity
DF	Dickey-Fuller
EDA	Exploratory Data Analysis
EDF	Empirical Distribution Function
EGARCH	Exponential generalised Autoregressive
ENV	Envelope
ES	Estimated Shortfall
EVT	Extreme Value Theory
FEV	Frequent Evidence
FTSE	Financial Times Stock Exchange Group
GARCH	generalised Autoregressive Conditional Heteroscedasticity
GARCH-M	GARCH-in-mean

GAS	generalised Autoregressive Score
GCC	Gulf Corporation Council
GCV	generalised Cross Validation
GDP	Gross Domestic Product
GED	generalised Error Distribution
GEVD	generalised Extreme Value Distribution
GFC	Global Financial Crisis
GHYP	generalised Hyperbolic Distribution
GIGD	generalised Inverse Gaussian distribution
GJR-GARCH	Glosten-Jagannathan-Runkle GARCH
GPD	generalised Pareto Distribution
HQIC	Hannan Quinn Information Criterion
IBOVESPA	Index of Brazil Sau Paulo
i.i.d	Independent and Identically Distributed
IMF	International Monetary Fund
IMOEX	Index of Moscow Exchange
JALSH	JSE African All Share
JB	Jarque-Bera
JSE	Johannesburg Stock Exchange
JSE-ALSI	Johannesburg Stock Exchange All Share Index
KPSS	Kwiatkowski, Phillips, Schmidt and Shin
K-L	Kullback-Leibler

LM	Lagrange Multiplier
LRT	Likelihood Ratio Tests
MA-GARCH	Moving Average GARCH
MCS	Multivariate Conditional Spearman
MEVT	Multivariate Extreme Value Theory
MGARCH	Multivariate GARCH
MLE	Maximum Likelihood Estimation
NASDAQ	National Association of Securities Dealers Automated Quotations
NDB	New Development Bank
NIFTY	National Stock Exchange of India
PACF	Partial Autocorrelation function
PGARCH	Power GARCH
PI	Prediction Interval
PINAW	Prediction Interval Normalised Average Width
PIW	Prediction Interval Width
POT	Peak Over Threshold
PP	Phillip Perron
QQ	Quantile to Quantile
RTSI	Russia Trading System index
SHCOMP	Stock Exchange Composite of China
SIC	Shibata Information Criterion

SSE	Shanghai Stock Exchange
TGARCH	Threshold GARCH
VaR	Value-at-Risk
VAR	Vector Autoregressive

Research Outputs

A list of research outputs from this thesis is given below.

Peer Reviewed Journal Publications

1. Mukhodobwane R. M., Sigauke C., Chagwiza W. and Garira W. (2020)., *Modelling volatility of the BRICS stock markets*, Journal of Statistics Optimization and Information Computing, 3, 749-772.

2. Mukhodobwane R M., Sigauke C., Chagwiza W. and Garira W. *Stochastic modelling of the BRICS stock markets risks*, Journal of Statistics applications and probability. (Accepted on 3 September 2020).

3. Mukhodobwane R M., Sigauke C., Chagwiza W. and Garira W. (2020). *Modelling and experimental analysis of the BRICS stock markets' extremal dependence* (To be sent to the journal for peer-review)

Chapter 1

Introduction

1.1 Overview

After series of global financial crises to date, empirical researchers are coming up with continuous modelling of rare events (risk) and tail dependence in the financial field. In particular, investors and traders are deeply concerned with daily activities and any unusual (or rare) movement in the markets where investment and trading occur.

This study uses both univariate and multivariate statistical structures of two extreme value theory (EVT) models for modelling rare events and their asymptotic dependencies in the financial returns of the BRICS regional markets.

1.1.1 The BRICS countries

The British Economist Jim O'Neill conceptualised the word BRICS as an acronym of five key emerging regional economies of Brazil, Russia, India, China and South Africa. Prior to the joining of South Africa in 2010, the first four nations were formerly called "BRIC". The term BRIC came into light in a paper of Goldman Sachs (Wilson and Puroshothaman, 2003) where long-term economic growth rates were projected for the four nations till 2050 (Nassif et al., 2016). The BRICS nations cover 40% (approximately) of the world's population, create an approximate 20% of

global output (IMF, 2013), and have an estimate of US$4.7tn in joint foreign reserves (Peng, 2015). The BRICS economies' Growth Domestic Product (GDP) amounted to $13.2tn in 2011, which is fairly over the Eurozone's of $13.1tn and not far off from the U.S. of $15.1tn (Emerson, 2012).

The BRICS are developing nations with high potentials and comparative good economic performance regionally and globally (Tian, 2016). During the sixth annual BRICS leaders' summit held in 2014, a decision was taken to introduce the BRICS Development Bank with US$100bn capitalization. In addition to this move, the BRICS Contingent Reserve Arrangement (CRA) was also initiated with US$100bn to start with. These initiatives were meant to augment the current IMF's (International Monetary Fund) global activities, hence strengthening global financial security (Tian, 2016).

In 2015, the BRICS proposed plans for economic, investment and trade cooperation until 2020 in view of having a unified mega market with a land-air-sea connection network (Tian, 2016). Presently, the BRICS Development Bank which is currently known as the "New Development Bank" (NDB), and the CRA are actively operating after joint approval by all BRICS nations. In particular, the BRICS states have reinforced their trade and economic relations through concesus approval of bilateral trade and cross-regional agreements (Tian, 2016).

1.1.2 Equity risk and extremal dependence

Risks are adverse market movements that are of major concern to investors and risk managers. The practice of advanced detection of risks, analysing them and taking preventive measures to alleviate the risks is known as risk management. Financial risks are posed by rare or extreme events existence at the tails of the returns' distributions and the behaviour of an investment portfolio during a financial crisis is an important element of risk management.

The study of risk in the tails of equity returns is a crucial area in stock markets research because the general view about investing is that the future could either yield good profits or worrisome losses as an aftermath of investment. Rare events may materialise as large negative or positive investment returns, major defaults, the collapse of risky asset prices, or a stock market crash (Uppal, 2013). Studies on extreme events usually involve areas such as extreme heat waves, catastrophic structural failures,, natural disasters such as heavy rains , storms and floods, currency crises, stock market crashes etc.

The relationship between the markets at the tail region will shed light into their extremal behaviour, which is termed extremal dependence in the literature. Good understanding of how the markets co-move in extremal dependence may result in better hedging strategies and portfolio handling for investors.

Two or more financial markets can move together at the tails of the returns distribution as a result of contagion effect, stock market characteristics and economic integration (Pretorius, 2002). Volatility spill-over as measured by shocks transmission from market to market increases in periods of crises and it provides evidence of contagion effects through financial markets (Hammoudeh et al., 2016). Contagion happens when investors rapidly change their positions, because they think a crisis in one market can sooner or later affect the others, which may ultimately lead to panic selling or herd behaviour effects. Fundamental linkages between equity markets usually originate from economic linkages existing between economies, and it indicates the extent to which nations have close trade, investment and other economic ties. Contagion effect is not quantifiable in itself but rather assessed with the residual from the co-movement that fundamentals cannot explain (Pretorius, 2002).

1.1.3 Portfolio diversification

Investors, traders and fund managers practice risk management in order to control and minimise their investment risks locally and internationally. Inadequate plans for risk management could be tragic on investment during financial mayhem, and could result in a miserable end for an investor. Based on this, portfolio diversification is of essence to traders, equity investors and fund managers so as to minimise their risk exposures. Diversification of portfolio can potentially eradicate risk if returns are uncorrelated (Levy and Sarnat, 1970). As suggested by economic theory, individual investors (in open economies) regularly aim to move their resources between several investments, regions, industries and countries for maximum returns benefits (K'onya, 2015).

This study examines these five BRICS regional equity markets to ascertain the degree to which they are integrated cross-regionally and determine the extremal dependencies among them. Stock markets interdependence and integration through trading blocs (like the BRICS) indicate improved co-movements of markets which may in turn reduce potential opportunities accessible via portfolio diversification (Yu and Hassan, 2006). In order to have a successful portfolio diversification, a negative or low co-movement between equity markets is required to enable potential poor performance in one side of the equity market to be comparatively hedged by good returns on the other side of the paired markets. Low correlations can reinforce the concept of international (portfolio) diversification by increasing expected returns following substantial reduction in portfolio risk (Vo and Daly, 2005).

1.1.4 Stock market's volatility

Studies on volatility of stock markets has grown considerably with the general belief that the world's equity markets are becoming strongly related and more dependent

(Fayyad and Daly, 2010). Modelling and forecasting of financial time series volatility have gained so much research attention in the literature due to their relevance in financial and economic applications like portfolio optimization, asset pricing and risk management. Furthermore, it is believed that EVT modelling can be more suitably studied in markets with volatility (Poon et al., 2004 and Djakovic et al., 2011).

Time series volatility implies a situation where the conditional variance of the return series varies over time. This study will examine the dynamics of market returns volatility in the stock markets of the five BRICS nations. These markets are selected because they, as a team, are progressively and increasingly becoming a propelling force on global economic growth (Tian, 2016) and they are of particular interest to investors and practitioners. In particular, the four BRIC markets are considered the largest and fastest growing economies among the global emerging markets and are also reputed in the literature as the global economic growth's engines (Sashi, 2016).

The data that are used in this study are the daily market returns for the five BRICS indices. These data are analysed using the "generalised Autoregressive Conditional Heteroscedasticity (GARCH) model", while the conditional mean equation of the GARCH process is modelled using an ARMA (Autoregressive Moving Average) model. Engle (1982) was the proponent of the ARCH (autoregressive conditional heteroscedasticity) model, but it was later generalised by Bollerslev (1986) as generalised ARCH (GARCH). The GARCH model can be used to determine the source and magnitude of volatility. The volatility analysis of this study is designed to describe the econometric modelling of the conditional mean and volatility of equity returns from these selected markets.

Several empirical research has been carried out in the past on cross-country equity market's returns and volatility. These, among others, include Kennedy and Sharma (1977) for London, New York and Bombay; Grieb and Reyes (1999) for Brazil and Mexico; Chiang and Doong (2001) for equity markets of seven Asian countries. This

study further extends the scope of research done in the literature by modelling volatility of these BRICS markets under the assumption of four error distributions of a normal, student's t, Generalised Error Distribution (GED), generalised hyperbolic (GHYP) distribution, and the skew versions of the normal, student's t, and GED.

1.1.5 Extreme value theory models

The EVT has been described as an effective tool for modelling the stochastic behaviour of tail risk occasioned by rare events. Hence, EVT models are usually applied where consideration is given to unusually small or large movements at the tails of the distributions using extreme realizations. EVT approach is currently preferred to the use of the traditional Value at Risk (VaR) approach because the latter is formulated on the assumption of normal distribution, which is contrary to the fat-tailed property (i.e. high kurtosis) of financial returns. The outcome of data arising from fat-tailed distribution like financial returns are severely understated by model of risk with a normal distribution assumption (Mandelbrot, 1963; Mandelbrot, 1997 and Taleb, 2007). Based on this, EVT based method of VaR analysis is usually recommended in preference to the normal distribution approach of VaR (see Bali, 2003) in modelling financial risk.

The EVT approach is also preferred to using correlation coefficients in analysing market co-movements because correlation coefficients give the same weight to extreme realizations, which is ascribed as a limitation in their use (Fernandez, 2006). Hence, correlation coefficients are inaccurate dependence measure if tail or extreme realizations display different dependence pattern from the rest of the (central) sample (Fernandez, 2006). Within the general linear statistical model, correlation is a central measure. Correlation coefficients are governed by the assumptions of normality, linearity, and homoscedascity. The result of using correlation coefficients are usually inaccurate in a non-linear relationship. Linearity in relationship exists when

a change in a variable causes a constant proportional change in another variable. Financial time series has non-constant volatility termed heteroscedastic (or ARCH) effects which cannot be adequately modelled by correlation coefficients. As a result of this, recent financial studies lay emphasis on the relevance of tail's rare events and researchers in the field are resorting to the use of extreme value theory (see Longin and Solnik, 2001; Longin, 2000; McNeil and Frey, 2000).

Before applying EVT models for the risk and extremal dependence modelling, volatility is filtered out since it is known that the removal of short-range linear dependence and ARCH effects or heteroscedasticity from the financial returns makes the EVT modelling more satisfactory (Smith, 2003).

Two approaches of EVT were used in this study for modelling equity risk and extremal dependence in the tails of stock markets of the five BRICS economies. The first approach is the "conditional multivariate extreme value" model and the second approach is the "point process" method. The first approach fits a generalised Pareto distribution (GPD) model to excesses (or exceedances) above a suitably high threshold, while the second approach has a non-homogeneous Poisson process (Coles, 2001) as the limiting distribution for excesses above the same chosen threshold as the first approach.

1.2 Research questions

This research should result in responses to the following questions:

1. Which of the BRICS markets is most prone to risk, considering their risk hierarchy in a descending order of magnitudes?

2. Is portfolio diversification strongly, moderately or weakly realistic between each of the selected pairs of the BRICS regional markets?

3. What effect do the trade and economic relations of the BRICS economies have on extremal dependencies of the paired markets?

4. What factors facilitate the extremal or tail (asymptotic) dependencies of the selected paired markets, is it fundamental factors (like economic/trade link) or contagion effect?

5. Which of the two selected EVT models of conditional extreme value (CEV) and point process will be more suitable for modelling the risk and extremal dependence of the BRICS markets?

1.3 Aim and objectives

1.3.1 Aim of the study

The main aim is to model risk and determine the level (or strength) of extremal dependence occurrence in the pairwise combinations of the BRICS markets with the use of univariate and multivariate EVT (MEVT) model via the bivariate threshold excess method of GPD and point process.

1.3.2 Objectives of the study

The objectives of this study are, to:

1. determine volatility hierarchy of the BRICS markets in a descending order of magnitudes,

2. determine in order of hierarchy the risks associated with each of the markets,

3. model the extremal dependence structure and its effect on portfolio diversification in the markets,

4. examine the influence of the trade and economic linkages of the BRICS nations on their extremal dependence,

5. estimate the return levels of rare events in the tails of the distributions in each of the five markets.

1.4 Significance of the study

Quite a number of studies have been carried out on the BRICS economies (see K'onya, 2015; Peng, 2015; Nassif et al., 2016; Sashi, 2016 and Tian, 2016), but there is no evidence in the literature that any work has been done on extremal dependence of the markets using EVT's conditional extreme value (CEV) model and point process approach. Hence, the findings of this study will be of paramount interest to various market participants, academics, policy markers, portfolio managers, practitioners, traders and investors for crucial investment's decision making. These findings will add to already existing facts in the literature and also generate new unprecedented facts to the literature on risk and extremal dependence of these evolving markets.

1.5 Research contributions

The major contribution of this thesis is in Financial Mathematics and Applied Statistics. The contributions are as follows:

- With focus on the BRICS equity markets, this study has been able to identify the most appropriate heavy-tailed distribution, out of seven stated error distributions, for modelling the volatility of financial returns when the true distribution is unknown in each of the markets. The seven error distributions were applied with ARMA(1,1)-GARCH(1,1) model to model the prevailing volatility

in each market and the best fat-tailed error distribution that describes each of the markets was experimentally analysed and selected. Adequate knowledge of the level of volatility in a financial market is crucial for portfolio risk management, hence volatility modelling in the sphere of a suitable error distribution is key to investors' decision making.

The seven error distributions used include the normal, skewed-normal, Student's t, skewed-Student's t, GED, skewed-GED and the generalised hyperbolic (GHYP) distribution. The merit of using as many as seven assumed error distributions is to obtain the most appropriate error distribution that can describe each of the stated markets, when the true underlying error distribution is unknown. This study gives new insights into the literature by providing a more robust explanation in more depth to the modelling and analysis of volatility of the BRICS stock returns by using these specified collective seven error distributions which as far as we know has not been done before this study.

- The study of risk in the tails of stock returns is critical to financial markets research because of the fear and concern of extreme losses sometimes encountered during investment. The risks in the BRICS stock markets were modelled using the univariate versions of two EVT approaches via the conditional extreme value model and point process. The modelling outcome shows that the Russian market is the most risk-prone, followed by the South African market, then the Chinese, Brazilian and Indian markets, respectively.

- For ease of portfolio diversification and investment, this study has been able to observe risk reduction and high rate of returns approach through extremal dependence modelling. The study examines the five BRICS markets to determine the extent to which they are related and ascertain the asymptotic dependencies

in the market pairs. The findings will crucially help investors and portfolio managers to practice risk management through portfolio diversification and minimise their investment risks domestically and internationally. The bivariate point process and conditional multivariate extreme value (CMEV) models were used for the extremal dependence modelling in these markets, and the general observations revealed three outcomes: weak, fairly strong and low negative dependencies. All these outcomes signify varied levels of low extremal or asymptotic dependence.

To the best of the author's knowledge, these unprecedented findings were not revealed in the literature before this study. With low correlation or simply weak asymptotic dependence and negative extremal dependence being the fundamental requirements for efficient portfolio diversification, investors and all market participants can seize the rich investment opportunities presented by these BRICS markets. These outcomes are highly beneficial to investors, portfolio managers and other market participants who are interested in maximising their investment returns and financial gains, and in the process mitigate possible investment downturns through international portfolio diversification in the BRICS stock markets.

- The five given datasets were also used to forecast future extremal dependencies between each of the ten pairs of the BRICS markets. These predictions may possibly be applied by investors and portfolio managers as strategic guides for better upcoming investment attempts.

1.6 Thesis structure

The rest of the project is organised as follows: in Chapter 2, relevant literatures are reviewed while Chapter 3 centers on the methodology of the various models used. The next three chapters focus on data modelling, estimations and analytical findings, where Chapter 4 provides the detailed volatility modelling, Chapter 5 reports the risk modelling and Chapter 6 covers comprehensive details of extremal dependence modelling of the markets. Chapter 7 states the study's summary, concluding remarks, and direction for future research work. The details are as follows:

- In Chapter 2, a detailed literature review on relevant aspects of the study is carried out. These aspects include surveys on financial market volatility, risk management, the frequentist approach, reviews of studies on BRICS markets, stock market linkages and extremal dependence.

- In Chapter 3, a comprehensive description of the models and statistical methods used for modelling the volatility, risks, and tail dependencies of the five BRICS stock markets' data are discussed. These encompass, amongst others, discussions on models for modelling volatility, methods for testing for stationarity, serial correlation and ARCH effects, model selection approaches, and the seven conditional error distributions. The chapter also includes detailed discussions on the applied univariate and multivariate EVT models, parameters estimation approaches and model diagnostics.

- In Chapter 4, modelling and statistical analysis of the BRICS stock markets' volatility is executed using GARCH (1,1) model under the assumptions of seven

error distributions that include the normal, skewed-normal, Student's t, skewed-Student's t, GED, skewed-GED and the generalised hyperbolic (GHYP) distribution. The chapter also provides analytical details on the best error distribution that describes each of the markets. Rugarch package (Ghalanos, 2015) in R statistical software was used for the volatility modelling. This work has been published in the Journal of Statistics Optimization and Information Computing Vol. 3, pp. 749-772, 2020.

- In Chapter 5, we modelled equity risks of the BRICS stock markets using the univariate versions of extreme value theory models via the conditional univariate extreme value's GPD and point process. R packages like the evd (Stephenson, 2015), eva (Bader and Yan, 2016), texmex (Southworth, Heffernan and Metcalfe, 2016), ismev (Heffernan et al., 2016), evmix (Scarrott et al., 2019), and extRemes (Gilleland, 2016) were used for the risk modelling. This work has been submitted to the Journal of Statistics Applications and Probability, and it has been accepted, awaiting publication.

- In Chapter 6, the extremal dependence structures of each of the ten pairs of the BRICS stock markets were modelled using multivariate extreme value theory via the CMEV and bivariate point process models. The fitting efficacies of these two models were compared to ascertain the one that gave a better fit. The chapter also includes the predictions of future extremal dependence of each pair of the markets. R packages involving the texmex (Southworth, Heffernan and Metcalfe, 2016), extRemes (Gilleland, 2016), ggplot2 (Wickham, 2016), GGally (Schloerke et al., 2018), gridExtra (Auguie, 2015) and evd (Stephenson, 2015) were used to carry out the fit. This work will be submitted to a journal for peer review.

- In Chapter 7, the summary of research findings, contributions, concluding remarks, and suggested areas for further study are presented.

Chapter 2

Literature Review

2.1 Introduction

This chapter summarises the findings and conclusions of some of the monographs and research works in the literature on modelling of financial market's volatility, risk management and extremal (tail) dependence. Adequate knowledge of the volatility and risk in a potential market and its interdependence with another market can provide relevant information for investors as a guiding tool during investment and portfolio diversification. Several methodologies and models on extreme events studies have been used in the literature to model risk and tail dependence but the EVT approach has been known to outperform the other models in terms of performance (Dacorogna et al., 1995).

2.2 Financial market volatility

Financial market's volatility is a reflection of market's uncertainty and it is generally proxied by the variances. The theoretical and empirical perceptions of global financial crisis and its effect on international markets have over the years attracted enormous attention from investors, academics and policymakers. It is mostly conceived (in

the literature by some researchers) that the volatility of equity price is exclusively caused by random arrival of new information that relates to stock's expected returns, while other researchers largely view trading as the cause of volatility (Chandra and Thenmozhi, 2014). Volatility has been described as one of the most significant features of financial returns where it is regarded as a general measure of the uncertainty in financial assets and can be used to characterise market-price's variability. The term is also referred to as the way market activity emotionally responds, the variability of financial assets' returns and the risk of assets' portfolio (Tinyakova, 2012).

Volatility modeling in financial markets is very vital for risk management and options trading. Some common attributes of financial returns' volatility have been described in the literature under certain well-defined structures known as "stylised facts" of volatility. A few of these stylised facts indicate that, first, volatility occurs in clusters in a process known as volatility clustering. This process occurs when the volatility in the series is low in certain periods of time and high at other time periods. Second, volatility is described as having an asymmetric reaction to negative (or bad) news and positive (or good) news, in a process called leverage effects. Third, volatility jumps are infrequent and rare, that is, volatility (normally) develops over time in a continuous way (Qiao and Wong, 2010).

Moreover, the causes of volatility by a host of factors including the financial performance of organizations, investor's behavior, political events that are ordinary and predictable like democratic elections, and information from news are provided in the literature. Andersen et al. (2006), Andersen and Bollerslev (1998), and Ross (1989) are financial economics' monographs containing empirical evidence on information flow that are traceable to macroeconomic news, with other public information that have a direct effect on the volatility of stock return (Chandra and Thenmozhi, 2014). Stock market volatility can also be traced to the advancement of behavioral finance and economics, where it is assumed that investors are not impeccably rational and

that stock price movements can be affected by such sentiment-based and irrational decisions of the investors (Daniel et al. 2002).

Nelson (1996) indicated that volatility is mainly generated by financial crises. After the author's analysis on several factors which may cause an increase in the volatility of equity market, he found that financial crises and recession are the most significant factors (Rejeb and Boughrara, 2014). According to Ben-Rejeb (2013); Bekaert, Harvey, and Ng (2005); Ang and Bekaert (2004); Hartmann, Straetsman, and De Vries (2001), a financial crisis is classically related to a volatility increase for majority of assets.

With track of time and globalization, emerging financial markets get bigger and are becoming more developed, hence investors are showing more interest and better attractions to them (Galloppo et al., 2015). Low volatility of financial returns is an indication of no dramatic and rapid fluctuations in the market. Volatility summarily helps financial market participants to understand (market) risk and describe the size of changes on the market within a certain time period (Galloppo et al., 2015)

Numerous researchers with focus on the financial sectors have used the ARCH model, GARCH model and several other variants of the GARCH type model to capture the high volatility in the log returns of financial time series data. A few of these variants include the exponential GARCH (EGARCH) model developed by Nelson (1991), the asymmetric Threshold GARCH (TGARCH) model independently developed by Glosten, Jaganathan, Runkle (1993) and Zakoian (1994), and the asymmetric power ARCH model of Ding et al. (1993). Other variants include the Component GARCH (CGARCH) model by Lee and Engle (1999), and the family GARCH model of Hentschel (1995), among others.

Makhwiting et al. (2012) applied ARMA-GARCH type models to model the volatility and financial market risk of the Johannesburg Stock Exchange's (JSE)

shares. From the different candidates joint ARMA-GARCH model used, the authors' findings from the study revealed that the most accurate volatility forecast is achieved by ARMA (0, 1)-GARCH (1, 1) model. The Autoregressive models with exogeneous variables and power transformed and threshold GARCH errors (ARX-PPTGARCH) was developed by Xia et al. (2015), with parameters estimated using the Bayesian approach. The performance of the model was tested comparatively with another model called the ARX-GARCH model. The findings from the comparison showed that important financial characteristics like heteroscedasticity and asymmetry (amongst others) were well captured by the ARX-PPTGARCH model (Sigauke, 2016).

Volatility in the JSE index was investigated by Mzamane (2013) with the use of the univariate and multivariate GARCH models. The author's findings empirically showed that leverage effect is present in the log returns of the JSE market's index. On the contrary, the findings on the study of the volatility of the JSE market's stock return by Niyitegeka and Tewari (2013) using GARCH type models showed that the leverage effects do not exist, and the volatility of the stock returns is persistent. The U.S. equity returns volatility has been highly examined, most especially by French et al. (1987). The volatility of most industrialised nations like Italy, the Netherlands, France, Germany, and London International Stock Exchange has also been modelled using GARCH models by Taylor and Poon (1992), and Koutmos (1998) amongst others. Majority of the studies found that risk premium is significantly associated with the conditional variance of excess returns.

In a research work on three African emerging equity markets in the sub-Saharan, Ogum (2002) applied a time-varying asymmetric-MA-TGARCH model and daily equity indices for the South African, Kenyan and Nigerian stock markets from 1985 to 1998. The study was able to establish that conditional mean and variance display asymmetric response to past shocks. GARCH-type models were applied by Floros

(2008) to describe financial market risk of the Israeli TASE-100 index and Egyptian CMA General Index from 1997 to 2007. With the use of daily return series of these two stock markets, the study displayed the ability of GARCH models in explaining stylised facts of stock returns like volatility clustering and leverage effects. The study further utilised GARCH-M model and found insignificantly positive association between increased expected return and increased expected risk with the conclusion that greater expected risk does not necessitate greater expected return.

Brooks (2007) modelled volatilities in emerging stock markets of 26 economies in the European regions, Middle-East, Latin America, Africa and Asia for the periods of 1995 to 2005 using a PGARCH model. The author observed standard leverage effects evidence in many of the markets, except Saudi Arabia, Bahrain and Chile, where the leverage effects are not present. The study further detected anti-leverage effects evidence in the markets of Zimbabwe, Egypt and Nigeria such that higher volatility is associated with good news or positive sign. Pescetto and Appiah-Kusi (1998) revealed, by using EGARCH model, that majority of the African equity markets are described by changing levels of volatility, with some periods displaying large volatilities. Their findings suggest the existence of asymmetries in the conditional volatility, but in opposite of the standard leverage effects, where good news influences the markets' volatility more than bad news of the same magnitude.

The conditional volatility of six emerging markets (of South Korea, India, Argentina, Pakistan, the Philippines, and Taiwan) was modelled by Kassimatis (2002) using an EGARCH specification. The author's findings showed that decreases in volatility are witnessed after most important events of financial liberalization. Prior to the findings of Kassimatis (2002), various ARCH/GARCH specifications were used by Kim and Singal (2000) to determine if liberalization is intensified by volatility in emerging markets or not. The authors found substantial decline of conditional volatility in the periods after liberalization.

2.3 The frequentist approach

It is observed in the literature that a major rivalry in applications and interpretation is usually experienced with the use of the frequentist and Bayesian approaches. Even though the core of the debate between these two statistical divisions centers on the definition of probability and differences in how prior information is used, the ramifications include the analysis, breadth of research design, and interpretation (Skrepnek, 2007).

The 19th century was dominated by the Bayesian statistics, while the 20th century has been described by the classical or frequentist (inference) approaches (e.g. maximum likelihood, random sampling, and probability values) originating from Fisher's work during the 1920s and Neyman and Pearson in the 1930s (such as the hypothesis testing with probability values) (Edwards, 1974; Skrepnek, 2007).

The focus of the deliberation in the two inferential methods is the approach taken concerning how the "current state of knowledge" (known as the "prior information") is used within analyses and of how probability is basically defined. The Bayesian context integrates the prior information "formally" into analyses and asserts that probability only involves (or simply relates to) a degree of belief that is reliant on the present (or current) state of knowledge. On the other hand, the frequentists framework most frequently only considers the current (or present) state of knowledge "informally" in analyses and states that probability explains the limit of a relative frequency for the observation of an event over time where an underlying and true state is given (Skrepnek, 2007).

Even though the two divisions of statistical thought make effort to reach the same goals with the use of different philosophies, within the scientific inquiry's scope, each may also be considered an appropriate and relevant method given a stated set of conditions. However, differences between the methods ultimately impact how

empirical research is designed, analysed and interpreted (Kennedy, 2003; Skrepnek, 2007).

Daziano and Bolduc (2012) focused on the estimation, covariance structure, specification, point-estimate analysis and identification of a logit model with endogenous latent features which avoids inconsistency problem. The authors used both the frequentist (via maximum simulated likelihood) and Bayesian approaches, and observed that despite the different philosophies underpinning the methods, the estimators provided are asymptotically equivalent. With a large sample size, both estimators summarily gave comparable and feasible results (Daziano and Bolduc, 2012).

Robert (2013) states that the methodological and philosophical difficulty with the frequentist evidence principle (FEV) is a strong restrictive reliance on the belief that one given model can be true or exact, while the Bayesian approach is criticised for its subjective prior modelling. In a monograph, Cox and Mayo (2010) stressed that because frequentist approach is based on the sampling distribution of the test, it is (more) objective (Robert, 2013).

The central views from which the controversy and differentiation exist between frequentist and Bayesian inference comprise the use of prior (information or) knowledge and the way it associates with beliefs that define probability. Although there are numerous definitions in place, one broadly used definition of probability in an orthodox sense is (the relative frequency of an event that occurs in an infinitively long series of replications or) as a limit of relative frequency in a large number of trials. In addition to this, another view follows a frequency perception of probability that includes the proportion of time that events of similar type will happen when an experiment is repeated several times (Kennedy, 2003; Grinstead, 1997; Freund, 1971). Following these definitions, it has been argued that the frequentist approach is being "objective" for the fact that results of a study does not give room for extraneous

contribution from other sources and are restricted to the immediate experiment being described or reported. This argument in a way makes the frequentist approach appealing among regulatory agencies and is one good reason that has made it popular with research scientists (Speigelhalter et al., 2004).

Contrary to the frequentist viewpoint, the Bayesian standpoint defines probabilities in the framework of a subjective degree of belief, which infers that an event's probability in an immediate experiment may include elements that can be taken from other sources and cannot necessarily be dismissed from an inferential framework (Freund, 1971; Kennedy, 2003; Skrepnek, 2007). However, we may indeed question the robustness of the analysis of a Bayesian if prior information is gotten from weak sources, which can potentially introduce bias and hereafter detracting from objectivity that is required within scientific inquiry (Skrepnek, 2007).

The robust Bayesian statistics has witnessed a significant success in the literature however, but its limitations regarding applied data analysis have become obvious (Bickel, 2015). If the set of credible priors is sufficiently small, a posterior's sensitivity to the prior in the set (Liu and Aitkin, 2008; Lavine, 1991; Berger, 1990) can show how reliable the results are with a negligible sensitivity, however, when the sensitivity is higher, no recourse is provided for the data interpretation (Berger, 1984). A worse scenario emerges where the set of credible priors may be very large and lead to abandoning the Bayesian method in favor of a frequentist approach without the guidance from concept (theory) that could make clear the choice between methods (Bickel, 2015). However, one relevant advantage of the Bayesian approach to that of the maximum likelihood estimation (MLE) approach is that the former is not dependent on regularity conditions required for extreme value distribution' maximum likelihood estimation of parameters (Coles, 2001; Beirlant et al., 2004; Sigauke, 2014)

It is well known, in spatial econometrics, that some frequentist approaches, like the generalised method of moments and instrumental variables estimation, may create

estimates that are outside the region of stationarity, while other approaches, such as the (maximum likelihood (ML), are not swayed by this downside (Elhorst, 2014; Hassan, 2017). Hence, in this study, the parameters estimation of the volatility, risks and extremal dependence modelling of the BRICS stock markets will be obtained using the frequentists framework via the MLE.

2.4 Risk management

Because of the rising need for financial risk management and the series of global (financial) crisis, forecasting of risk has become a critical issue for stakeholders in basically all financial markets across the globe. The traditional approach used for modelling risk was through the use of the value-at-risk (VaR) and expected shortfall (ES) methods, whose concepts are based on the normal distribution assumption. Furthermore, this traditional VaR's estimates are based upon the concept of homoscedasticity, which assumes that the standard deviation of returns does not change over time. Because of this, Engle (2001) claims that a much better estimates can be obtained from models that are explicitly based on the concept of heteroscedasticity, which allows the standard deviation to change over time (Al-Janabi, 2012).

A VaR model measures how much a portfolio can lose within a given period of time, with a probability prespecified (Hong et al., 2009). That is, it measures the risk in a market by evaluating how much a portfolio's value could decline with probability of $b\%$ above a certain time horizon u because of changes in market rates or prices, while the ES models the expected size of a loss that can exceed the VaR. Hence, the VaR estimation is simply interested in knowing "how bad can things become?", while the ES estimate focuses on measuring (or assessing) the expected loss if things actually go bad (Karmakar and Paul, 2015).

As opposed to the "normal" concept, financial returns are known to exhibit fat

tails and normal distribution cannot effectively model such return's behaviour. Fat tails in the financial returns indicates that extreme outcomes take place much more frequently than what the normal distribution assumption can predict (Karmakar and Paul, 2015; Sigauke et al., 2010). The deficiency of the VaR and ES models (due to their normal assumption) can be modified using an EVT based method of VaR approach (Bali, 2003). As stated in Section (1.1.5), modelling of risk using model frameworks with a normal distribution assumption are known to understate the result originating from data like the financial returns with fat tails (Al-Janabi, 2012; Taleb, 2007; Mandelbrot, 1997; Mandelbrot, 1963).

It has been argued by several authors like de Jesus, Ortiz, and Cabello (2013); Santos et al. (2013); McAleer et al (2013); Allen et al. (2013); Santos and Alves (2012) that extreme events contained in the tail distribution of losses can be taken into account explicitly by EVT (Herrera and Schipp, 2014). As an advantage over classical approaches, the EVT is a parametric method which allows for extrapolation of the tail behaviour to extreme levels. Furthermore, the EVT method does not use the whole dataset but limits it modelling approach to the tail behaviour of a loss distribution using only extreme values (Herrera and Schipp, 2014).

Karmakar and Paul (2015) used the two-stage GARCH-EVT approach of McNeil and Frey (2000) to model the tails of (return) distributions and compute intraday VaR and ES measures for 16 stock markets across North America, Africa, Asia, Europe, Latin America and Australia. The authors (Karmakar and Paul, 2015) compared the efficacy of the conditional EVT method with other competing models and observed that EVT outperformed the others. McNeil and Frey used the GARCH model to filter the return series to get a nearly i.i.d. (independent and identically distributed) residuals (in the first stage), and they fitted EVT model on the standardized residuals in the second stage (Karmakar and Paul, 2015).

Several other authors have used GARCH-EVT joint approach with other models

on various data sets and have observed that the GARCH-EVT method performed better than the competing models (for VaR estimation) when compared. GARCH-EVT model was used by Marimoutou et al. (2009), Ghorbel and Trabelsi (2008), Cotter (2007) etc. to measure VaR in various markets and they observed that the EVT did better than other well-known modeling techniques in forecasting of VaR's estimates (Karmakar and Paul, 2015). Bali and Neftci (2003) applied the duo of a Student's t distributed GARCH model and the GARCH-EVT model to U.S. short-term interest rates and the authors discovered that more accurate estimates of VaR were yielded by the GARCH-EVT model than did the Student's t distributed GARCH model. Karmakar and Shukla (2015) also compared the accurateness of GARCH-EVT method for calculating VaR with other rival models using data from 6 global emerging and developed equity markets. Their findings showed that GARCH-EVT method performed best in the in the estimation of VaR.

In another study, the point process approach of extreme value models was used by Smith (2003) to model the daily returns for General Electric, Pfizer and Citibank from 1982 to 2001. Volatility was first filtered from the data using GARCH (1, 1) model, after which the returns were modelled to analyse the VaR (value at risk). The researcher resolved that in order to obtain a satisfactory extreme value modelling outcome, it is better to remove heteroscedasticity and short-range serial correlation from the return before applying EVT models.

Many EVT and GARCH based models were estimated by Ergun and Jun (2010) to predict intraday VaR for the returns of S & P 500 equity index futures. The researchers' findings showed that the EVT-GARCH based models which take into account conditional kurtosis and skewness provide accurate forecast of VaR. Chavez-Demoulin and McGill (2012) used EVT-Hawkes process to measure high frequency intraday VaR. The authors observed that a suitable estimate of high quantile risk measures for the U.S. market's financial time series was provided by the process.

This study applies the GARCH-EVT approach to model volatility and risk in the BRICS stock market. The GARCH model will be used in modelling and filtering out heteroscedasticity in the returns, while the univariate EVT via the threshold excess model and the point process will be used to model the risk in each of the markets. In statistics and econometrics, probabilities of left tail are closely associated with the likelihoods of extreme downward movements in the market (see Embrechts et al., 1997). Even though both the "gains (i.e., positive or right tails)" and "losses (i.e., left or negative tails)" of equity return distributions are of invaluable modelling interest from a risk management view, much more studies on extreme equity returns have focused on losses when compared to gains, and large booms are usually considered less important than large crashes (Karmakar and Paul, 2015). This study focuses on modelling the losses or negative tails of the distribution of the BRICS equity markets. This is because the left tail is of more interest from a risk management viewpoint (Ardia et al., 2011), and a more practical risk measure should be related to large adverse movements, or large losses in the market (Hong et al., 2009).

2.5 Market linkages and extremal dependence

a) Stock markets linkages:

Several articles in the literature have shown that linkages between international equity markets can be described either by fundamentals or in relations to contagion hypothesis. The first hypothesis highlights the role of common fundamental factors (Adler and Dumas, 1983; Stulz, 1981; Solnik, 1974; Jawadi et al., 2015). Fundamentals hypothesis postulates that shocks are spread through stable linkages, and transmission of volatility is the same in periods of calm and crisis. The effects of contagion result when enthusiasm for stocks in one equity-market leads to enthusiasm for stocks in other equity-markets, irrespective of the evolution of market fundamentals (Jawadi

et al., 2015).

The transmission of volatility is high during market contagions as investors make efforts to discover potential changes in price using fluctuations observed in other markets. In such scenario, a shock stemming from one market may present a disrupting impact on other markets, which in turn disrupts other different markets. Occasionally, the domino effect (where the shock in one market triggers/sets-off shock in another market(s)) ensues regardless of the development of market fundamentals (Maghyereh and Awartani, 2012; Longstaff, 2010; Bekaert et al., 2005; Forbes and Rigobon, 2002). Contagion is the increase in the probability of crisis beyond what could be expected by the linkages between fundamentals (Fazio, 2007).

Chan-Lau et al., 2004 described contagion as the probability of seeing large return observations concurrently across different financial markets or the co-exceedances (of extreme observations) instead of as an increase in correlations. Several monographs in the literature have documented that financial markets' prices co-movement increases significantly during stress period, like the Mexican crisis in 1994, the 1992-93 Exchange Rate Mechanism crisis, the 1997 financial crises in East Asia, and in 1998 in Brazil and Russia. Chan-Lau et al. (2004) classified extreme (tail) events as those returns that exceed a large threshold value (using a 95th quantile as the cut-off point), and they used EVT methods to measure contagion as the co-exceedances or joint behavior of financial returns' extremal observations across different markets. EVT measures of contagion is referred to as extreme contagion measures by Chan-Lau et al., 2004.

The EVT approach for modelling contagion captures well the belief that small shocks are differently transmitted across financial markets than large shocks. The application of global extreme contagion measures to the analysis of extreme positive and negative returns can be referred to as bull and bear markets contagion respectively (Chan-Lau et al., 2004).

b) Extremal dependence:

Extremal dependence concept can be used to explain the co-movement between markets at the tail (extreme) region of the price or return distributions.

Chan-Lau et al. (2004) presented measures of global extreme contagion created from bivariate extremal dependence measures to quantify both positive and negative stock returns contagion at the intra- and inter-regional levels for several emerging and mature stock markets during the past decade. The emerging stock markets are Thailand, Taiwan Province of China, Singapore, the Republic of South Korea, Philippines, Indonesia, Hong Kong SAR, Malaysia, Mexico in Latin America, Brazil, Chile, and Argentina. The mature stock markets include those of the United States, France, United Kingdom, Japan and Germany. The authors observed that the measures of contagion using extremal dependence and correlation approaches are not highly correlated, except for the Latin America stock markets. Their findings suggest that results may be misleading when correlations are used as a proxy for contagion.

Extremal dependence study was carried out by Fernandez, (2006) to determine whether very large depreciation or appreciation in a six-nation's nominal exchange rate might spill-over to the euro, and vice versa. The study period was 1999-2002 and the examined countries were Japan, U.S, Poland, Czech Republic, UK and Switzerland. Their findings showed no asymptotic dependence between paired exchange rates.

Using 1995-2016 stock data, risk spillovers were analysed (by Warshaw, 2019) across the stock markets of North American (i.e. the U.S., Canada and Mexico), where upside (downside) risk denotes potential extreme short (long) position losses. The dependence structure of each pair of the markets were modelled using the generalised autoregressive score (GAS) copulas, after which an upside and a downside Conditional Value-at-Risk (CoVaR) were estimated. In contrast with the notion that International stock markets are usually considered more likely to jointly crash than

boom, the author observed a symmetric conditional tail dependence for each of the paired stock markets. The author further discovered a larger co-movement under extreme economic conditions in all the three market pairs due to a significantly higher symmetric conditional tail dependence following the Global Financial Crisis (GFC). In addition, the study revealed significantly larger upside and downside risk spillovers for U.S.-Mexico and Canada-Mexico pairs following the GFC as compared to the period before the crisis.

Dynamic correlations were also estimated, by Lahrech and Sylwester (2013), for the Canadian, U.S. and Mexican stock market returns, using a trivariate extension of the ADCC-GARCH model (Cappiello, Engle, and Sheppard, 2006; Warshaw, 2019). The study revealed a relative stability in the US-Canadian markets dynamic correlation, but the Canadian-Mexican and U.S.-Mexican pair-wise time varying correlations exhibited upward trends in the post-NAFTA period (Warshaw, 2019).

Using Markov Switching copulas for the periods January 1915 to February 2017, Ji et al. (2020) analysed the risk spillovers of equity market from the U.S. to the other G7 nations. The authors observed that risk spillovers are asymmetric and significant, and that spillover that originated from the U.S. is stronger than the ones from the rest of the G7 nations. The study further showed that upside risk spillover is weaker than downside risk spillover (Warshaw, 2019).

The theory of co-integration analysis was used by Alagidede (2008) with the aim of analysing the market linkages among South Africa, Egypt, Nigeria and Kenya. The findings showed that these markets do not considerably move together, despite the economic reforms existing among them. These four African markets (South Africa, Egypt, Nigeria and Kenya) were also modelled by Samuel et al. (2018) with daily returns data for the period of September 2000 to August 2015 using the bivariate-threshold-excess model and point process approach. The researchers observed that the markets displayed asymptotic independence or (very) weak asymptotic dependence

and negative dependence.

With the use of EVT models via the univariate and multivariate versions of Block maxima, point process and threshold excess models, Lipika (2018) conducted a study on the climate data of Western Cape province of South Africa. The researcher used data from 1965 to 2015 and concluded that the block maxima method was outperformed by the duo of threshold excess and point process models. Furthermore, it was observed that the point process method outperformed the threshold excess model in the bivariate analysis due to more availability of data observations, but both models performed equally well in the univariate analysis.

In a research work by Joe et al. (1992), the non-parametric approach of polar coordinates and two parametric dependence models of point process were compared using 504 pairwise combinations of sulphate and nitrate concentrations in the same U.S. region. The logistic and bilogistic models were the parametric dependence models used and the study showed that the bilogistic model was better than the logistic model for the dependence modelling under both non-parametric and parametric approaches.

Boyer et al, 1997 and Embrechts et al., 1999 posited that using simple correlation analysis to study financial market dependence can be deceptive. While Boyer et al. (1997) showed numerically and analytically that stationary distributed data could produce spurious correlation breakdowns, Embrechts et al. (1999) believed that the use of correlations as measures of dependence is only justified for multivariate normal distributions (Chan-Lau et al., 2004). Under a non-normal (distribution) situation, correlations results fail to (accurately) show the multivariate dependence structure. This means that absolute unity correlation does not mean the variables are perfectly dependent, and correlation value of zero does not indicate independence (Chan-Lau et al., 2004).

Poon, Rockinger, and Tawn (2001) used simple methods developed by Coles, Heffernan, and Tawn (1999) to estimate the extremal dependence of two series and test

whether they are asymptotically independent. These methods were applied to daily equity market returns in Germany, France, Japan, the U.S. and the UK. The authors divided the sample into three different periods to enhance a more tractable analysis of extremal dependence in the bull and bear markets. Their findings showed that the bear markets experience much stronger extremal dependence than the bull markets, and that most of the pairwise combinations of equity returns are characterised by asymptotic independence.

Many authors have used the narratives of volatility spillover as a proxy for the concept of extremal (tail) dependence. A lot of the extant articles (e.g., Hong, 2001; Lin et al., 1994; King et al., 1994; King and Wadhwani, 1990; Hamao et al., 1990; Cheung and Ng, 1990, 1996; Engle and Susmel, 1993; Engle et al., 1990 and Granger et al., 1986) in the literature have focused on risk measurement using volatility with emphasis on the effects of volatility spillover (Hong et al., 2009). Volatility in itself is a significant instrument in macroeconomics and finance, but in practice, it can only effectively characterise small risks (Gourieroux and Jasiak, 2001; Hong et al., 2009). In situations where extreme market movements occasionally arise, volatility alone cannot adequately capture risk. Measures of volatility based on distributions of asset return cannot generate accurate market risks estimates during volatile periods (Longin, 2000; Bali, 2000). Furthermore, volatility consist of both losses and gains in a symmetric manner, while financial risk is clearly related to losses but not profits. Hence, a more sensible risk measure should relate to large adverse market movements, or large losses (Hong et al., 2009).

The pattern of the dynamics of tail movements across equity markets is vital for effective risk measurement and management. In an integrated market for instance, risk spills from market to market, and thus accounting for that transmission when risks are measured and managed in local markets is very importance. Contrarily, spillovers may be ignored safely when financial markets are completely segmented,

because risk cannot transmit across markets (Maghyereh and Awartani, 2012).

An essential part of information set required by policymakers and financial managers is an in-depth comprehension of the direction and magnitude of linkages and spillover effects. The knowledge of interdependence between markets is vital to financial managers to determine diversification and hedging of their international investment. Also, a policymaker sees financial instabilities, like equity market crashes and a bank collapse as major problems that directly impact a nation's welfare (Ghini and Saidi, 2017). During the period 1988–2000, Worthington and Higgs (2004) used the multivariate GARCH model to study the transmission of equity returns and volatility in Asian emerging markets and developed markets. Their findings revealed the magnitude and source of spillovers by indicating that the mean spillovers to the emerging markets from the developed markets are not homogenous across the emerging equity markets, and direct spillovers are normally higher than indirect spillovers, particularly for the emerging markets.

Qiao (2010) used the BEKK GARCH model, which is an MGARCH (multivariate GARCH) model formulated by BEKK (i.e. Baba et al., 1990) for the analysis of the linkages between the I.T. (information technology) equity market in the U.S. and those in Hong Kong, Sweden, Finland, Japan, Canada and France. The authors' findings showed that U.S. I.T. has a mean spillover effect to Hong Kong and Japan but not to Sweden, Finland, Canada and France I.T. markets, and that the I.T. markets of France, Sweden and Canada also create substantial mean spillover effects to the U.S. I.T. market. Ghini and Saidi (2017) used a bivariate VAR-BEKK GARCH model to examine the return and volatility linkages among the Moroccan (MASI) equity market and those of the U.S. (NASDAQ), Germany (DAX), France (CAC), and the UK (FTSE) equity markets prior to and during the financial crisis. The researchers used data from January 2002 to December 2012 and observed varying degrees of interdependence and spillover effects between the Moroccan emerging equity market

and the four selected major stock markets before and after the global financial crisis.

Jawadi et al. (2015) studied volatility spillover between the U.S. market and the trio of European equity markets: Frankfurt, London and Paris. The study was carried out around the Subprime crisis' period using the Threshold GARCH model, and the authors discovered evidence of weak volatility transmission between the two selected regions before the Subprime crisis. During the period after the crisis however, they observed returns and volatility spillover from European markets to the U.S. and vice versa at different times of the trading day. The findings indicate the dependence of the two regions became more pronounced during the recent Subprime crisis. Nam et al. (2008) investigated the spillover of price and volatility from the U.S. stock market to the markets of Taiwan, Malaysia, South Korea, Singapore and Hong Kong. The latter (nations) are five Pacific-Basin markets, and the authors' finding revealed that the impact of U.S. shocks on Asian-markets' volatility decreased significantly as a result of the 1997 financial crisis.

Kanas (1998) studied volatility spillover among Paris, London and Frankfurt exchanges at the period of the 1987 U.S. crash. For the period January 1984 to December 1993, the author discovered mutual (i.e. both directional) spillover between Paris and London, and between Frankfurt and Paris, and unidirectional spillover (emanating) from London to Frankfurt (Jawadi et al., 2015). The author further revealed that these three European stock markets became increasingly connected to the U.S. equity market during the period following the October 1987 crash. Following the October 1987 crash, Hamao et al. (1990) applied a univariate GARCH model to study the interdependence among the stock markets of Tokyo, New York and London. Having decomposed the daily close-to-close returns into their close-to-open and open-to-close components, the authors found the spillover of price volatility from New York to London, London to Tokyo and New York to Tokyo. However, spillover effects were not detected in the reversed directions.

During the period 1988–2000, Worthington and Higgs (2004) used the multivariate GARCH model to study the transmission of equity returns and volatility in Asian emerging markets and developed markets. Their findings revealed the magnitude and source of spillovers by indicating that the mean spillovers to the emerging markets from the developed markets are not homogenous across the emerging equity markets, and direct spillovers are normally higher than indirect spillovers, particularly for the emerging markets.

The effects of spillover between the U.S. and German stock markets were examined by Baur and Jung (2006), with particular attention around the opening time of the two stock markets. The authors' conclusion did not see any spillover evidence from previous afternoon U.S. returns to the trading of German's morning. Using the ARCH models' family, Susmel and Engle (1994) investigated hourly volatility spillover between London and New York, but the authors did not discover strong spillover of volatility between the two stock markets. Miralles-Marceloa et al. (2010) studied the linkages between the U.S. Dow-Jones and Spanish IBEX with data spanning 2nd January 2003 to 31st December 2004 using GARCH models. The authors found spillover evidence from previous daytime U.S. returns to Spanish overnight returns.

The knowledge of spillover effects on return and volatility may be helpful in portfolio management, specifically in market selection and strategic asset allocation (Maghyereh and Awartani, 2012). Qiao and Wong (2010) studied the effect of the GARCH model on U.S. equity markets for different periods. Using a bivariate GARCH method, the authors examined the effects of spillover effects of turnover and volume on the conditional volatility. Their findings did not see any effect of turnover or volume spillover on conditional volatility among the companies studied. The result indicated that the volatility of equity return of a company may not essentially be influenced by the turnover and volume of other companies.

Suliman (2011) examined the propagation of major financial crisis in the GCC

(Gulf Cooperation Council) from 1960 to 2002. The study showed that the Asian 1997 financial crisis and the U.S. 1987 equity market crash were propagated to the GCC nations. With the use of high frequency data, Jawadi et al. (2013) investigated the U.S. stock market openings' impact on linkages between the French and UK stock markets in a non-linear and linear framework. The authors estimated a threshold autoregressive model and a multivariate linear specification and found significant time varying and positive dependence of the UK and French equity markets' returns on the U.S. returns during overlapping hours.

A spectral measure method was proposed by Starica (1999) to estimate the joint probabilities of extreme returns in series created by models having constant conditional correlation models, a family that incorporates GARCH models. The study revealed a high level of dependence among extreme returns of high-frequency data on European Union currencies. Bae, Karolyi, and Stulz (2003) used a multi-nomial logistic regression to estimate contagion across countries within a region and across regions. The authors' findings revealed that contagion could be explained by changes in exchange rates, interest rates, and conditional stock return volatility. Longin and Solnik (2001) used monthly data spanning the period January 1959 to December 1996 and found that extreme returns' pairwise correlation between the U.S. and that of Japan, Germany, France, and the United Kingdom (the G-5) does not increase in bull markets but increases in bear markets.

Costinot, Roncalli, and Teiletche (2000) proposed that financial markets' dependence was better modeled using copulas instead of correlation approach. The authors observed that the probability of joint exceedance for the French CAC-40 and the U.S. Dow Jones equity market indices intensely increased using copulas instead of the bivariate normal distribution. They also found a stronger evidence of contagion with the use of copulas than those obtained using correlation analysis during the analysis of the dependence between stock market returns and exchange rates of the Asian

crisis in 1997.

Sohel-Azad (2009) used Gregory-Hansen (1996) cointegration test to investigate whether three East Asian equity markets of South Korea, Japan and China are jointly and/or individually efficient, and whether contagion occurs between the cointegrated markets. The author's findings revealed that the South Korean and Japanese equity markets have individual market efficiency, while the Chinese does not. Furthermore, the outcome on the test for cointegration indicated strong rejection of market efficiency for the three markets. With a simple case of contagion, the authors found that despite the presence of a long-term association among the three selected markets, the contagion hypothesis only holds between South Korean and Japanese equity markets, denoting the benefits of short-run portfolio diversification from these two markets.

Adequate understanding (and relevant information) of interdependencies among markets are pertinent in enabling bankers, investors, and policymakers to better model and predict the dynamics of equity price. Furthermore, such information helps to reduce systemic and financial risk, guides investors and risk managers in specifying their investment choices in terms of international diversification, trading strategies, and hedging (Jawadi et al., 2015). More germane information on the potential benefits derivable from international diversification of portfolio with regards to risk reduction can be seen in Solnik (1974).

An essential part of information set required by policymakers and financial managers is an in-depth comprehension of the direction and magnitude of linkages and spillover effects. The knowledge of interdependence between markets is vital to financial managers to determine diversification and hedging of their international investment (Ghini and Saidi, 2017).

2.6 Reviews of studies on BRICS

With the BRICS economies on a global spotlight, the dynamic analysis of their markets' volatilities, risks and tail dependence is paramount to international investors, policymakers and all market participants who are interested in portfolio diversifications in their stock markets. Mensi et al. (2016), examined the asymmetric linkages between the BRICS three country risk ratings (i.e. economic, financial and political risk) and their stock markets from January 1995 to August 2013 with the use of dynamic panel threshold models. Their findings indicated asymmetry in most of the analysis, however the signs and significance of the risk rating effects on the BRICS market returns vary across the upper and lower regimes.

Kang et al. (2016) carried out studies on the relationship between the BRICS commodity futures markets and their stock markets with reference to testing asymmetries, dynamic conditional correlations (DCCs) and long memory volatility properties in the markets. The trivariate fractionally integrated asymmetric power autoregressive conditional heteroscedasticity (DCC-FIAPARCH) model was used and the studies spanned from 1997 to 2013. Their findings identified the long memory properties of volatility in the pairs of the BRICS commodity and stock markets. In addition to this, asymmetries and DCCs were also observed in the pairs.

Ijumba (2013) investigated the levels of interdependence and dynamic linkage among the BRICS countries. The study employed a Vector Autoregressive (VAR), univariate GARCH (1,1) and multivariate GARCH models. The results from VAR showed that there is unidirectional linear dependence of Indian and Chinese stock markets on Brazilian market. On the other hand, the univariate GARCH (1,1) model revealed the presence of volatility persistence in all the BRICS markets' stock returns. China was found to be the most volatile, followed by Russia, and South Africa was the least volatile. Multivariate GARCH also showed that there is volatility persistence

among BRICS stock markets.

Zhang et al. (2013) used the causality-in-variance test developed by Herwartz and Hafner (2006) to study volatility transmissions (spill-overs) between the bond and equity markets of the BRICS and G7 nations from December 1988 to December 2012. The findings showed, at 1% significance level, a unidirectional spill-overs from the bond to the stock markets in the UK, Germany and U.S. For South Africa, France and Brazil, the findings observed a bidirectional volatility spill-overs between the bond and equity markets, but no definite directional conclusion was made for Canada, Italy, Japan, China and India.

The industries co-movements and pact of the member states in the BRICS markets were investigated by Lee et al. (2017) using the (BRICS) industry weekly data from 1997 to 2013. The researchers witnessed a large increase in the co-movements of the BRICS markets' industries effective from 2003, and this was possibly due to Goldman Sachs report on the BRICS economies' rapid development. The study used GJR-GARCH and EGARCH models to determine asymmetries in the conditional correlations of the BRICS markets' returns. The outcome of their work indicated signs of asymmetries on threshold and leverage effects with strong reaction to good news. It was further observed that among all the sampled industries, the BRICS financial industries had the highest co-movements.

Romero and Kasibhatla (2013) studied the volatility dynamics and out-of-sample forecasts of each of the four BRIC's (markets') returns using GARCH models. The study was conducted under the assumptions of three distributions of a normal, student's t, and GED from January 2000 to December 2010. Their findings showed statistically significant indication of asymmetric impact of bad and good news in each of the markets. Furthermore, the student's t error distribution was the best to describe the Brazilian, Russian and Indian markets, while the Chinese market was best characterised by the GED.

Afuecheta et al (2020) contributed on behaviour of rare events of the BRICS stock markets. Their focus was on the extreme behaviour of the five countries' stock markets from 1995 to 2015. The authors used five distributions: the generalised extreme value distribution (GEV), the generalised logistic distribution (GL), the generalised Pareto distribution (GP), the Student's t-one parameter exponential distribution (STE) and the Student's t-two parameter Weibull distribution (STW). The overall fit of these distributions was compared using different criteria: log-likelihood, the Akaike information criterion (AIC), the Bayesian information criterion (BIC), the consistent Akaike information criterion (CAIC), the corrected Akaike information criterion (AICc), the Hannan–Quinn criterion (HQC). The outcome indicated that the GEV distribution gave the best fit. The estimates of value at risk, $\mathrm{VaR}_p(\mathrm{X})$ and expected shortfall, $\mathrm{ES}_p(\mathrm{X})$ from the BRICS stock markets were computed, and it was found that Russia and Brazil have the largest risks. The authors also modelled the tail dependence of the BRICS economies by using various copula models, namely: Galambos, H*üsler*-Reiss, Gumbel, normal and Student's t. models. The Gumbel copula was found to be the best model with the best fit.

Jegadeshwaran and Sangeetha (2018) explored the relationship among the BRICS nations' stock markets. The main purpose was to determine how the stock markets of these emerging countries are related to each other. The study employed daily closing prices of BRICS stock markets indices namely IBOVESPA, RTSI, SENEX, SSE Composite Index and JSE ALSI. The results showed that all the variables are positively correlated, RTSI and JSE ALSI were found to have a high positive relation and IBOVESPA and SSE CI have the lowest correlation coefficient during the period under review. In addition, the study employed Granger Causality test in order to identify the presence of the significant variables among the BRICS stock markets indices. The results showed unidirectional relationships between IBOVEPSA and all other stock markets and it was found that there is no relationship between JSE ALSI

and RTSI.

Babu et al. (2015) investigated the co-movement of BRICS nations' capital markets. This was done by investigating the short–run and long-run integrations and linkages of BRICS countries' stock markets indices, namely, BSE Senex, FTSE/JSE Top 40 Index, IBOVESPA, RTS Index and SSE Composite, during the study period April 2004 - March 2014. The study employed GARCH (1, 1) model, Johannsen Co-integration test, Vector Error Correction model, and Granger Causality test in order to study the stock markets linkage. The results that were found from Johannsen Co-integration test revealed that all the samples indices of BRICS stock markets were co-integrated with each other. The study concluded that BRICS indices were engaged in a long time relationships and only RTS Index recorded both short-run and long-run relationships with other BRICS sample indices. The conclusion further reveals that global investors could use the opportunity for portfolio diversification, both under short-run and long-run periods in BRICS stock markets.

For more comprehensive understanding of the use of the univariate and multivariate approaches of extreme value theory in risk management see Longin and Solnik (2001); Poon et al. (2003).

2.7 Contributions to the literature

It can be seen from the stated literatures that many authors, like Taylor and Poon (1992), Koutmos (1998), Hamao et al. (1990), Miralles-Marceloa et al. (2010), Jawadi et al. (2013), Chan-Lau et al. (2004), Warshaw (2019), Karmakar and Paul (2015), Bali and Neftci (2003) among others, have done several research-works on modelling the financial return volatility, risks and extremal dependence of mature stock markets like those of the United States, U.K, Germany, Japan, France, Switzerland, Spain, Australia etc. Also, some authors like Chan-Lau et al. (2004), Karmakar and Paul

(2015), Samuel et al. (2018), Kim and Singal (2000), Kassimatis (2002) among others have modelled the volatility and risks in the returns of certain emerging stock markets like those of Thailand, Taiwan Province of China, Singapore, the Republic of South Korea, Philippines, Indonesia, South Africa, Hong Kong SAR, Malaysia, Mexico, Brazil, Chile, Argentina etc., including the extremal dependence between some of these emerging markets and mature markets.

As for the collective markets of the BRICS economies, quite a few authors, like Ijumba (2013), and Romero and Kasibhatla (2013) as earlier described, have modelled their markets' return volatility. However, before this study, none of these authors has modelled the volatility with as many as seven error distribution assumptions under the GARCH model to obtain the most appropriate distribution for describing each market when the underlying distribution is unknown. In other words, in Chapter 4, the return volatility of the BRICS stock markets is modelled under the assumption of seven error distributions that include the normal, skewed-normal, Student's t, skewed-Student's t, GED, skewed-GED and the generalised hyperbolic (GHYP) distribution. In particular, this study gives new insights into the literature by providing a more robust explanation in more depth to the modelling and analysis of volatility of the BRICS stock returns by using the specified seven error distributions, which as far as the author knows, has not been done before this study. Moreover, as opposed to that major study carried out by Ijumba (2013) for the entire five BRICS markets using weekly data that covered only 2 years of markets' activities from 2010 to 2012, this study uses a wider coverage of activities from 2010 to 2018 using daily closing data, which can potentially give more accurate results.

Also, before this study, as far as the author knows, no work has been done on modelling the BRICS stock markets risks using the univariate EVT models of the conditional extreme value (CEV) and point process before this study. In other words, this study has been able to model the BRICS stock return risks using the univariate

versions of the CEV and point process models as detailed in Chapter 5. Furthermore, quite a few authors like Mensi et al. (2016), Kang et al. (2016), Ijumba (2013), Lee et al. (2017), Zhang et al. (2013), Romero and Kasibhatla (2013), Afuecheta et al (2020) and Babu et al. (2015) have modelled the extremal dependence of the collective BRICS stock markets. However, none of the authors has used the combined multivariate versions of the point process models through the logistic, negative logistic, Husler-Reiss, Bilogistic, negative bilogistic and Coles-Tawn (or Dirichlet) models, and the CMEV model before this study to the best of the author's knowledge. Hence, this study robustly models and estimates the asymptotic (extremal) dependencies in the ten pairs of the BRICS stock markets as detailed in Chapter 6.

2.8 Conclusion

In this study, we used data from the five BRICS stock markets with emphasis on modelling the volatility and risk in each of the markets using a univariate analysis, and comparing them in descending order of hierarchy. Also, we model the tail or extremal dependence in the pairwise combinations of the markets using a multivariate approach.

Although evidence from the literature suggest that some studies have been done on modelling volatility of the BRIC(S) markets (see Zhang et al., 2013; Romero and Kasibhatla, 2013; Kang et al., 2016; Lee et al., 2017) and a few studies on modelling their interdependence or co-movement (see Ijumba, 2013; Babu et al., 2015; Jegadeshwaran and Sangeetha, 2018), no evidence is currently available on modelling their extremal dependence using the conditional extreme value (CEV) model and point process approach. This is the gap this study intends to bridge by modelling the risk in each market and the asymptotic dependence of the paired markets using the CEV model and point process approach.

Chapter 3

Methodology

3.1 Introduction

This chapter gives an in-depth description of the models and methods used in the analysis of the volatility, risks and tail dependence of the five BRICS markets data. The modelling is implemented using various EVT *R*-packages like texmex (Southworth, Heffernan and Metcalfe, 2016), ismev (Heffernan et al., 2016), evmix (Scarrott et al., 2019), evd (Stephenson, 2015), eva (Bader and Yan, 2016), POT (Mathieu and Dutang, 2016) and extRemes (Gilleland, 2016) in *R* statistical software obtainable via CRAN repository (https://cran.r-project.org/). The volatility aspect will be modelled using rugarch package developed by Ghalanos (2015) in *R*. Estimations of parameters will be executed with the use of maximum likelihood estimation approach. After this, the goodness of fit of the estimated parameters will be examined, followed by checking for model adequacy.

3.2 Time series heteroscedasticity

The assumption of homoscedasticity in least square models has for a long time been the foundation of applied econometrics, and it assumes that the expected value of

the error terms when squared is the same at any time. Autoregressive moving average (ARMA) models also share this idea of homoscedasticity when modelling the conditional expectation of a process with reference to the past. But suppose we are interested in modelling the unusual volatility of financial data like the recent daily returns of stock in an equity market. This might yield a next day's return that reflects more variation than the previous and this cannot be captured by ARMA model due to its constant conditional variance, σ^2. This type of situation requires a model that can accommodate and model the non-constant volatility since the series moves through time. Such scenario is referred to as the autoregressive conditional heteroscedastic (ARCH) effects of financial time series process.

The ARCH effect or simply heteroscedastic effect of financial returns is usually witnessed in the form of "volatility clustering", that is, a process when small (big) changes follow small (big) changes in volatility. While the autoregressive conditional heteroscedastic (ARCH) model and the general autoregressive conditional heteroscedastic (GARCH) model filter (capture) volatility, ARMA model fixes the autocorrelation in the residuals of a financial return.

3.3 ARCH model

The ARCH model was first proposed by Engle (1982) for modelling the changing variance of a time series due to the non-constant conditional variance of financial asset's return series (r_t), given past returns. For the conditional variance, the ARCH (1) model is specified as:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2, \tag{3.3.1}$$

where α and ω are unknown parameters and ε_{t-1}^2 is a lagged innovation term. Also $\omega \geq 0$, and $\alpha \geq 0$ are sufficient conditions to ensure that $\sigma_t^2 \geq 0$. The σ^2 are

conditional variances because they are assumed to depend on past realisations of the error process (O'Donnell and Rayner, 2009), or simply, a linear function of the past squared innovations (Engle, 1982; Stavros and Evdokia, 2004). That is, the model sets out the variance of the current error term conditional on past realisations of the squares of the error terms. In other words, the variance at time t (σ_t^2) can be modelled conditionally on (i.e., using values of the) past squared observations. Conditioning is often expressed as regressions of a variable's future value on the past and present values of the same variable (Fabozzi, 2013).

This model is generalised to an ARCH(v) model as:

$$\sigma_t^2 = \omega + \sum_{j=1}^{v} \alpha_j \varepsilon_{t-j}^2, \tag{3.3.2}$$

where α_j are parameters with $j = 1, ..., v$. Here, σ^2 will be positive if $\omega > 0$ and $\alpha_j > 0$ for all j. These are sufficient conditions, but they are not necessary for positivity of the conditional variances (Engle, 1982; Bollerslev, 1986; O'Donnell and Rayner, 2009).

3.4 Parsimonious parameterisation of the GARCH model

Engle (1982), introduced a volatility process with time varying conditional variance; the ARCH process. However in practice, empirical evidence shows that a large lag order or high ARCH order has to be selected for the ARCH modelling. This high ARCH order implies the estimation of many parameters, and that usually leads to tedious calculations. To reduce this computational burden, Bollerslev (1986) extended the ARCH model of Engle by including past conditional variances. This was actualised by proposing the generalised ARCH (GARCH) model as a natural solution to

the challenge faced with the high ARCH orders. The GARCH model is based on an infinite ARCH specification, and it intensely reduced the number of estimated parameters from an infinite number to just a few.

Thus, a GARCH specification often leads to a more parsimonious representation of the conditional variance process and provides added flexibility over the linear ARCH model when parameterizing the conditional variance. It gives parsimonious models that are easy to estimate and, even in its simplest form, has proven remarkably successful in forecasting conditional variances (Engle, 2001).

3.5 The GARCH model

Modelling of the magnitudes of the BRICS markets returns volatility will be done via the use of GARCH model developed by Bollerslev (1986). Furthermore, Smith (2003) observed that it is more reliable to remove volatility from the returns before fitting EVT models to the residuals (of the returns) to make the output more reasonable. The conditional variance of the GARCH model is stated as a linear function of its own lags, and the model is usually specified by its conditional variance and conditional mean equations. The simplest model specification is GARCH (1, 1) model with the mean equation defined as:

$$\text{Mean equation}: \ r_t = \mu + \varepsilon_t, \tag{3.5.1}$$

where r_t denotes the return series, ε_t is the part of the time series return that is random and it is known as the residuals, μ denotes the mean function usually expressed as an ARMA process, i.e.,

$$\mu_t = \sum_{i=1}^{p} \phi_i r_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i} \tag{3.5.2}$$

If the objective of the model is forecasting (or predicting) of future prices or returns for instance, for a given observation, the residual is the difference between the predicted value by the model and the observation (the observed value). In other words, the residual is the error of the model, which is referred to as the error term (Fabozzi, 2013). In this study, this volatile residual (ε_t) of the return will be modelled assuming the following distributions: normal, Student's t, generalised error distribution (with their skew variants) and the generalised hyperbolic distribution (GHYP).

The variance equation of the GARCH (k, v) can be stated as

$$\varepsilon_t = z_t \sigma_t$$
$$z_t \sim N(0, 1)$$
$$\sigma_t^2 = \omega + \sum_{j=1}^{v} \alpha_j \varepsilon_{t-j}^2 + \sum_{i=1}^{k} \beta_i \sigma_{t-i}^2, \tag{3.5.3}$$

i.e., the conditional variance σ_t^2 of the GARCH (k, v) model is a linear function of the past conditional variances and past squared innovation. The ω represents the intercept and z_t denotes the standardised residual returns. The residuals are random variables with variance 1 and mean 0 (Smith, 2003) and they are known to be independent and identically distributed, i.i.d., (McNeil and Frey, 2000).

3.6 Test for stationarity

Before using the GARCH model for volatility analysis, the data series will be tested for stationarity via the unit root tests. The test is relevant to avoid a situation where non-stationarity in the data leads to problems in statistical inference. Stationarity is achieved by differencing the logged data. A process is non-stationary when the linear stochastic process has a unit root. A stationary data is stated as $I(0)$, while an integrated order one data series that is non-stationary is written as $I(1)$.

The stationarity of the series will be examined and the results compared using two unit root tests via the Augmented Dickey Fuller (ADF) test and Phillips-Perron test, and one stationarity test via the KPSS test.

3.6.1 Augmented Dickey Fuller (ADF) test

The Augmented Dickey Fuller (ADF) test of Dickey and Fuller (1979) will be used to carry out the test for stationarity on the price and return series (data). This test is an augmented form of the Dickey-Fuller (DF) test for a more complicated and larger set of models of time series (than the DF test which is valid only if the series is a first-order autoregressive process (i.e. AR (1)) (Box and Jenkins, 1970) as stated in equation 3.6.1)

$$y_t = \rho y_{t-1} + \delta' \mathbf{A}_t + \varepsilon_t, \qquad t = 1, ..., T, \tag{3.6.1}$$

where $\mathbf{A}_t$ is the vector of the deterministic terms, the residuals $\varepsilon_t\{t = 0, 1, ...\}$, i.e., the non-systematic components of the model, are assumed to be white noise (a sequence of independent and identically distributed random variables that are uncorrelated with zero mean and finite variance), while ρ and δ' are parameters to be estimated. Equation (3.6.1) can be estimated to obtain the standard DF test (in equation 3.6.2) after taking the first difference by subtracting y_{t-1} from both sides of the equation, where $\varphi = \rho - 1$.

$$\begin{aligned} y_t - y_{t-1} &= \rho y_{t-1} - y_{t-1} + \delta' \mathbf{A}_t + \varepsilon_t \\ \Delta y_t &= (\rho - 1) y_{t-1} + \delta' \mathbf{A}_t + \varepsilon_t \\ \Delta y_t &= r_t = (\rho - 1) r_{t-1} + \delta' \mathbf{A}_t + \varepsilon_t \\ r_t &= \varphi r_{t-1} + \delta' \mathbf{A}_t + \varepsilon_t, \qquad \text{for, } t = 1, ..., T. \end{aligned} \tag{3.6.2}$$

That is, for financial time series data, the returns (r_t) is the change in the price index series y_t. This explains the r_t obtained from taking the first difference of the logarithms of the price index (P) in equation (4.3.1) (Section 4.3) where $r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) = \ln P_t - \ln P_{t-1} = y_t - y_{t-1}$. The ADF test formulation permits higher order autoregressive processes because many financial and economic time series have a more complicated dynamic structure than what a simple AR(1) model can capture (Rossi, 2014). Hence, the ADF test corrects this difficulty of the DF test by constructing a parametric correction for higher-order correlation where it is assumed that the series follows an AR(p) process (see Equation 3.6.3). The null and alternative hypotheses of the ADF test may be written as:

$$H_0 : \rho = 1, \text{ data is not stationary } I(1)$$
$$H_1 : |\rho| < 1, \text{ data is stationary } I(0).$$

The ADF test is executed with the assumption that the data dynamics take an ARMA structure, and the test is centered on estimating the test regression:

$$r_t = \varphi r_{t-1} + \delta' \mathbf{A}_t + \sum_{j=1}^{p} \Psi_j \Delta r_{t-j} + \varepsilon_t, \tag{3.6.3}$$

where $\mathbf{A}_t$ is an optional exogenous regressors and may comprise of a constant, or a linear trend, or a constant and trend. The p lagged difference terms, Δr_{t-j}, are used to approximate the ARMA structure of the errors, with the value of p set to make the homoscedastic error serially uncorrelated. The Δ is the first difference operator, while δ', $\varphi = (\rho - 1)$, and Ψ are the parameters to be estimated. The test statistic is based on the least squares estimates (in equation 3.6.4)

$$t_{ADF} = \frac{\hat{\varphi}}{SE(\hat{\varphi})}, \tag{3.6.4}$$

where $\hat{\varphi}$ is the least square estimate of φ and $SE(\hat{\varphi})$ is its standard error estimate. A decision to reject the null hypothesis can be reached if the critical value is greater than the test statistic value, where it will then be concluded that the series is not integrated of order 1, i.e. "no unit-root is present". A practical difficulty associated with the ADF test is in the choice of lags p. If p is too low, autocorrelation will affect the test, and if p is too large, the power of test will be lower. Hence, Schwert (1989) gave the suggestion of choosing the maximum lag $p_{max} = 12(T/100)^{1/4}$ (Arltová and Fedorová, 2016).

Although there may seem to be restriction in the assumption that follows an autoregressive (AR) process, it has been demonstrated by Said and Dickey (1984) that the asymptotic of the ADF test is valid even with a moving average (MA) component. Hence, the general ARMA(p, q) models can be accommodated (Rossi, 2014) as long as sufficient lagged difference terms are included in the test regression (EViews 9.0, 2016).

The limiting distribution of test statistics for the ADF models is identical with that of the DF test statistics' distribution and for $T \to \infty$ is tabulated in MacKinnon (1991) and Dickey (1976). The DF test statistic follows the Dickey-Fuller distribution, and the distribution's critical values (obtained by a simulation) have been tabulated in Dickey (1976) and Fuller (1976) (see Arltová and Fedorová, 2016). The asymptotic distributions of the unit root tests are generally non-standard and non-normal (Rossi, 2014).

3.6.2 Phillips-Perron (PP) test

An alternative approach of testing for a unit root, where serial correlation can be controlled, was developed by Phillips and Perron (1988). This approach is known as the Phillips-Perron (PP) method and it is based on the Dickey-Fuller test to test the null hypothesis that a time series is integrated of order 1 (i.e. H_0: $\rho = 1$) in

$r_t = (\rho - 1)r_{t-1} + \delta' \mathbf{A}_t + \varepsilon_t$. Like the ADF test, the PP test controls for a higher order serial correlation in the return series beyond the AR(1) limit of the Dickey-Fuller (DF) t-test by making a non-parametric correction to the t-test statistic, hence it invalidates the DF test. Furthermore, using this standard Dickey-Fuller test with non-parametrically modified test statistics, the PP test addresses the difficulty usually encountered in the selection of lag p in a regression model for a unit root testing (Arltová and Fedorová, 2016).

The PP method adjusts the t-ratio of the coefficient to ensure that serial correlation has no effect on the test statistic' asymptotic distribution. The PP test is built on the statistic:

$$\tilde{t}_{\varphi} = t_{\varphi} \left(\frac{\gamma_0}{f_0} \right)^{1/2} - \frac{T(f_0 - \gamma_0)(SE(\hat{\varphi}))}{2 f_0^{1/2} s}, \tag{3.6.5}$$

where t_{φ} is the t-ratio of φ and $\hat{\varphi}$ is the estimate, while s is the standard error of the test regression. The term f_0 represents an estimator of the residual spectrum at frequency zero, and γ_0 is a consistent estimate of the error variance in $r_t = (\rho - 1)r_{t-1} + \delta' \mathbf{A}_t + \varepsilon_t$, and it is calculated as $(T - k)s^2/T$, with k denoting the number of regressors (EViews 9.0, 2016). The null hypothesis that a time series is integrated of order 1 (i.e. non-stationary) can be rejected if p-value < 0.05.

When performing the PP test, the two choices to be made are, first, to indicate whether the test regression should include (or not include) a constant, a constant and a linear time trend. Second, a method for estimating f_0 would have to be chosen, and likely choices can be based on autoregressive spectral density estimation, or kernel-based sum-of-covariances (EViews 9.0, 2016). The asymptotic distribution of the ADF statistic and that of the PP modified t-ratio is the same (Rossi, 20014). However, the performance of the Phillips–Perron test is worse in finite samples than the ADF test (Davidson and MacKinnon, 2004).

3.6.3 The Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test

The two unit root tests (of the ADF and PP) mentioned in Sections 3.6.1 and 3.6.2 are based on testing the null hypothesis that the time series y_t is (a unit root) non-stationary, i.e. it is integrated of order one $I(1)$, against the alternative of stationarity, $I(0)$. However, KPSS test (Kwiatkowski, Phillips, Schmidt and Shin, 1992) described the scenario in the opposite way by testing the null hypothesis that the time series y_t is stationary, $I(0)$, around a deterministic trend (i.e. trend-stationary) against the alternative of a unit root, $I(1)$. The KPSS time series is expressed as the sum of deterministic trend (d_t), random walk (rw_t), and stationary random error (ε_t) (Arltová and Fedorová, 2016). It is based on the model:

$$\begin{aligned} y_t &= d_t + rw_t + \varepsilon_t \\ rw_t &= rw_{t-1} + u_t, \end{aligned} \tag{3.6.6}$$

where d_t is the deterministic part of the model (Kwiatkowski, Phillips, Schmidt and Shin, 1992), rw_t is a random walk with variance σ_u^2, ε_t are i.i.d. $N(0, \sigma_\varepsilon^2)$, and u_t are i.i.d. $N(0, \sigma_u^2)$.

The KPSS test is based on a Lagrange multiplier (LM) test of the hypothesis that the random walk has a variance of zero, i.e. $H_0 : \sigma_u^2 = 0$, which denotes that rw_t is a constant, against the alternative hypothesis $H_1 : \sigma_u^2 > 0$. The test statistic is stated as:

$$\text{LM} = \sum_{t=1}^{T} s_t^2 / \hat{\sigma}_\varepsilon^2, \tag{3.6.7}$$

where s_t is a cumulative residual function i.e. $s_t = \sum_{t=1}^{T} \hat{\varepsilon}_t, \quad t = 1, 2, ..., T$, and $\hat{\sigma}_\varepsilon^2$ is the variance's estimate of process ε_t from the equation (3.6.6). The reported critical values for the LM test were derived by a simulation and are based upon the

asymptotic results presented in KPSS (Table 1, p. 166) (EViews 9.0, 2016). If the null hypothesis is rejected, then the series has a unit root. Moreover, if the results of the ADF and PP tests above show a unit root but the result of the KPSS test show a stationary process, one should be careful and opt for the latter result (Rossi, 2013).

In conclusion, since financial returns modelling displays a more complex structure than what the traditional AR (1) model-based DF test can analyse, this study applies the ADF and PP tests for the unit root test. In addition, a stationarity test via the KPSS is also applied and the results compared to that of the ADF and PP tests.

3.7 Test for serial correlation

After the data is made stationary, it is required to test for the presence of short-range linear dependence termed serial correlation or autocorrelation in the residual after fitting an ARMA model, which may occur as a result of the relationship between a variable (like financial data) and its lagged version over various intervals of time. This test can be implemented with the use of the Weighted Ljung-Box test of Fisher and Gallagher (2012). This weighted Portmanteau test statistic is given by

$$\widetilde{Q}_W = n(n+2)\sum_{k=1}^{m}\frac{(m-k+1)}{m}\frac{\hat{r}_k^2}{n-k} \tag{3.7.1}$$

The statistic $\widetilde{Q}_W$ can be interpreted as a weighted Ljung-Box test, where $\hat{r}_k^2$ is the residual sample autocorrelation of order k, n represents the sample size, and m is the tested number of lags . The residual at lag one is given the most weight, 1, while the residual at lag m is given the least weight, $1/m$.

The test statistic is asymptotically distributed as χ^2 random variable with $p+q$ degrees of freedom; where p and q are respectively the autoregressive order and moving average terms estimated in the fitted model. If $p-\text{value} < 0.05$, the null hypothesis is rejected, i.e. H_0 : No serial correlation in the residuals.

The test will be carried out together with the AC (autocorrelation) and PAC (partial autocorrelation) functions of the residuals.

3.8 Test for ARCH effects

Following the filtering of short-range dependence from the data series, it is usually required to test the presence of heteroscedasticity or ARCH effects in the data before fitting the GARCH models on the residuals of the return series. The presence of ARCH effects can be tested using ARCH LM Test or McLeod-Li test. Both methods will be used in this study.

3.8.1 ARCH LM test

The presence of heteroscedasticity in the residuals of the five markets' return series will be tested with the application of the Lagrange Multiplier (LM) test for ARCH effects introduced by Engle (1982).

To test for ARCH effects in the returns' conditional variance, the test procedure is in two phases: First, we fit an ARMA (p, q) model to the return series of each of the indices:

$$r_t = \alpha_o + \sum_{i=1}^{p} \phi_i r_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{3.8.1}$$

where ϕ and θ are parameters to be estimated. The residual ε_t is obtained by running the linear regression on the model. Next, a regression of the squared residuals (ε_t^2) in equation (3.8.2) is run on a constant φ_0 and lagged squared residuals up to order k to test for heteroscedasticity (ARCH effects).

$$\hat{\varepsilon}_t^2 = \varphi_0 + \left(\sum_{j=1}^{k} \varphi_j \hat{\varepsilon}_{t-j}^2 \right) + \psi_t \tag{3.8.2}$$

The Engle's LM test statistic is asymptotically χ^2 distributed with ν degrees of freedom, where ν is the number of estimated parameters, i.e. $v = k + 1$, and ψ_t is the error term associated with the regression model. The test is calculated as the product of the number of observations (M) and (R^2) from the test regression, i.e. (MR^2), where R is the sample multiple correlation coefficient (Wang et al., 2005). The presence of heteroscedasticity can be tested using these hypothesis:

$H_0: \ \varphi_1 = \varphi_2 = \cdots = \varphi_p = 0$
against the alternative hypothesis
$H_1: \ \varphi_1 \neq 0, \ \varphi_2 \neq 0, \ \cdots, \ \varphi_p \neq 0$
i.e. $\alpha_j \neq 0$ for some j

This is the null hypothesis of "no heteroscedasticity" up to order p in the residual. A decision can be made to reject the null hypothesis if the test statistic value is greater than the critical value. The presence of ARCH effects will indicate that volatility varies with time, which in turn validates and creates a platform for the use GARCH modelling.

3.8.2 McLeod-Li test

McLeod-Li test can also be used as an alternative to detect the presence of ARCH effects by plotting the p-values of the test using a number of lags. McLeod and Li (1983) proposed a formal test for ARCH effect based on the Ljung-Box test. They considered the autocorrelation functions of the squares of the time series and tested whether the first n squared residuals' autocorrelations from an ARMA model can be used to test for ARCH effects. The Ljung-Box Q-Statistics of McLeod-Li test is given by

$$Q = S_z(S_z + 2)\sum_{l=1}^{n} \frac{\hat{\kappa}_l^2(\varepsilon^2)}{S_z - l}, \tag{3.8.3}$$

where S_z is the sample size and $\hat{\kappa}_l^2$ is the squared sample autocorrelation of squared residual series at lag l. The asymptotic distribution of the test statistics is χ^2 with n degrees of freedom, under the null hypothesis of "no heteroscedasticity in the data".

3.9 Conditional distributions

This section describes the seven main univariate error distributions used in this study for modelling the markets' volatilities, and they include the normal, skewed-normal, Student's t, skewed-student's t, GED, skewed-GED and the generalised hyperbolic distribution (GHYP).

3.9.1 The normal distribution

The normal distribution is characterised entirely by its mean and variance (which are its first two moments). It is a symmetric and unimodal (i.e., singe-peaked) distribution having zero excess kurtosis and zero skewness. A random variable X can be described as normally distributed with mean μ and variance σ^2. The density is stated as (Ghalanos, 2015):

$$f(x) = \frac{e^{\frac{-\frac{1}{2}(x-\mu)^2}{\sigma^2}}}{\sigma\sqrt{2\pi}}. \tag{3.9.1}$$

When the residuals ϵ is standardised by σ^2 (following a mean filtration process), it produces the standard normal density expressed as

$$f\left(\frac{x-\mu}{\sigma}\right) = f(z) = \frac{1}{\sigma}\left(\frac{e^{-\frac{1}{2}z^2}}{\sqrt{2\pi}}\right) \tag{3.9.2}$$

3.9.2 Skewed distributions by inverse scale factors

Fernandez and Steel (1998) proposed the introduction of skewness into symmetric and unimodal distributions by introducing inverse scale factors in the negative and

positive real half lines (Ghalanos, 2015). Let a skew parameter be represented by ξ (when $\xi = 1$, it shows that the distribution is symmetric), the density of a random variable X can be stated as

$$f(x|\xi) = \frac{2}{\xi + \xi^{-1}} \left[f(\xi x) H(-x) + f(\xi^{-1} x) H(x) \right], \tag{3.9.3}$$

where $H(.)$ denotes the Heaviside function and $\xi \in \mathbb{R}^+$. The function in equation (3.9.4) generates the absolute moments that are required for deriving the central moments (Ghalanos, 2015).

$$M_r = 2 \int_0^\infty x^r f(x) dx. \tag{3.9.4}$$

Then the mean and variance are defined as

$$E(x) = M_1(\xi - \xi^{-1}) \tag{3.9.5}$$

$$Var(x) = (M_2 - M_1^2)(\xi + \xi^{-2}) + 2M_1^2 - M_2 \tag{3.9.6}$$

The Normal, Student and GED distributions have skew variants which have been standardised to unit variance and zero mean by using the moment conditions given above (Ghalanos, 2015). Details of the skewed-normal, skewed-student's t and skewed-GED are given in Sections 3.9.3, 3.9.5 and 3.9.7 respectively.

3.9.3 The skewed-normal distribution

The skewed-normal (SN) is a parametric class of probability distributions with a shape parameter ξ that regulates the skewness. Hence, it is an extension of the normal distribution that allows for a continuous variation from normality to non-normality (Ashour and Abdel-hameed, 2010).

Let $\phi(\cdot)$ and $\Phi(\cdot)$ be the standard normal density function and (its) distribution function respectively for a random variable X with ξ as the shape parameter (Eling,

2014; Ashour and Abdel-hameed, 2010). The probability density function (pdf) of the skewed-normal distribution is then given as (see Azzalini, 1985)

$$f(x) = 2\phi(x)\Phi(\xi x), \quad (-\infty < x < \infty), \ \xi \in \mathbb{R} \tag{3.9.7}$$

The density function in equation (3.9.7) can be written as (Ashour and Abdel-hameed, 2010)

$$f(x) = 2\frac{1}{\sqrt{2\pi}}\exp(-x^2/2)\int_{-\infty}^{\xi x}\frac{1}{\sqrt{2\pi}}\exp(-t^2/2)dt, \tag{3.9.8}$$

where $X \sim SN(0, 1, \xi)$ for $-\infty < x < \infty$ and any given $\xi \in \mathbb{R}$. When $\xi = 0$, the SN distribution reduces to the normal distribution, but it becomes the half-normal distribution if $\xi \to \pm\infty$. The skewness of the distribution increases as ξ increases (in absolute value). The square of a random variable that follows an SN distribution is a Chi-square variable with one degree of freedom ($X^2 \sim \chi_1^2$) irrespective of the value of ξ (Azzalini, 1985).

For practical numerical work, the scale and location parameters can be incorporated by transforming linearly as $Y = \mu + \sigma X$, where μ and σ are the location and scale parameters respectively. This then follows the skewed-normal distribution $Y \sim SN(\mu, \sigma^2, \xi)$, with $\sigma > 0$ (Eling, 2014).

An alternative representation of the skewed-normal distribution that is useful for financial (return) modeling was presented by Pourahmadi (2007). In this way, the expression $Y \sim SN(\mu, \sigma^2, \xi)$ is written as a weighted average of a half-normal and a standard normal variable:

$$Y = \mu + \sigma X = \mu + \sigma\left(\delta|Z_1| + \sqrt{1-\delta^2}Z_2\right), \tag{3.9.9}$$

with $\delta = \xi/\sqrt{1+\xi^2} \in [-1, 1]$. The Z_1 and Z_2 are independent $N(0, 1)$ random variables. The Y becomes $N(\mu, \sigma^2)$ if $\delta = 0$. If equation (3.9.9) is interpreted in

the language of financial economics, the return Y is driven (with the inclusion of the location parameter μ) by a Gaussian element Z_2 modulated by $\sigma\sqrt{1-\delta^2}$ and a half-Gaussian element $|Z_1|$ modulated by $\sigma\delta$ (Eling, 2014). The skewness of the distribution becomes more pronounced to the left (right), the closer the value of δ is to $-1(+1)$. The impact of δ can then be highlighted by the mean, variance, skewness and kurtosis of Y as follows

$$\begin{aligned} \mathrm{E}(Y) &= \mu + \sigma\sqrt{2/\pi}\delta, && (3.9.10)\\ \mathrm{Var}(Y) &= \sigma^2(1-2\delta^2/\pi), && (3.9.11)\\ \mathrm{Skewness}(Y) &= (4-\pi)/2(\delta(2/\pi)^{1/2})^3/(1-2\delta^2/\pi)^{3/2}, && (3.9.12)\\ \mathrm{Excess\ \ Kurtosis}(Y) &= 2(\pi-3)(\delta(2/\pi)^{1/2})^4/(1-2\delta^2/\pi)^2. && (3.9.13) \end{aligned}$$

There is a quadratic link for the variance, whereas the mean is a linear increasing function in the skewness parameter δ. Unlike the normal distribution, the values of the skewed-normal distribution ranges from -1 to 1 and can therefore be calibrated to skewed data. However, the range of possible skewness values is still comparatively limited (Eling, 2014).

3.9.4 The Student's t distribution

Like the normal distribution, the Student's t distribution is also unimodal, symmetric and bell-shaped, but with thicker (or heavier) tails than the normal distribution. As an alternative to the normal distribution, the Student's t distribution can be used to fit the standardised innovations (residuals). It is wholly described by a shape parameter ξ, but its 3-parameter representation are used for standardization as follows:

$$f(x) = \frac{\Gamma\left(\frac{\xi+1}{2}\right)}{\sqrt{\sigma\xi\pi}\Gamma\left(\frac{\xi}{2}\right)}\left(1+\frac{(x-\mu)^2}{\sigma\xi}\right)^{-\left(\frac{\xi+1}{2}\right)}, \qquad (3.9.14)$$

where Γ is the Gamma function, while ξ, μ, and σ are the shape, location and scale parameters respectively. The mean (and mode) of the t distribution signifies the location parameter μ while the variance is stated as (Ghalanos, 2015)

$$\text{Var}(X) = \frac{\sigma\xi}{(\xi - 2)}. \tag{3.9.15}$$

It is required that $\text{Var}(X) = 1$ for standardization purposes, hence

$$\sigma = \frac{(\xi - 2)}{\xi} \quad (\text{from } \frac{\sigma\xi}{(\xi - 2)} = 1). \tag{3.9.16}$$

With $\frac{(\xi-2)}{\xi}$ substituted into equation (3.9.14), the standardised Student's t distribution gives

$$f\left(\frac{x-\mu}{\sigma}\right) = f(z) = \frac{1}{\sigma}\frac{\Gamma\left(\frac{\xi+1}{2}\right)}{\sqrt{(\xi-2)\pi}\Gamma\left(\frac{\xi}{2}\right)}\left(1 + \frac{(z)^2}{(\xi-2)}\right)^{-\left(\frac{\xi+1}{2}\right)}. \tag{3.9.17}$$

The Student's t distribution has excess kurtosis equal to $6/(\xi - 4)$ for $\xi > 4$ and zero skewness.

3.9.5 The skewed-Student's t distribution

The skewed version of the student's t distribution simply termed the "skewed-student's t" distribution was introduced by Branco and Dey (2001), and further developed by Azzalini and Capitanio (2003). This distribution allows regulating both kurtosis and skewness, and that is why it is particularly valuable for modeling capital market data (Eling, 2014). Using the transformation in equation (3.9.18), the standardised student's t skewed distribution can be stated as

$$X = \frac{Z}{\sqrt{M/v}}, \quad M \sim \chi^2(v), \tag{3.9.18}$$

where Z represents an independent $SN(0, 1, \xi)$; but if $N(0, 1)$ is used instead, it would yield the standard t (Eling, 2014). The parameter v denotes the degrees of freedom.

The random variable Y has a skew-t distribution, i.e. $Y \sim ST\ (\mu,\ \sigma^2,\ \xi,\ v)$, with parameters $(\mu,\ \sigma^2,\ \xi,\ v)$ following linear transformation from X, such that $Y = \mu + \sigma X$. The mean and variance of the distribution can be calculated as follows (more complex presentation of the kurtosis and skewness can be seen in Azzalini and Capitanio, 2003)

$$\mathrm{E}(Y) = \mu + \sigma w \delta \tag{3.9.19}$$

$$\mathrm{Var}(Y) = \sigma^2 \left(\frac{v}{v-2} - w\delta^2 \right), \tag{3.9.20}$$

where $w = \sqrt{\frac{v}{\pi}} \frac{\Gamma\left(\frac{1}{2}(v-1)\right)}{\Gamma\left(\frac{1}{2}v\right)}$, and δ is a skewness parameter. The variance is a quadratic function on δ, while the mean is a linear increasing function in δ. When the skewed-student's t distribution is compared with the skew normal distribution, the former is known to take more extreme values for both skewness and kurtosis (Eling, 2014).

3.9.6 The generalised error distribution

The GED is symmetric with a 3-parameter distribution that belongs to the family of exponential distributions with conditional density stated as

$$f(x) = \frac{\xi e^{-\frac{1}{2}\left|\frac{x-\mu}{\sigma}\right|^{\xi}}}{2^{1+\xi^{-1}}\sigma\Gamma(\xi^{-1})}, \tag{3.9.21}$$

where μ, ξ, and σ are the location, shape and scale parameters respectively. The GED is also a unimodal distribution where the location parameter μ is the mean of the distribution. The kurtosis, Kur and variance, Var are stated as:

$$Kur(x) = \frac{\Gamma(5\xi^{-1})\Gamma(\xi^{-1})}{\Gamma(3\xi^{-1})\Gamma(3\xi^{-1})} \tag{3.9.22}$$

$$Var(x) = \sigma^2 2^{\frac{2}{\xi}} \frac{\Gamma(3\xi^{-1})}{\Gamma(\xi^{-1})} \tag{3.9.23}$$

The density becomes flatter and flatter as ξ decreases and it approaches the uniform distribution as the limit $\xi \to \infty$. When $\xi = 1$ and $\xi = 2$, the distribution becomes the Laplace and the normal respectively. A unit standard deviation in equation (3.9.24) can be obtained by rescaling the density during standardization

$$Var(x) = \sigma^2 2^{\frac{2}{\xi}} \frac{\Gamma(3\xi^{-1})}{\Gamma(\xi^{-1})} = 1 \tag{3.9.24}$$

$$\therefore \sigma = \sqrt{2^{-\frac{2}{\xi}} \frac{\Gamma(\xi^{-1})}{\Gamma(3\xi^{-1})}}$$

When this is substituted into the scaled density of z, it becomes

$$f\left(\frac{x-\mu}{\sigma}\right) = f(z) = \frac{\xi e^{-\frac{1}{2}\left|\sqrt{2^{-\frac{2}{\xi}} \frac{\Gamma(\xi^{-1})}{\Gamma(3\xi^{-1})}} z\right|^{\xi}}}{\sqrt{2^{-\frac{2}{\xi}} \frac{\Gamma(\xi^{-1})}{\Gamma(3\xi^{-1})}} 2^{1+\xi^{-1}} \Gamma(\xi^{-1})} \tag{3.9.25}$$

3.9.7 The skewed-GED distribution

The skewed variant of the GED is called the "skewed-GED (SGED)" distribution. The standardised SGED distribution for a random variable X has a density function expressed as (Lee and Pai, 2010)

$$f(x_t|\xi,\delta) = \mathcal{C} \,.\, \exp\left(-\frac{|x_t - \eta|^{\xi}}{[1 - sign(x_t - \eta)\delta]^{\xi}\theta^{\xi}}\right), \tag{3.9.26}$$

where $\mathcal{C} = \frac{\xi}{2\theta.\Gamma\left(\frac{1}{\xi}\right)}$, $\theta = \Gamma\left(\frac{1}{\xi}\right)^{1/2} \Gamma\left(\frac{3}{\xi}\right)^{-1/2} \mathcal{S}(\delta)^{-1}$, $\eta = 2\delta \cdot A \cdot \mathcal{S}(\delta)^{-1}$, $\mathcal{S}(\delta) = \sqrt{1 + 3\delta^2 - 4A^2\delta^2}$ and $A = \Gamma\left(\frac{2}{\xi}\right) \Gamma\left(\frac{1}{\xi}\right)^{-1/2} \Gamma\left(\frac{3}{\xi}\right)^{-1/2}$.

The fat-tails and height of the density function are directed by the shape parameter ξ with constraint $\xi > 0$, whereas δ denotes the density's skewness parameter with $-1 < \delta < 1$. *Sign* represent the sign function. The SGED distribution takes the form of the standard normal distribution when $\xi = 2$ and $\delta = 0$. The density function skews to the left (right) with negative (positive) skewness (Lee and Pai, 2010).

3.9.8 The generalised hyperbolic distribution

The generalised hyperbolic distribution (GHYP) is a normal mixture of variance-mean in which the generalised inverse Gaussian distribution (GIG) is the mixing distribution. It is a continuous probability distribution introduced by Ole Barndorff-Nielsen (1977), with the pdf (probability density function) given based on the second kind of modified Bessel function represented by K_λ (Ole, 2013). Furthermore, it has semi-heavy tails property that makes it relevant and regularly used in risk management and modelling of financial markets data.

Standardization and estimation of the density requires estimating two invariant parameters, i.e. the location and scale parameters (denoted by υ, $\aleph$), that denote the shape and skewness combined. Following this, is a series of transformation stages to scale and translate the 2 parameters $(\upsilon, \aleph)$ into the parametrization of $(\varsigma, \vartheta, \hbar, \imath)$ which generates standard formulae for the likelihood function (Ghalanos, 2015).

Definition 3.9.1. Definition of the standardised generalised hyperbolic distribution

Let a random variable with variance 1 and mean 0 be represented by X_t, such that X_t is distributed as GHYP $(\upsilon, \aleph)$. Now let the scaled version of X_t be represented by Z with variance (1), such that Z is also distributed as GHYP $(\upsilon, \aleph)$. We can state

the density $f(.)$ of z as

$$f\left(\frac{x_t}{\sigma};\upsilon,\aleph\right)=\frac{1}{\sigma}f_t(z;\upsilon,\aleph)=\frac{1}{\sigma}f_t(z;\tilde{\varsigma},\tilde{\vartheta},\tilde{\hbar},\tilde{\imath}) \tag{3.9.27}$$

As applied in the rugarch package of Ghalanos (2015) for the volatility modelling, the density can only be expressed in the $(\varsigma,\vartheta,\hbar,\imath)$ parametrization. The transforming steps of the parametrization from $(\upsilon,\aleph)$ to the $(\varsigma,\vartheta,\hbar,\imath)$, while also standardizing for mean 0 and variance 1 is as follows. Let

$$\upsilon=\hbar\sqrt{\varsigma^2-\vartheta^2} \tag{3.9.28}$$

$$\aleph=\frac{\vartheta}{\varsigma}, \tag{3.9.29}$$

which can be written in terms of ϑ and ς after some substitution as,

$$\vartheta=\varsigma\aleph \tag{3.9.30}$$

$$\varsigma=\frac{\upsilon}{\hbar\sqrt{(1-\aleph^2)}}. \tag{3.9.31}$$

It is required for standardization that,

$$E(X)=\imath+\frac{\vartheta\hbar}{\sqrt{\varsigma^2-\vartheta^2}}\frac{K_{\lambda+1}(\upsilon)}{K_\lambda}=\imath+\frac{\vartheta\hbar^2}{\upsilon}\frac{K_{\lambda+1}(\upsilon)}{K_\lambda(\upsilon)}=0 \tag{3.9.32}$$

$$\therefore\imath=-\frac{\vartheta\hbar^2}{\upsilon}\frac{K_{\lambda+1}(\upsilon)}{K_\lambda(\upsilon)} \tag{3.9.33}$$

$$\mathrm{Var}(X)=\hbar^2\left(\frac{K_{\lambda+1}(\upsilon)}{\upsilon K_\lambda(\upsilon)}+\frac{\vartheta^2}{\varsigma^2-\vartheta^2}\left(\frac{K_{\lambda+2}(\upsilon)}{K_\lambda(\upsilon)}-\left(\frac{K_{\lambda+1}(\upsilon)}{K_\lambda(\upsilon)}\right)^2\right)\right)=1 \tag{3.9.34}$$

$$\therefore\hbar=\left(\frac{K_{\lambda+1}(\upsilon)}{\upsilon K_\lambda(\upsilon)}+\frac{\vartheta^2}{\varsigma^2-\vartheta^2}\left(\frac{K_{\lambda+2}(\upsilon)}{K_\lambda(\upsilon)}-\left(\frac{K_{\lambda+1}(\upsilon)}{K_\lambda(\upsilon)}\right)^2\right)\right)^{-\frac{1}{2}} \tag{3.9.35}$$

Now expressing $\vartheta^2/(\varsigma^2-\vartheta^2)$ as

$$\frac{\vartheta^2}{\varsigma^2-\vartheta^2}=\frac{\varsigma^2\aleph^2}{\varsigma^2-\varsigma^2\aleph^2}=\frac{\varsigma^2\aleph^2}{\varsigma^2(1-\aleph^2)}=\frac{\aleph^2}{(1-\aleph^2)} \tag{3.9.36}$$

as a result of this, the formula for $\hbar$ can be re-written in terms of the parameter estimates $\hat{\upsilon}$ and $\hat{\aleph}$ as,

$$\hbar=\left(\frac{K_{\lambda+1}(\hat{\upsilon})}{\hat{\upsilon}K_{\lambda}(\hat{\upsilon})}+\frac{\hat{\aleph}^2}{(1-\hat{\aleph}^2)}\left(\frac{K_{\lambda+2}(\hat{\upsilon})}{K_{\lambda}(\hat{\upsilon})}-\left(\frac{K_{\lambda+1}(\hat{\upsilon})}{K_{\lambda}(\hat{\upsilon})}\right)^2\right)\right)^{-\frac{1}{2}} \tag{3.9.37}$$

To transform into the $(\tilde{\varsigma},\tilde{\vartheta},\tilde{\hbar},\tilde{\imath})$ parametrization, equation (3.9.37) is substituted into equation (3.9.31), and then simplified as follows:

$$\begin{aligned}
\tilde{\varsigma}&=\frac{\hat{\upsilon}\left(\frac{K_{\lambda+1}(\hat{\upsilon})}{\hat{\upsilon}K_{\lambda}(\hat{\upsilon})}+\frac{\hat{\aleph}^2\left(\frac{K_{\lambda+2}(\hat{\upsilon})}{K_{\lambda}(\hat{\upsilon})}-\frac{(K_{\lambda+1}(\hat{\upsilon}))^2}{(K_{\lambda}(\hat{\upsilon}))^2}\right)}{(1-\hat{\aleph}^2)}\right)^{\frac{1}{2}}}{\sqrt{(1-\hat{\aleph}^2)}}\\
&=\frac{\left(\frac{\hat{\upsilon}K_{\lambda+1}(\hat{\upsilon})}{K_{\lambda}(\hat{\upsilon})}+\frac{\hat{\upsilon}^2\hat{\aleph}^2\left(\frac{K_{\lambda+2}(\hat{\upsilon})}{K_{\lambda}(\hat{\upsilon})}-\frac{(K_{\lambda+1}(\hat{\upsilon}))^2}{(K_{\lambda}(\hat{\upsilon}))^2}\right)}{(1-\hat{\aleph}^2)}\right)^{\frac{1}{2}}}{\sqrt{(1-\hat{\aleph}^2)}}\\
&=\left(\frac{\frac{\hat{\upsilon}K_{\lambda+1}(\hat{\upsilon})}{K_{\lambda}(\hat{\upsilon})}}{(1-\hat{\aleph}^2)}+\frac{\hat{\upsilon}^2\hat{\aleph}^2\left(\frac{K_{\lambda+2}(\hat{\upsilon})}{K_{\lambda+1}(\hat{\upsilon})}\frac{K_{\lambda+1}(\hat{\upsilon})}{K_{\lambda}(\hat{\upsilon})}-\frac{(K_{\lambda+1}(\hat{\upsilon}))^2}{(K_{\lambda}(\hat{\upsilon}))^2}\right)}{(1-\hat{\aleph}^2)^2}\right)^{\frac{1}{2}}\\
&=\left(\frac{\frac{\hat{\upsilon}K_{\lambda+1}(\hat{\upsilon})}{K_{\lambda}(\hat{\upsilon})}}{(1-\hat{\aleph}^2)}\left(1+\frac{\hat{\upsilon}\hat{\aleph}^2\left(\frac{K_{\lambda+2}(\hat{\upsilon})}{K_{\lambda+1}(\hat{\upsilon})}-\frac{K_{\lambda+1}(\hat{\upsilon})}{K_{\lambda}(\hat{\upsilon})}\right)}{(1-\hat{\aleph}^2)}\right)\right)^{\frac{1}{2}}
\end{aligned} \tag{3.9.38}$$

Lastly, the remaining parameters are obtained recursively from $\tilde{\varsigma}$ and the earlier results,

$$\tilde{\vartheta}=\tilde{\varsigma}\hat{\aleph} \tag{3.9.39}$$

$$\tilde{\hbar} = \frac{\hat{v}}{\tilde{\varsigma}\sqrt{1-\hat{\aleph}^2}}, \tag{3.9.40}$$

$$\tilde{\imath} = \frac{-\tilde{\vartheta}\tilde{\hbar}^2 K_{\lambda+1}(\hat{v})}{\hat{v}K_{\lambda}(\hat{v})}. \tag{3.9.41}$$

3.10 Summarised stages of the GARCH modelling

The stages involved in the modelling of the BRICS markets' financial returns are as follows: (1) we test and make sure that the returns are stationary using two unit root tests (via the ADF and PP tests) and one stationarity test (through the KPSS test); (2) next, the stationary returns are tested for serial correlations and necessary adjustments are made to correct any significant linear dependence observed using appropriate ARMA models; (3) after correcting for autocorrelation in the returns, the resulting (approximately) independent and identically distributed (i.i.d.) residuals are tested for heteroscedasticity (ARCH effects) using the ARCH LM and McLeod-Li tests; (4) if ARCH effects are observed in the returns' residuals, appropriate GARCH (k, v) models will be fitted under the assumption of seven error distributions to ascertain the error distribution that can best describe each of the markets. The fitted GARCH model will be used to filter out heteroscedasticity in the residuals and model the level of volatility inherent in each of the BRICS market.

The sections that follow give a detailed discussion into the univariate and multivariate EVT for modelling the risk in each of the BRICS markets and extremal dependence in the pairwise combination of the markets.

3.10.1 Model selection

A model is an approximation of the reality displayed in a given set of observed data (Shibata, 1986; Rao and Wu, 2001). Myung (2004) further stated that a statistical

model M is a parametric collection of probability distributions, indexed by model parameters:

$$M = \{f(x|\theta) \mid \theta \in \Omega\}, \tag{3.10.1}$$

where $f(x|\theta)$ is a family of density functions parameterised by θ (i.e. the parameter of interest), f itself is the probability function on the p-dimensional space $\boldsymbol{R}^p$, and Ω is the parameter space (Shibata, 1986; 2002). The methods of model selection can be used to identify useful models, based on predictive accuracy or generalizability (Myung, 2004), and a model that is described as useful should be able to fit the data well (Breheny, 2013).

Representative statistical models are usually built by researchers to test hypotheses and theories, and then ascertain how well each of the models fit collected data. The credibility of a single theory can be tested by fitting one statistical model (that represents the theory) to the data, after which the explanatory or predictive power of the model is evaluated, and its parameters interpreted. In many statistical analyses, the mechanisms of a phenomenon are sometimes represented by contending theories, hence they have two or more contending or candidate models.

To determine the theory that best describes the observed data, the candidate models are compared using an objective method. The model that is considered the best by the stipulated method is chosen as the "best" model, and it favorably indicates that the theory can be represented by the model (Christensen, 2018). Some of the model selection methods include the likelihood ratio tests (LRT), coefficient of determination (R^2), adjusted coefficient of determination (R^2 adj), generalised cross-validation (GCV), the information criteria, among others.

The LRT is used to determine the model that is better between two contending nested models. A nested model is one model, say, a condensed model (or sub-model) derived from another model, i.e. a broad model. The broad model contains all the parameters of the sub-model, and the number of parameters contained in the broad

model must be greater than the ones in the sub-model (Cahill, 2003; Kim et al., 2017). Theorem (3.10.1) summarises the LRT as follows.

Theorem 3.10.1. *Suppose the broad model with parameter* $\theta^0 = (\theta_1, \theta_2)$ *and the sub-model with parameter* θ_2 *are represented by* M_b *and* M_s *respectively. Let the deviance statistic* D *for comparing the two models be expressed as*

$$D = 2\{l_b(M_b) - l_s(M_s)\},$$

where $l_b(M_b)$ *and* $l_s(M_s)$ *are the log-likelihood's maximised values for models* M_b *and* M_s*. The validity of the sub-model* (M_s) *in relation to the broad model* (M_b) *can be tested at the* α *level of significance by rejecting* M_s *for* M_b *if* $D > c_\alpha$*. Here,* c_α *is defined as the (1 -* α*) quantile of the* χ^2_k *distribution, where* $k = k_b - k_s$ *is the degrees of freedom, and* k_b *and* k_s *are numbers of parameters for* M_b *and* M_s *respectively.*

Model comparison approaches that are based on the null hypothesis testing context, like the likelihood ratio tests (LRT) and R^2 change tests are simple to conduct and reasonably intuitive, but there are settings for which they are not suitable (Christensen, 2018).

The coefficient of determination R^2 is not a good criterion for model selection because it always tends to overfit by choosing the most complex model (Breheny, 2013). That is, it always increases when a variable (whether relevant or completely irrelevant) is added to the model, and it can be made equal to one (i.e. its maximal value) when a sufficient number of regressors are included (Dufour, 2007). The R^2 method can be summarised as follows:

Theorem 3.10.2. *For a linear regression model of the type:*

$$y_t = x_{t1}\varphi_1 + x_{t2}\varphi_2 + \cdots + x_{tk}\varphi_k + \varepsilon_t, \quad t = 1, ..., T, \tag{3.10.2}$$

where y_t represents the dependent variable and $x_{t1}, ..., x_{tk}$ are the explanatory variables, ε_t is a random disturbance that is assumed to be uncorrelated with (or independent of) the explanatory variables. For a classical linear model, the disturbances $\varepsilon_1, ..., \varepsilon_T$ are assumed to be independent and identically distributed (i.i.d) based on $N(0, \sigma^2)$ distribution, and the regressors can also be assumed to be taken as fixed.

The coefficient of determination R^2 is the proportion of the variance of the dependent variable in the regression model in equation (3.10.2) *and it is explained as*

$$R^2 = 1 - \frac{\hat{V}(\varepsilon)}{\hat{V}(y)}, \tag{3.10.3}$$

where $\hat{V}(\varepsilon) = \sum_{t=1}^{T} \hat{\varepsilon}_t^2/T$, $\hat{V}(y) = \sum_{t=1}^{T}(y_t - \overline{y})^2/T$, $\overline{y} = \sum_{t=1}^{T} y_t/T$, and the least squares residuals are $\hat{\varepsilon}_1, ..., \hat{\varepsilon}_T$.

An extended version of R^2 called "R^2 adj" was proposed by Thei (1961) to address and adjust the problem of overfitting in R^2 criterion by replacing $\hat{V}(y)$ and $\hat{V}(\varepsilon)$ with the corresponding unbiased estimators $v_y^2 = \sum_{t=1}^{T}(y_t - \overline{y})^2/(T-1)$ and $v^2 = \sum_{t=1}^{T} \hat{\varepsilon}_t^2/(T-k)$ to give

$$R^2 \text{ adj} = 1 - \frac{v^2}{v_y^2} = 1 - \frac{T-1}{T-k}(1-R^2) = R^2 - \frac{k-1}{T-k}(1-R^2), \tag{3.10.4}$$

where k is the number of independent regressors, that is, the number of explanatory variables in the model (Dufour, 2011), excluding the constants, T is the sample size or the number of observations in the dataset. However, as a limitation, while the R^2 adj does not encourage overfitting, it does not really impose penalty on it either, like information criteria do (Breheny, 2013). It can also be seen from the equation in (3.10.4) that an increase in the number of regressors may increase R^2 adj. Furthermore, R^2 adj is known to be inconsistent such that the rule for maximizing it does not choose the true model with a probability converging to 1 (see equation

3.10.7) (Gouriéroux and Monfort, 1995; Dufour, 2007). More so, R^2 and R^2 adj are not particularly suitable for choosing between models that have different numbers of variables (Breheny, 2013).

Theorem 3.10.3. *Let* $\boldsymbol{f}_\lambda = (f_\lambda(x_1), ..., f_\lambda(x_n))^T$, *where* T *is the transpose vector and denote* $A(\lambda)$ *as the smoothing matrix that is defined by* $\boldsymbol{f}_\lambda = A(\lambda)\mathbf{y}$. *To use the GCV, the core computational effort is the calculation of the degrees of freedom, i.e. the trace* $trA(\lambda)$, *for the spline. With the way* f_λ *is represented in Reinsch (1967; 1971), the smoothing matrix can be stated in the form of the inverse of a particular banded matrix of bandwidth* $2m + 1$. *Hutchinson and de Hoog (1985) developed an efficient* $O(m^2n)$ *algorithm to calculate the diagonal elements of* $A(\lambda)$ *by using the band structure and Cholesky decomposition, and thus evaluate* $trA(\lambda)$. *The confidence intervals for the spline estimate can also be obtained by the diagonal elements of* $A(\lambda)$, *known as the leverage values. Using the local support basis for* f_λ *in Lyche and Schumaker (1973), the technique in O'Sullivan (1985) produces another* $O(m^2n)$ *algorithm for the computation of* $trA(\lambda)$ *and the leverage values (see Section 3.8.1 of Gu, 2002). For the GCV criterion built on QR, there are also efficient* $O(m^2n)$ *algorithms factorization.*

The generalised cross-validation (GCV) (Craven and Wahba, 1979; Wahba, 1990) is used as described in Theorem (3.10.3) (Lukas et al., 2010) to estimate the predictive quality of a model. The selected model's size is viewed as smoothing parameter which balances closeness of fit and complexity. A model's complexity is a measure of the number of selected variables and can be understood as a penalty (Jansen, 2015). The GCV generally performs well when the sample size is large, but can be unreliable with smaller sample size, and even for large sample size, it occasionally gives a parameter value (estimate) that is far too small (Lukas et al., 2010).

Selection of models using information criteria, underpinned by the information-theoretic framework are more flexible than methods grounded in the null hypothesis testing (Hamaker et al., 2011). As an advantage, information criteria do not necessitate the contending models to relate with each other in any particular way, and the contending (candidate) models do not need to be nested. More so, the use of information criteria can accommodate the simultaneous comparison of any number of models, i.e. models comparison is not limited to two at a time, nor is it compulsory to test models incrementally (like Model 1 versus Model 2, Model 2 versus Model 3, etc.). In conclusion, for the fact that information criteria are not based on the framework of null hypothesis testing, designation of any model as a null model is not needed for comparison purposes (Christensen, 2018).

Common information criteria as defined in equations (3.10.6), (3.10.9), (3.10.10) and (3.10.11) include the Akaike information criterion (AIC), Bayesian information criterion (BIC), Hannan-Quinn information criterion (HQIC) and Shibata information criterion (SIC) respectively. An information criterion is a function of the value of log-likelihood (the goodness of fit component) and the number of the model's parameters (the model complexity component). In general, the smaller (bigger) the information criterion (the log-likelihood), the better the model's fit.

3.10.2 Akaike information criterion

The first information criterion was introduced by Akaike (1974), and it is known as the Akaike information criterion (AIC) as stated in equation (3.10.6). Information criteria's logic is rooted in information theory. The first mathematical definition of information was proposed by Claude E. Shannon (Shannon and Warren, 1949) and discussed in subsequent years. One of the first monographs on information theory was authored by S. Goldman and published by Prentice-Hall (see Goldman, 1953). AIC is based on Kullback-Leibler (K−L) distance (Kullback and Leibler, 1951), in

which information theory is connected to random variable distributions. Following the "K−L distance" concept provided in Burnham and Anderson (2002), the distance (or discrepancy) between two models are briefly conceptualised as follows:

Let $f(x)$ denote one of the two models, known as the "generating" or "true" model from which the actual realizations occur. The second (approximating) model, denoted by $g(x)$, is a model stated in line with the theory of the researcher, about the true model. Now based on information theory, the K−L distance (in equation 3.10.5) is described as the information lost when $g(x)$ is used to approximate $f(x)$ (Konishi and Kitagawa, 1996). If $g(x)$ and $f(x)$ are exactly the same, then the K−L discrepancy between the two models is 0; otherwise, the discrepancy is greater than 0 (Christensen, 2018). That is, we can approach model selection by trying to make small the K−L information (Kullback and Leibler, 1951; Konishi and Kitagawa, 1996)

$$I\{g(x); f(x|\hat{\theta})\} = \int g(x)\log g(x)dx - \int g(x)\log f(x|\hat{\theta})dx, \tag{3.10.5}$$

which measures the divergence or discrepancy of $g(x)$ relative to $f(x|\hat{\theta})$, where $\hat{\theta}$ denotes the parameter estimate of a p-dimensional vector of unknown parameters.

The relative quality of statistical models from a given collection of models for a set of data can be estimated by the Akaike information criterion (AIC). Hence, the AIC is a model selection estimator that estimates the quality of individual model in relation to each of the other models.

$$\text{AIC} = \frac{-2l}{n} + \frac{2p}{n}, \tag{3.10.6}$$

where n denotes the sample size, l is the log-likelihood of the maximum likelihood $L(\Theta)$, and Θ is the unknown parameter vector. The $2p$ is used to penalise overparameterization (overfitting), where p denotes the number of estimated parameters (Li et al., 2009), and it is multiplied by 2 (i.e. $2p$) to simplify the algebra (Dufour, 2007).

The model with the lowest value of AIC should be the one selected.

In statistics, when a data generating process is represented by a statistical model, such representation will practically never be exact because some of the information will be lost. The amount of information lost by a given model can be estimated by the AIC such that the lesser (higher) the information a model loses, the higher (lesser) the model's quality. The AIC therefore rewards goodness of fit (measured by the likelihood function) and penalises overfitting (i.e. overparameterization) in the model. The goodness of fit of a model improves with increasing number of parameters, but an imposed penalty acts as a check to curtail overparameterization in the model.

One of the information criterion's desirable property is efficiency, and the AIC is known to be efficient, i.e. it asymptotically minimises a loss function (Claeskens and Hjort, 2008; Chaurasia and Harel, 2013). Efficiency of the AIC is also described by Christensen (2018) in the context of selecting the best model when the dimension (i.e. the number of parameters) of the generating model is infinite, and that is impossible to fit in practice.

Another desirable property of information criterion in large samples is "consistency" as described in equation (3.10.7). The AIC however is criticised for not being asymptotically consistent because it does not select the most parsimonious true model with probability converging to one, as the sample size becomes larger (Shibata, 1976). That is, the AIC has a high probability of picking a model with "too many parameters" as the sample size $n \to \infty$ (Dufour, 2007; Bozdogan, 1987; Christensen, 2018). It is further believed that this inconsistency comes to bear because the computation of the AIC does not directly involve sample size (Janssen and De Boeck, 1999; Forster, 2004; Li et al., 2009).

$$Pr[I_c(M_{SC}) = I_c(M_T)] \to 1 \quad (\text{as } n \to \infty), \tag{3.10.7}$$

where I_c is the information criterion, M_{SC} is a selected candidate model, and M_T

denotes the true model.

Shono (2005) further describes the inconsistency using the relative weight of the second term of the right side in equation (3.10.8) (i.e. the penalty term component $2p$ of the AIC). The author showed that this component becomes very small compared to that of $(-2) \times$ (maximum log-likelihood l), when the sample size is large. As a result of this, AIC is asymptotically equivalent to maximum log-likelihood as

$$\text{AIC} = -2l + 2p \approx -2l \ \ (\text{as } n \to \infty). \tag{3.10.8}$$

Because of this, AIC tends to choose the complicated model. Hence, consistency property is not satisfied by AIC (Shibata, 1989) except in special scenarios where the ratio p (i.e. number of unknown parameters)/n (sample size) is kept fixed (Shono, 2005).

3.10.3 Bayesian information criterion

The Bayesian information criterion (BIC) was proposed by Schwarz (1978) as an asymptotic approximation to a transformation of the Bayesian posterior probability of a candidate model. It is computed as stated in equation (3.10.9) based on the empirical log-likelihood and does not need the priors to be specified (Chaurasia and Harel, 2013). The BIC is consistent such that, as the sample size increases, the criterion will select a true model of finite dimension if it is included among the candidate models (Christensen, 2018). That is, the BIC chooses the true model with probability one (equation 3.10.7), with the assumption that the set of selected candidate models contains the true model (see Nishii, 1984; Shibata, 1981).

Using consistency property, the BIC discourages overparameterization by imposing more heavy penalties on model complexity than does the AIC (Li et al., 2009). Hence, it tends to favour highly parsimonious models which is neither overly complex

nor overly simple but usually lies between the two extremes (Breheny, 2013).

$$\text{BIC} = \frac{-2l}{n} + \frac{p \log_e(n)}{n}, \tag{3.10.9}$$

where $\log_e$ represents the natural logarithm. To be consistent, a criterion is expected to select the correct model regardless of the size of the sample, which indicates that such criterion should select the same model even with an increase in the sample size (Buckland et al., 1997). This can be guaranteed by increasing the penalty term as the sample size increases (Pollmann, 2015).

3.10.4 Hannan-Quinn information criterion

The Hannan-Quinn information criterion (HQIC) expressed in equation (3.10.10) was proposed by Hannan-Quinn (1979) and Hannan (1980), and it is a model selection criterion that is based on the law of iterated logarithm (Javed and Mantalos, 2013). It also satisfies the consistency property (like the BIC) as the sample size increases (and becomes large) (Shono, 2005).

$$\text{HQIC} = \frac{-2l}{n} + \frac{2p \log_e(\log_e(n))}{n}. \tag{3.10.10}$$

3.10.5 Shibata information criterion

The Shibata information criterion in equation (3.10.11) follows the same pattern as the BIC and HQIC, where the sample size n is included in the complexity component (see Ghalanos, 2015). It is least frequently used in the literature in comparison with the AIC, BIC and HQIC.

$$\text{SIC} = \frac{-2l}{n} + \log_e\left(\frac{n+2p}{n}\right) \tag{3.10.11}$$

The coefficients for p in the model complexity component of each information criteria (AIC, BIC, HQIC and SIC) formula shows the degree of penalty imposed on the number of model parameters.

Although consistent and efficient information criteria denote different approaches to model selection, they are often used jointly in applied work and are usually presented together with each other in software (Christensen, 2018). The AIC and BIC are commonly used in the literature (Chaurasia and Harel, 2013), while the HQIC and especially the SIC (see Ghalanos, 2015) seems to have not been that frequently used in practice.

Steps involved in model selection

The following three-step processes take place during selection of models using information criteria after the set of candidate models is defined. First, each candidate model is fitted to the same data, and it must be ensured that any transformation applied to the outcome variable is maintained across all contending models (Burnham and Anderson, 2002).

Second, the desired information criteria are obtained for each model. Each of the stated contending models has a unique contribution to the value of an information criterion through the estimation of its log-likelihood and the model's number of parameters. Furthermore, when computing information criteria that depends on the size of the data sample, such sample size is consistent across candidate models. Third, the candidate models are compared by ranking them with regards to the (values of the) information criteria used. The model with the lowest (or least value closest to zero) is chosen as the "best" model (Christensen, 2018). The choice of a most adequate model that can return accurate forecasting results is very essential in financial data analysis.

3.11 Overview of extreme value theory

The Value-at-Risk (VaR) and correlation coefficients methods are the classical risk and extremal dependence models, but their use in financial returns modelling has been challenged due to their normal distribution assumption. Alternative models like the point process, CEV, copulas and extreme value copulas based on EVT are found to be more efficient in modelling the non-linear financial returns. Hence, extreme value theory approach is usually used in computing tail (extreme) risk measures because it provides the fundamental requirements for statistical modelling of asymptotic rare events.

As a brief description on copulas, let $\mathbf{X}$ be a two-dimensional random vector such that $\mathbf{X} = (X, Y) \stackrel{\mathrm{d}}{=} \mathbf{X}_T$, with F_X and F_Y as the continuous distributions of the components X and Y respectively. Let the quantile function for any distribution function H be defined by

$$H^{\leftarrow}(T) = \inf\{x \in \mathbf{R} : H(x) \geq T\}, \quad T \in (0, 1). \tag{3.11.1}$$

Then

$$P(X \leq F_X^{\leftarrow}(x),\ Y \leq F_Y^{\leftarrow}(y)) = C(x, y). \tag{3.11.2}$$

The linking function C is called a copula and it denotes a distribution function on [0, 1] with uniform margins. The copula function models the dependence between variables X and Y, while $F_X(X)$ and $F_Y(Y)$ have uniform distributions on (0, 1). The use of copulas for extreme value statistics modelling can be found in monographs by de Haan and de Ronde (1998), and Galambos (1987) among others. The purpose of the copula is to transform the vector $\mathbf{X}_T$ into identical marginal distribution and then model the dependence between the margins. Copula transforms the margins to standard uniform or Frétchet distribution, which makes it easy to estimate the joint

distribution of $\mathbf{X}$.

The main difficulties in MEVT is the estimation of the spectral measure (i.e. the likelihood of the extremes' directions) for datasets with very large values, and the estimation of the index of regular variation (which is the likelihood of the distance from the source where extremes exist). It is believed by Mikosch, 2005 that these two quantities cannot be estimated in a simpler way with the introduction of copulas.

Another possible challenge is in the choice of a copula that can be viewed as reasonable for modelling involving multivariate extreme value distributions. Because copulas connote any dependence structure between two (or more) random variables, it is not obvious a priori which copulas follow the extreme value theory property (Mikosch, 2005). Although extreme value copulas, like the Gumbel copula, have been proposed as possible copulas for modeling extreme events, the fit of these copulas needs some justification to prove that the data come from a mechanism that can generate extreme value (Mikosch, 2005).

The EVT approach (via the point process and CEV models) can explicitly account for tail events since its focus of modelling interest is confined to only extreme values within the tail of the distribution, and not on the entire observations in the dataset (Herrera and Schipp, 2014). Hence, this study will use these two EVT based models of the conditional multivariate extreme value (with details in Sections 3.14 to 3.18.4) and point process (described in Sections 3.22 to 3.24) to analyse the risk and extremal dependence of the markets.

The risk in each of the selected markets will be modelled by the univariate version of the two selected EVT models of the CEV and point process. Furthermore, the EVT models become more useful when forecasting rare events movement using multiple variables concurrently via the MEVT modelling framework. This study will limit the MEVT to the bivariate case, where pairwise combinations of the BRICS markets will be modelled one after the other.

3.12 The history of EVT

Fisher and Tippett (1928) are the pioneers in the work of extreme value limiting distribution of samples of independent and identically distributed (i.i.d.) random variables' maximum distributions. Todorovic and Zelenhasic (1970) are forerunners in the use of the "peaks-over-threshold" (POT) approach of EVT, while classical articles for fitting EVT distributions, specifically in environmental studies can be traced to the works of Hosking, Wallis and Wood (1985), and Prescott and Walden (1980). Measures on the mathematical development of the EVT in relation to extreme order statistics can be found in the works of Pickands (1975), Hill (1975), and Weissman (1978).

The development of the limiting distribution of largest order statistics are also traceable to the work of Leadbetter et al. (1983). The description of GPD as a stable distribution of threshold exceedances are seen in Davidson and Smith (1990), and Smith (1984). Sibuya (1960) narrates the asymptotic dependence and independence of variables of the MEVT, while the multivariate point proceses are described in de Haan (1985), and de Haan and Resnick (1977). General overview on EVT can be seen in Reiss and Thomas (2007), de Haan and Ferreira (2006), Bierlant et al. (2004), and Coles (2001).

To ensure comprehensive understanding of the statistical theories of EVT models, this study will begin with the description of the two fundamental approaches involving the traditional block maxima method (BMM) and the peaks-over-threshold (POT) method. The peaks-over-threshold (POT) method is tailored down to the conditional extreme value (CEV) model and the point process approach since both methods are based on the choice of a reasonably high quantile threshold.

3.13 Block maxima method

3.13.1 The univariate case

With block maxima model, extreme losses are divided into identical blocks where maximum loss in each block is the largest observation. This procedure is termed "block maxima" with each local block containing maximum loss. Thus, the extreme losses can be characterised by the local block maxima which are the data to be fitted or modelled using this method. Theorem (3.13.1) describes the generalised extreme value distribution (GEVD) as the asymptotic approximation of the block maxima observations as stated by Fisher and Tippett (1928) and Gnedenko (1943).

Theorem 3.13.1. *Let X_1, X_2, ... be i.i.d (independent and identically distributed) random variables whose distribution function is F and let $M_j = \{max\}\ X_l$, for $(j-1)q < l \leq jq$, that is, the block maxima, defined for $q = 1, 2, ...$ and $j = 1, 2, ..., k$. Therefore, the total number of observations becomes $n = q \times k$, i.e. k blocks of size q. The distribution F is assumed to be in the domain of attraction of some extreme value distribution, known as the generalised extreme value distribution (GEVD)*

$$G_\xi(x) = \exp\left[-\left(1+\xi\frac{x-\mu}{\sigma}\right)_+^{-\frac{1}{\xi}}\right] \tag{3.13.1}$$

for $\left\{x|1+\xi\left(\frac{x-\mu}{\sigma}\right) > 0\right\}$,

where ξ, μ and $\sigma > 0$ denote the tail shape parameter, the location parameter representing the center of the distribution, and scale parameter (i.e. the size of the deviations about μ) respectively.

For suitably selected sequence of normalising constants $a_q > 0$ and b_q, and all x

$$\begin{aligned}\lim_{q\to\infty} P\left(\frac{M_j - b_q}{a_q}\right) &= \lim_{q\to\infty} F^q(a_q x + b_q) \\ &= G(x), \quad j = 1, 2, ..., k,\end{aligned} \tag{3.13.2}$$

where $(M_j - b_q)/a_q$ is a normalised maximum. It is required that both $k = k_n \to \infty$ and $q = q_n \to \infty$ as $n \to \infty$ so as to attain meaningful limit outcomes (Ferreira and de Haan, 2015). When normalised, G can have its place in one of the following three non-degenerate families:

$$G_G(x) = e^{-e^{-x}}, \quad x \in \mathbb{R}, \qquad \text{i.e. the Gumbel distribution (Type 1) with } \xi \to 0,$$

$$G_F(x)_\alpha = \begin{cases} 0, & x \leq 0 \\ e^{-x^{-\alpha}}, & x > 0, \ \alpha > 0 \end{cases} \quad \text{i.e. the Fréchet distribution (Type 2) with } \xi > 0,$$

$$G_W(x)_\alpha = \begin{cases} e^{(-x)^\alpha}, & x \leq 0 \\ 1, & x > 0, \ \alpha > 0 \end{cases} \quad \text{i.e. the Weibull distribution (Type 3) with } \xi < 0.$$

In risk management analysis, the distribution of the block maxima M_j has a light tail with finite upper bound for the Weibull distribution, an exponential tail (that is unbounded) for the Gumbel distribution and a heavy tail that includes polynomial decay with no upper limit for the Fréchet distribution (Gilli and Kellezi, 2006).

As a limitation, the block maxima method can miss some of the required high realizations and retains some lower (central) ones (Coles, 2001). For this reason, preference is usually given to the peaks over threshold (POT) method because it uses all necessary high realizations, which makes it more reliable.

3.13.2 Formal hypothesis tests for a selected domain of attraction

Any selected domain of attraction, i.e., the Weibull, Gumbel or Fréchet model, for the estimated shape parameter under the GEVD can be tested using formal hypothesis testing procedures such as the likelihood-based methods. The test is necessary especially for practical use because the choice of a wrong model can be terrible (in practice) since it will lead to the choice of erroneous parameters for design. The three models have very different physical meanings due to their tails description. Hence, the need for proper identification of one of the three subfamilies models is very important practically as it is theoretically (Castillo et al., 2005).

3.13.3 The likelihood-based methods

Based on a given set of data $\mathbf{x} = \{x_1, ..., x_n\}$ with sample size n, these methods will be used to test the hypothesis:

$$H_0 : \xi = 0 \text{ (Gumbel)}$$
$$H_1 : \xi \neq 0 \text{ (Fréchet or Weibull).}$$

Given a parameter vector $\ell(x; \vartheta)$ for the log-likelihood with a function $\vartheta = (\mu, \sigma, \xi)$, where the three parameters represent the location, scale and shape respectively. Let the maximum likelihood estimates of ϑ under H_0 and H_1 be $\hat{\vartheta}_0 = (\hat{\mu}_0, \hat{\sigma}_0, 0)$ and $\hat{\vartheta}_1 = (\hat{\mu}_1, \hat{\sigma}_1, \hat{\xi}_1)$ respectively. These likelihood-based methods can be described under two asymptotically equivalent tests of the "likelihood ratio test" and "Wald tests" (Castillo et al., 2005).

3.13.4 The likelihood ratio test

The likelihood ratio (LR) test as stated in equation (3.13.3) compares $\ell(\mathbf{x}; \hat{\vartheta}_0)$ with $\ell(\mathbf{x}; \hat{\vartheta}_1)$. That is, it compares the evaluation of the likelihood at $\hat{\vartheta}_0$ with the evaluation of the likelihood at $\hat{\vartheta}_1$.

$$LR = 2\{\ell(\mathbf{x}; \hat{\vartheta}_1) - \ell(\mathbf{x}; \hat{\vartheta}_0)\}, \tag{3.13.3}$$

where LR is a χ^2 with 1 degree of freedom under the null hypothesis H_0.

For a better accuracy to the approximation of asymptotic distribution of the likelihood ratio test LR, a modification suggested by Hosking (1984) is given as

$$LR_{**} = \left(1 - \frac{2.8}{w}\right) LR, \tag{3.13.4}$$

where w is the sample values i.e. the number of exceedances or cluster-maxima above the threshold u. A decision for the rejection of H_0 can be made at a level of significance α if

$$LR_{**} > \chi_1^2(1-\alpha), \tag{3.13.5}$$

The critical values $\chi_1^2(1-\alpha)$ is the $(1-\alpha)$ quantile of the χ^2 distribution with 1 degree of freedom (Castillo et al., 2005).

3.13.5 Wald tests

The Wald tests are based on the comparison of the estimate $\hat{\xi}_1$ with its standard error, $\hat{\varphi}_{\hat{\xi}_1}$, obtained from the square-root of the third diagonal element of the inverse of the Fisher information matrix in equation (3.13.6).

$$\mathbf{I}_{\vartheta} = \begin{bmatrix} g_{11} & g_{12} & g_{13} \\ g_{21} & g_{22} & g_{23} \\ g_{31} & g_{32} & g_{33} \end{bmatrix}. \tag{3.13.6}$$

Prescott and Walden (1980) stated the elements of $\mathbf{I}_{\vartheta}$ as follows:

$$g_{11} = E\left(-\frac{\partial^2}{\partial\mu^2}\right) = \frac{n}{\sigma^2}p, \tag{3.13.7}$$

$$g_{22} = E\left(-\frac{\partial^2}{\partial\sigma^2}\right) = \frac{n}{\sigma 2\xi 2}\{1 - 2\Gamma(2-\xi) + p\}, \tag{3.13.8}$$

$$g_{33} = E\left(-\frac{\partial^2}{\partial\xi^2}\right) = \frac{n}{\xi^2}\left\{\frac{\pi^2}{6} + \left(1 - \eta - \frac{1}{\xi}\right)^2 + \frac{2q}{\xi} + \frac{p}{\xi^2}\right\}, \tag{3.13.9}$$

$$g_{12} = g_{21} = E\left(-\frac{\partial^2}{\partial\mu\partial\sigma}\right) = \frac{n}{\sigma^2\xi}\{p - \Gamma(2-\xi)\}, \tag{3.13.10}$$

$$g_{31} = g_{13} = E\left(-\frac{\partial^2}{\partial\mu\partial\xi}\right) = -\frac{n}{\sigma\xi}\left(q + \frac{p}{\xi}\right), \tag{3.13.11}$$

$$g_{23} = g_{32} = E\left(-\frac{\partial^2}{\partial\sigma\partial\xi}\right) = \frac{n}{\sigma\xi^2}\left[1 - \eta - \frac{1-\Gamma(2-\xi)}{\xi} - q - \frac{p}{\xi}\right], \tag{3.13.12}$$

where the Gamma function is stated as

$$\Gamma(u) = \int_0^{\infty} x^{u-1}e^{-x}dx$$

The Psi function is expressed as

$$\phi(u) = \frac{d\log\Gamma(u)}{du} \tag{3.13.13}$$

Also,

$$p = (1-\xi)^2\Gamma(1-2\xi), \tag{3.13.14}$$

$$q = \Gamma(2-\xi)[\phi(1-\xi) - (1-\xi)/\xi], \tag{3.13.15}$$

and the Euler's constant $\eta = 0.5772157$.

There are two types of information matrix, namely: the observed I_O version and the expected I_E version as stated in equations (3.13.16) and (3.13.17) respectively.

$$I_O = -\frac{1}{n}\left(\frac{\partial^2 \ell(\mathbf{x};\vartheta)}{\partial\vartheta\partial\vartheta^T}\right) \tag{3.13.16}$$

$$I_E = -\frac{1}{n}E\left(\frac{\partial^2 \ell(\mathbf{x};\vartheta)}{\partial\vartheta\partial\vartheta^T}\right) \tag{3.13.17}$$

The Wald test statistic is expressed as

$$W = \frac{1}{n}\frac{\hat{\xi}_1^2}{\hat{\varphi}^2_{\hat{\xi}_1}} \tag{3.13.18}$$

Based on the version of the information matrix used for the standard error, the Wald test can be computed with standard error from either version of the information matrix. As it is with the likelihood ratio test, the χ^2 distribution with 1 degree of freedom is the limit distribution of all statistics under the H_0 of the Wald test. Hence, the Wald tests and likelihood ratio test are asymptotically optimal for the two-sided hypothesis test in Section (3.13.3) since they are asymptotically equivalent. The null hypothesis H_0 can be rejected when the Wald test statistic W is greater than the critical values $\chi^2_1(1-\alpha)$.

For a one-sided hypothesis test as shown equations (3.13.19) and (3.13.20) , these two test statistics (W and LR_{**}) are also asymptotically optimal but have their asymptotic limit under the H_0 as the standard normal distribution. That is, the standard normal one-tailed critical values may be used for the H_0 rejection procedure. The square root of any of the applicable statistics can be used and the square root's sign will be that of the ξ_1, where "1" is based on the observed information matrix.

$$H_0 : \xi = 0 \text{ (Gumbel)} \quad \text{versus} \quad H_1 : \xi < 0 \text{ (Weibull)} \tag{3.13.19}$$

$$H_0 : \xi = 0 \text{ (Gumbel)} \quad \text{versus} \quad H_1 : \xi > 0 \text{ (Frechet).} \tag{3.13.20}$$

For instance, under the LR (or LR_{**}) test statistic, negative deviations denotes $\xi < 0$ and positive deviations indicate $\xi > 0$. H_0 can be rejected if

$$\sqrt{LR_{**}} > \Phi(1-\alpha), \tag{3.13.21}$$

where $\Phi(1-\alpha)$ denotes the $(1-\alpha)$ quantile of the standard normal distribution (Castillo et al., 2005).

3.14 Peaks over threshold approach

This approach can be described by defining an extreme event as a value that exceeds a high enough threshold and uses only excess observations (i.e. peaks or exceedances) above the threshold for statistical inference. Every point over the chosen threshold is considered an extreme observation set aside for risk modelling.

This method will be demonstrated under two categories of threshold selection models: conditional extreme value model (CEV) and point process. The limiting distribution of the first model is the generalised Pareto distribution (GPD) while the second model can be approximated by a non-homogeneous Poisson distribution.

Semi-parametric models and fully parametric models are the two analytical frameworks that use POT class of models. The former is built around the Hill estimator and other related estimators (Danielsson, Hartmann & de Vries, 1998), while the latter is based on the GPD (Embrechts, Resnick & Samorodnitsky, 1998). When applied correctly, both methods are equally relevant but this study will prefer the use of the

fully parametric style because it is easier to estimate using the maximum likelihood approach.

3.15 The generalised Pareto distribution (GPD)

Theorem 3.15.1. *Suppose X is a random variable that holds in equation* (3.13.2)*, Pickands (1975) showed that as the threshold moves towards the endpoint $u \to u_{end}$ of this variable of interest, the limiting distribution of normalised excesses is the non-degenerate GPD described as:*

$$\Pr\{X \leq x | X > u\} \to H(x), \qquad u \to u_{end} \tag{3.15.1}$$

The GPD's expression as the limiting distribution of the threshold-excess model, where $y = x - u$, is stated as:

$$H(y) \quad = \quad \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma} y\right)^{-\frac{1}{\xi}}, & \text{for} \quad \xi \neq 0; \\ 1 - e^{-\frac{y}{\sigma}}, & \text{for} \quad \xi = 0, \end{cases} \tag{3.15.2}$$

where $H(y)$ is the GPD for $y \in [0, -\frac{\sigma}{\xi}]$ if $\xi < 0$ and $y \in [0, (x_F - u)]$ if $\xi \geq 0$. The scale and shape parameters are denoted by σ (where $\sigma > 0$) and ξ respectively.

3.16 Threshold-excess model

Given that the family of a class of approximating distribution to the tail of an arbitrary distribution function F of a random variable X is

$$G(x) = 1 - \jmath \left[1 + \xi \left(\frac{x - u}{\sigma_u}\right)\right]^{-\frac{1}{\xi}}, \qquad x > u, \tag{3.16.1}$$

where $\xi \neq 0$, $\jmath = \Pr(X > u)$, and $\sigma > 0$ for a family defined on $\{x - u : x - u > 0 \text{ and } (1+\xi(x-u)/\sigma_u) > 0\}$. This indicates that for a sufficiently large threshold, and on the condition that an individual observation exceeds the threshold u (i.e. $x > u$), $F(x) \approx G(x)$ with parameters $\jmath$, ξ, and σ (Coles, 2001).

3.17 Combining interval forecasts of the two models

Most existing approaches focus on point forecasting, where the prediction errors that exist cause uncertainties to the forecast results. But unlike point forecasting that focuses on prediction of single observations, interval forecast broadly quantifies the uncertainty surrounding a forecast. Hyndman and Fan (2010) stated that the full distribution of the forecasts can be given by probabilistic forecasting, which captures uncertainties in the forecasts (Sigauke, 2017). Quan et al. (2014) further stated that point forecasts face uninformative and unreliable problems when the level of uncertainty increases in data, but that prediction interval (PIs) or interval forecast are proposed to quantify uncertainties connected with the point forecasts. A decision maker may get interval forecasts for an uncertain quantity from multiple forecasters or analysts to acquire more information (Gaba et al., 2017).

If point forecasting is used for predicting the return levels, then at time s, the task will require forecasting the value of the return level for time $s + h$, where h denotes the predicting horizon. For interval forecast however, at time s the task will require forecasting an interval of values for time $s + h$ with a certain probability described as the confidence level (Rana et al., 2013).

More information can be obtained about the variability of a target variable and the related uncertainty when an interval of values is forecast via a prediction Interval

with a certain probability than just forecasting a single value. Prediction interval (PI) has a lower and upper bound between which the predicted value is projected to lie with a certain set probability (Chatfield, 2000; Rana et al., 2013). Based on this, interval forecasts are more suitable than point forecasts for quantification of the certainty of the point forecasts and for risk management in financial markets, inventory management and stock manufacturing (Chatfield, 2000; Rana et al., 2013).

Since the interval forecast from each of the two selected models contains useful independent information, it is usually unwise to discard one model in preference to the other. Hence, it has been empirically established that combining interval forecasts from individual models can yield a more robust and more accurate forecast (Bates and Granger, 1969; Clemen, 1989; de Menezes et al., 2000; Timmermann, 2006; Devaine et al., 2012; Gaillard, 2015; Nowotarski and Weron, 2014; Sigauke, 2017). Furthermore, empirical studies have shown that it is not possible to select the most reliable model from a diversity of models. This argument is supported by the findings of Aggarwal et al. (2009), who compared the results from 47 publications with the conclusion that there is no systematic evidence that one model consistently out-performs the other(s) (Nowotarski and Weron, 2014). This further supports the need for combining interval forecasts to obtain a single forecast which may possibly yield a big improvement. That being the said, it has also been argued that combining two forecasts does not (always) guarantee an improvement over the better of the two individual forecasts (Bates and Granger, 1969).

The following three methods or heuristics for combining two or more sets of forecasts as described in Sections (3.17.2), (3.17.3) and (3.17.4) will be used in this study. The three heuristics (Average, Median, and Envelop) are very useful because their computations require absolutely no distribution assumptions (Gaba et al., 2017).

3.17.1 Heuristics for combination of interval forecasts

For a random variable X, assume there is 100(1 - α)% forecast intervals $[L_j, U_j]$, j = 1, . . ., s, made available by s forecasters. Let the 100(1 - α)% combined forecast interval for X derived from the s individual intervals with heuristic h be represented by $[L_h, U_h]$ (Gaba et al., 2017).

3.17.2 Average (Ave)

The *Ave* heuristic involves a simple average of the intervals' endpoints stated as

$$L_{Ave} = (1/s)\sum_{j=1}^{s} L_j \text{ and } U_{Ave} = (1/s)\sum_{j=1}^{s} U_j. \tag{3.17.1}$$

The simple average (or arithmetic mean) method is often used to summarise data, be it in combining probability forecasts or point forecasts, because of its robustness, good performance and simplicity (Gaba et al., 2017), and it is broadly used in economic and business forecasting (Clemen, 1989; Genre et al., 2013; Nowotarski and Weron, 2014). *Ave* is in agreement with averaging quantiles that have been known for good performance if it is assumed that the intervals are symmetric in probability, such that $G_j(L_j) = 1 - G_j(U_j) = \alpha/2$, where G_j denotes forecaster j's cumulative distribution function for X. To compare the performance of the three selected heuristics, the *Ave* will be used as a benchmark against the other two (i.e. Med, and Env) parsimonious and simple heuristics.

3.17.3 Median (Med)

Another frequently used method for summarizing data is known as the median, represented as

$$L_{Med} = \text{Median}\{L_1, ..., L_s\} \text{ and } U_{Med} = \text{Median}\{U_1, ..., U_s\}. \quad (3.17.2)$$

The literature has shown that this heuristic has many desirable properties when used for combining probability distributions through its cumulative distribution function (cdf) (Hora et al., 2013). However, it is not as sensitive to extreme values as the mean (Gaba et al., 2017).

3.17.4 Envelop (Env)

The envelop method is an aggregate heuristic measure of summarizing data such that no forecasts are discarded or rejected however extreme they are (Gaba et al., 2017). It can be expressed as

$$L_{Env} = \min\{L_1, ..., L_s\} \text{ and } U_{Env} = \max\{U_1, ..., U_s\}. \quad (3.17.3)$$

Enveloping will normally produce a wider interval than the other two heuristics, except in a scenario where the individual intervals are alike. Based on this, it can be used to address the problem of overconfidence that may occur in individual forecasts. Overconfidence is described as a situation where distributions are too tight, which can cause inaccurate or incorrect calibration because of being too narrow (Gaba et al., 2017).

3.17.5 Prediction interval's quality assessment

In order to assess the quality of the constructed PIs, their width and coverage probability are usually considered. A good PI is expected to have a small or narrow width and a high coverage probability. Assessment based on coverage probability have been extensively used to measure the quality of PIs, but little attention has been given to the use of width for quality measures (Khosravi et al., 2011; Rana et al., 2013).

Hence, this study will focus on measures based on the width using the "Prediction Interval Normalised Averaged Width (PINAW)" metric in equation (3.17.4) to evaluate the quality of the constructed prediction intervals (PIs) for both the individual and combined forecasts.

$$\text{PINAW} = \frac{1}{nR}\sum_{i=1}^{n} a_i, \tag{3.17.4}$$

where n denotes the total number of predictions and a_i is the prediction interval width (PIW) stated as

$$\text{PIW}_i = U_i - L_i \quad i = 1, ..., n. \tag{3.17.5}$$

The range R, used to standardise the average width of the PIs in percentage, is the difference between the maximum a_i and minimum a_i.

The level of uncertainty of the datasets determines the widths of the PINAW. Wider PINAW are produced with a higher level of uncertainty, while lower uncertainty level yield smaller PINAW. Narrow PINAW indicates a high quality PIs is constructed, hence the width of PIs is required to be as small as possible (Quan et al., 2014).

3.18 Conditional multivariate extreme value modelling

3.18.1 The multivariate case

This study applies the multivariate analysis approach of Heffernan and Tawn (2004) for the extremal dependence modelling. Before estimating the dependence structure,

this process uses conditional multivariate approach by first fitting the marginal variables with the GPD models. As the GPD model is used for approximating exceedances above a threshold, this dependence structure is also conditioned on a variable exceeding a large enough threshold. To illustrate this on the BRICS stock markets, given the threshold exceedance of one of the markets' variables, the conditional multivariate approach can describe the conditional distribution of the remaining four markets, with the use of a regression type model.

3.18.2 Marginal transformation

Before using the regression type structure for modelling dependence, the original data scale must be marginally transformed to standard Laplace or Gumbel margins. We transform to the Laplace margin (see Theorem 3.18.1) because it simplifies the regression model's structure more than when transformed to the Gumbel margins (Southworth et al., 2016).

Theorem 3.18.1. *Let a p-dimensional random variable having arbitrary marginal distributions be represented by* $\ddot{\boldsymbol{X}} = (X_1, ..., X_p)$. *Let an estimate of the* i^{th} *marginal distribution function* $(i = 1, ..., p)$ *be denoted by* $\hat{F}_i$, *and let the standardised marginal distribution has its distribution function denoted by* G_s. *A transformed variable* $\ddot{\boldsymbol{Y}} = (Y_1, ..., Y_p)$ *having standardised marginal distributions is obtained from the original random variable* $\ddot{\boldsymbol{X}}$, *using the probability integral transform as follows:*

$$\ddot{Y}_i \;=\; (G_s^{-1}(\hat{F}_i(X_i))),\; i = 1, ..., p. \tag{3.18.1}$$

3.18.3 Regression model structure

Following the marginal transformation of the variables' data, Theorem (3.18.2) gives the description of the regression type structure used by the conditional extreme value

(CEV) model for the dependence modelling.

Theorem 3.18.2. *Let the variable on which to condition be represented by* $\ddot{Y}_i$, $i \in \{1, ..., p\}$. *Then* $\ddot{\mathbf{Y}}_{-i}$ *signifies the remainder of the vector* $\ddot{Y}$ *not including the ith component. The approach used by Heffernan and Tawn is conditioned on* $\ddot{\mathbf{Y}}_i$ *being above some high threshold* u, *and the dependence of the remaining* $\ddot{\mathbf{Y}}_{-i}$ *is modelled conditional on the observed value of* $\ddot{Y}_i$ *exceeding the threshold* u *i.e.* $\ddot{Y}_i > u$. *The specific choice of* G_s *in equation (3.18.1) dictates the form of the regression type model for the conditional dependence structure.*

3.18.4 Laplace margins

The Laplace distribution function is denoted by G_s and $\ddot{Y}$ are marginally Laplace distributed. Furthermore, on the condition that $\ddot{\mathbf{Y}}_i$ variable exceeds a high enough threshold u, the model of Heffernan and Tawn for the remaining variables $\ddot{\mathbf{Y}}_{-i}$ is given in equation (3.18.2).

$$\ddot{\mathbf{Y}}_{-i} = \alpha_{|i}\ddot{\mathbf{Y}}_i + (\ddot{Y}_i)^{\beta_{|i}}\Re_{|i}, \tag{3.18.2}$$

where $\Re_{|i}$ is a vector of residuals and $(p-1)$ dimensional parameter vectors $\alpha_{|i}$ and $\beta_{|i}$ satisfying $(\alpha_{|i}, \beta_{|i}) \in [-1, 1]^{p-1} \times (-\infty, 1)^{p-1}$. Here, $\alpha_{j|i}$, $\alpha_{|i}$ related with $\ddot{\mathbf{Y}}_i, (i \in 1, ..., p, j \neq i)$, then $-1 \leq \alpha_{j|i} < 0$ and $0 < \alpha_{j|i} \leq 1$ correspond respectively to negative and positive association between $\ddot{\mathbf{Y}}_j$ and $\ddot{\mathbf{Y}}_i$'s large values (Southworth et al., 2016).

3.19 Threshold selection

In order to select a sufficiently high threshold for the univariate risk modelling, cautious trade off between bias and variance must be ensued. This is necessary to avoid

having too high threshold with few realizations with which to make inferences (Ferro, 2003), and which can also result in increase of the parameter estimate's variance because of the reduced sample (Hu and Scarrott, 2018), or too low threshold to avoid bias where non-extreme or central observations are selected in place of extreme ones. In practice, the threshold is required to be suitably high to ensure a reliable asymptotic GPD approximation, hence reducing the bias (Scarrott and MacDonald, 2012). This study will use two threshold selection approaches described in the following sections.

3.19.1 Extreme value mixture models

The mixture models approach is built to provide an objective estimate of a suitable threshold with uncertainty quantification (Scarrott and MacDonald, 2012). The approach is applied in this study because it is believed that the traditional fixed threshold approach is subjective, and does not account for the uncertainty involved in the choice of a threshold and in the resultant shape parameter estimates. The threshold is treated, by most mixture models, as a parameter that can be estimated with the use of standard inference schemes, hence they can (potentially) account for the related uncertainty on tail inferences (Hu and Scarrott, 2018).

The mixture model is a combination of a bulk model under the threshold and GPD above the threshold. That is, the model operates by dividing the distribution into two parts: the bulk and the tail. The bulk distribution contains high density non-extreme observations with low information about the tail of the distribution. The tail fraction on the other hand contains low density observations with high (asymptotic) information (Scarrott and MacDonald, 2012). The Weibull, gamma and normal are some of the distributions used for the bulk model.

Extreme value mixture models are implemented in the literature under the coverage of a full range of parametric, semi-parametric and non-parametric approaches

for the bulk component (Hu and Scarrott, 2018). To obtain a suitable threshold selection for this study, the mixture models' approach will be narrowed down to the non-parametric extremal mixture models of MacDonald et al. (2011). This mixture models is the Kernel GPD model with tail modelling that follows a GPD and the bulk model under the threshold is the standard kernel density estimator.

The non-parametric method is given preference over the usual parametric bulk approach because it is more robust to bulk model than the parametric technique (Yang, 2013). Furthermore, if the population distribution is unknown, which is more usually likely the situation in financial returns modelling, the non-parametric extreme value mixture models will provide the best tail estimator (than the parametric and semi-parametric), but they add to the computational complexity however, and over-fitting has to be carefully avoided (Hu and Scarrott, 2018). Hu and Scarrott, 2018 further indicated that flexible extreme value mixture models apply non-parametric density estimators beneath the threshold, following MacDonald et al. (2013); MacDonald et al. (2011); Tancredi et al. (2006). The standard Kernel GPD model' distribution function is given as:

Bulk model based tail fraction approach:

$$F(x|X,\gamma,u,\sigma_u,\xi,\vartheta_u) = \begin{cases} H(x|X,\gamma) & x \leq u, \\ (1-\vartheta_u) + \vartheta_u \times \ddot{G}(x|u,\sigma_u,\xi) & x > u, \end{cases} \tag{3.19.1}$$

where $\vartheta_u = 1 - H(u|X,\gamma)$.

Parameterised tail fraction approach:

$$F(x|X,\gamma,u,\sigma_u,\xi,\vartheta_u) = \begin{cases} (1-\vartheta_u)\frac{H(x|X,\gamma)}{H(u|X,\gamma)} & x \leq u, \\ (1-\vartheta_u) + \vartheta_u \times \ddot{G}(x|u,\sigma_u,\xi) & x > u, \end{cases} \tag{3.19.2}$$

where $H(x|X,\gamma)$ signifies the kernel density estimator's distribution function with parameter γ. The GPD's distribution function is $\ddot{G}(x|u,\sigma_u,\xi)$ and ϑ_u denotes the bulk

model-based tail fraction. The u, σ_u, and ξ represent the threshold, scale parameter and shape parameter respectively.

To estimate the tail fraction, the bulk model-based tail fraction benefits from borrowing information from the generally ample bulk data. The main challenge with this bulk model however, is that it exposes estimation of the tail to the bulk model's misspecification (Hu and Scarrott, 2018). The parameterised tail fraction approach was introduced by MacDonald et al. (2011) with an extra parameter (ϑ_u) for the tail fraction, and it can reduce the effect of the misspecification of the bulk model on the tail estimates. Furthermore, the bulk model-based tail fraction is included in the parameterised tail fraction approach as a special case, where $\vartheta_u = 1 - H(u|X, \gamma)$, and it should be clear that either of the two specifications gives a proper density (Hu and Scarrott, 2018).

3.19.2 Threshold stability plot

Threshold stability plot is a fixed threshold approach type that will be used (in this study) as a support to verify the outcome of the mixture models. This threshold selection method provides the conditions for the lowest threshold selection over which extremal index estimates are nearly constant (Heffernan and Southworth, 2013).

The GPD displays threshold stability properties (Hu and Scarrott, 2013), where parameter estimates of ξ (i.e. the shape) and σ (scale) are plotted against several threshold values u, to create a threshold stability plot. This threshold stability is one of various properties of GPD that have conventionally been applied in the fixed threshold approach to choose a suitable threshold (Hu and Scarrott, 2018). The plot is used in Coles (2001) as a diagnostic for threshold selection.

The plot is underpinned with the concept that if the excesses (or exceedances) of a high threshold u_o follow a GPD with parameters ξ and σ_{u_o}, then the GPD is valid for exceedances over all thresholds $u > u_o$, with shape and scale parameters ξ_u and σ_u

respectively. The two distributions in this situation have identical shape parameters ($\xi_u = \xi$) (Coles, 2001), but the scale parameter is defined as

$$\sigma_u = \sigma_{u_o} + \xi(u - u_o). \tag{3.19.3}$$

That is, the shape parameter is invariant to threshold, but the scale parameter is a linear function of the differences in the threshold (Hu and Scarrott, 2018). The difficulty with the scale parameter σ_u in equation (3.19.3) is that it changes with u except when $\xi = 0$, but this can be alleviated when the generalised Pareto scale parameter is reparametrised (Coles, 2001) as

$$\sigma^* = \sigma_u - \xi u \tag{3.19.4}$$

This new parametrised scale parameter is constant with regards to u by virtue of equation (3.19.3) (Coles, 2001; Bommier, 2014).

The locus of points is used to define the plot as

$$\{(u, \sigma^*);\ u < x_{\max}\} \text{ and } \{(u, \xi_u);\ u < x_{\max}\}, \tag{3.19.5}$$

where the maximum of the observations is denoted by $x_{\max}$ (Bommier, 2014).

Accordingly, estimates of both ξ and σ^* need to be constant above u_o, if u_o is a valid threshold for the GPD asymptotic approximation of the exceedances (Coles, 2001; Bommier, 2014). However, these estimates of the quantities (ξ and σ^*) will not be precisely constant due to sampling variability, but stability should be seen after their sampling errors are allowed (Coles, 2001; Northrop and Coleman, 2014). The argument behind the stability plot therefore recommends that both the ξ and σ^* should be plotted against (a range of possible thresholds, Hu and Scarrott, 2018) u, together with each quantity's confidence interval, and the threshold u_o selected as the lowest value of u for which ξ and σ^* remain near-constant (Coles, 2001).

3.19.3 Estimation of parameters

Various techniques can be used to estimate the parameters of the GPD fitted to threshold exceedances. The techniques include among others moment-based methods, graphical methods based on probability plots' versions, the Bayesian method, and the MLE. Each method has its merits and drawbacks, but the MLE is known to be adaptable to complex model-building with outstanding utility and that makes it particularly attractive (Coles, 2001). Further evidence in the literature shows that MLE gives good estimates when the shape parameter $\xi > -1/2$, this makes the technique more appropriate for estimation of financial return data with positive tail index of $\xi > 0$ (Bensalah, 2000). However, there are theoretical limitations associated with the use of the likelihood methods for generalised extreme value modelling. These limitations are potential difficulties known as the regularity conditions that are needed to validate the usual asymptotic properties connected with the maximum likelihood estimator. Smith (1985) summarised the conditions as follows:

- The maximum likelihood estimators are not likely to be obtainable when $\xi < -1$;
- Maximum likelihood estimators do not have the standard asymptotic properties when $-1 < \xi < -1/2$ but they are generally obtainable;
- The maximum likelihood estimators are regular in a way of exhibiting the usual asymptotic properties when $\xi > -1/2$.

The Bayesian estimators can be used in preference of the maximum likelihood estimators when the regularity conditions are violated.

Given a sample $x = x_1, ..., x_m$ with m number of observations, the log-likelihood function $l(\xi, \sigma|x)$ of the parameter values associated with the GPD can be expressed as the logarithm of the joint probability density as follows:

$$l(\xi,\sigma|x) = \begin{cases} -m\log\sigma - \left(\frac{1}{\xi}+1\right)\sum_{i=1}^{m}\log\left(1+\frac{\xi}{\sigma}x_i\right) & \text{if } \xi \neq 0 \\ -m\log\sigma - \frac{1}{\sigma}\sum_{i=1}^{m}x_i & \text{if } \xi = 0. \end{cases} \tag{3.19.6}$$

With $\xi \neq 0$, the log-likelihood is conditional on $(1+\sigma^{-1}\xi y_i) > 0$ for $i = 1, ..., n$; otherwise $l(\xi,\sigma) = -\infty$. Numerical techniques of the log-likelihood are required because analytical maximization is not possible. Furthermore, care must be taken to avoid numerical instabilities when $\xi \approx 0$ in equation (3.19.6) (where $\xi \neq 0$), and to also ensure that the algorithm does not fail as a result of evaluation outside of the acceptable parameter space (Coles, 2001).

The maximum likelihood estimators will be used in computing the scale ($\hat{\sigma}$) and shape ($\hat{\xi}$) parameter estimates that maximise the log-likelihood function of all the sampled observations above the threshold u. Standard likelihood theory is used to obtain (approximate) standard errors and confidence intervals for the GPD (Coles, 2001). The value-outcome of these estimated parameters will give clues on the heaviness of the tails of the returns distribution.

3.20 Extremal index and declustering

The process of modelling extreme observations using a threshold exceedance model generally faces the problems of dependence that usually exists because of short-term clustering of exceedances (Coles, 2001). The EVT approach fits a Poisson process using the point process model and the GPD (via the CEV model) to "independent" excesses (or exceedances) above some appropriately high threshold. It is empirically believed that threshold excesses do not essentially exist separately, but are clustered together most of the time (Fukutome et al., 2015). Hence, to ensure that the exceedances are independent, the time series undergoes the process of declustering,

where the dependent observations are filtered to obtain a set of threshold exceedances that are approximately independent (Coles, 2001; Smith, 1989).

Declustering is a process that is conditional on the strength of dependence between extremes of a sequence that is stationary. This strength of dependence can be summarised by a constant θ called the extremal index as described in Theorem (3.20.1) (Ferro, 2003). In the limit of large sample size n and for large enough thresholds, Ferro and Segers (2003) showed that the inter-exceedance times converge to a mixture distribution having the extremal index θ as the parameter (Fukutome et al., 2015). That is, the extremal index θ is the parameter for the asymptotic distribution of inter-exceedances times. For a stationary series, extremal index θ (loosely defined in equation 3.20.1) is the tendency of the process to experience clustering at extreme levels (Coles, 2001).

$$\theta = (\text{limiting mean cluster size})^{-1}, \tag{3.20.1}$$

where limiting indicates clusters of excesses of increasingly high thresholds (Coles, 2001).

For some definitions, the inter-exceedance times are the time intervals between consecutive exceedances, while the time intervals separating clusters of exceedances are called inter-cluster times, and the intervals between exceedances within a cluster are known as intra-cluster times (Fukutome et al., 2015). The inter-cluster times converge to an exponential distribution, occurring with probability θ, and the mean of the exponential distribution (called the mean cluster size or mean number of exceedances in a cluster) is $1/\theta$, while the intra-cluster times tend to 0 and exist with a probability of $(1-\theta)$. That is, it is required in theory that the intra-cluster times tend to zero, and the inter-cluster times must follow an exponential distribution (Fukutome et al., 2015).

In connection with equation (3.20.1), extremal index is the reciprocal of the limiting mean cluster size or alternatively, the reciprocal of the extremal index θ is the limit of the mean of the cluster-size distribution (Leadbetter, 1983; Ferro, 2003) expressed as

$$\theta^{-1} = \lim_{n\to\infty} \sum_{j=1}^{\infty} j\varrho_n(j), \tag{3.20.2}$$

for the cluster-size distribution defined as

$$\varrho_n(j) = P\{N_{r_n}(u_n) = j | N_{r_n}(u_n) > 0\} \quad \text{for} \quad \text{integer} \;\; \text{j} \geq 1, \tag{3.20.3}$$

where $N_r(u) = \sum_{i=1}^{r} I(X_i > u)$ denotes the number of exceedances in a block of length r and the integer part of x be denoted by $\lfloor x \rfloor$ (Ferro, 2003). An average of $1/\theta$ exceedances must occur in each cluster, since the expected number of exceedances is not affected by the strength of the dependence (Ferro, 2003).

A cluster occurs when exceedances are separated by less than a fixed interval known as the run parameter. Clusters can be identified or defined in a process by using blocks and runs declustering schemes (Leadbetter et al., 1989).

Runs declustering is based on the assumption that excesses are of the same clusters if they are disconnected or separated by lesser than a certain number, called the run length r, of non-extreme observations (Ferro and Segers, 2003; Ferro, 2003). The run length r is the interval length of clusters or the minimum gaps between the clusters (Coles, 2001). Fukutome et al. (2015) described the runs declustering on the assumption that the independent excesses are separated by a minimum number (or distant) called the run parameter of non-exceedances. Two exceedances can be thought of as forming a cluster if they are disconnected by a number of non-exceedances lesser than the run parameter.

Blocks declustering on the other hand entails selecting a block length b and dividing the sequence $\{1, ..., n\}$ into $p = \lfloor n/b \rfloor$ blocks, $\{(i-1)b + 1, ..., ib\}$ for $1 \leq i \leq p$.

In this scheme, any exceedances located within the same block are stipulated as belonging to the same cluster (Ferro, 2003).

Given that $M_{i,j} = \max\{X_{i+1}, ..., X_j\}$, the extremal index's estimators that correspond to the runs $\bar{\theta}_n^R$ and blocks $\bar{\theta}_n^B$ declustering schemes (described in Smith and Weissman, 1994) are stated in equations (3.20.4) and (3.20.5) respectively (Ferro, 2003).

$$\bar{\theta}_n^R(u;r) = \frac{\sum_{i=1}^{n} I(X_i > u, M_{i,i+r} \leq u)}{\sum_{i=1}^{n} I(X_i > u)}, \tag{3.20.4}$$

$$\bar{\theta}_n^B(u;b) = \frac{\sum_{i=1}^{p} I\{M_{(i-1)b,ib} > u\}}{\sum_{i=1}^{pb} I(X_i > u)}, \tag{3.20.5}$$

Consideration is given to extremal index because neglecting it can lead to the risk of underestimating the marginal quantiles or overestimating the return level (Beirlant et al., 2004), which may eventually lead to wrong statistical inference. The value of extremal index required for declustering can be conveniently assumed in the range $0 < \theta \leq 1$ (Beirlant et al., 2004). An extremal index of $\theta = 1$ signifies independence in a process, while a situation with $\theta < 1$ denotes dependence, which suggests that the process will need to be declustered at a chosen value of run length r. The strength of dependence increases as θ tends to 0. After declustering the clustered exceedances above the selected threshold, it is advisable to fit the EVT models on the largest exceedances (called the cluster maxima) within the clusters, and not on all the exceedances when modelling peaks or rare events (Davison and Smith 1990; Beirlant et al., 2004; Fukutome et al., 2015).

Ferro (2003) described extremal index θ through analogous limiting distribution of the block maxima (in Theorem 3.20.1). The description is based on the restriction $C(u_n)$ (on serial dependence at only extreme levels) as detailed in Theorem (3.20.2), and it follows the mixing condition presented by Leadbetter (1983).

Theorem 3.20.1. *Let $\{\hat{X}_i\}_{i\geq 1}$ be the independent random variables' sequence whose marginal distribution is F and make $\hat{M}_n = max\{\hat{X}_1, ..., \hat{X}_n\}$. If there occur sequences of (normalizing) constants $a_n > 0$ and b_n such that*

$$P\left(\frac{\hat{M}_n - b_n}{a_n} \leq x\right) \xrightarrow{w} G(x) \quad as\ n \to \infty$$

for a non-degenerate distribution function G (known as the generalised extreme value distribution function), if condition $C(u_n)$ (in Theorem 3.20.2) holds with $u_n = a_n x + b_n$ for each x such that $G(x) > 0$, and if $P\{(M_n - b_n)/a_n \leq x\}$ can converge for some x, then

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \xrightarrow{w} G^{\theta}(x) \quad as\ n \to \infty$$

for some constant $\theta \in [0, 1]$, where θ is the extremal index, which can be estimated using the two declustering schemes in equations (3.20.4) *and* (3.20.5).

Theorem 3.20.2. *Define $M(I) = max\{X_i : i \in I\}$, and for thresholds sequence u_n, make $I_{j,p}(u_n) = \{\{M(I) \leq u_n\} : I \subseteq \{j, ..., p\}\}$ the set of every intersection of the events $\{X_i \leq u_n\}, j \leq i \leq p$.*

Condition *$C(u_n)$. For all $A \in I_{1,p}(u_n)$, $B \in I_{p+l,n}(u_n)$ and $1 \leq p \leq n - l$,*

$$|P(A \cap B) - P(A)P(B)| \leq \phi(n, l),$$

where $\phi(n, l_n) \to 0$ as $n \to \infty$ for some $l_n = o(n)$. It is ensured by condition $C(u_n)$ that any two events of the type $\{M(I_1) \leq u_n\}$ and $\{M(I_2) \leq u_n\}$ can become nearly independent with increase in n when the index sets $I_i \subset \{1, ..., n\}$ are separated by $l_n = o(n)$, which is a relatively short length interval. Thus, the long-range dependence between such events are limited by $C(u_n)$ (Ferro, 2003).

If a stationary sequence of random variables $X_{ii\geq 1}$ with F as the marginal distribution function has extremal index θ, then

$$\bar{F}(u)S(u) \xrightarrow{\mathrm{d}} S_\theta \quad \text{as } u \to \omega, \tag{3.20.6}$$

where $\bar{F} = 1 - F$ is the survival function of F, and F is the marginal distribution function of a sample $\{X_i\}_{i=1}^n$ from a sequence of random variables that are stationary. The ω is the upper end-point expressed as $\omega = \sup\{x : F(x) < 1\}$, S_u is the inter-exceedance time (in equation 3.20.7) by the process $\{X_i\}_{i>1}$ with distribution defined in equation (3.20.8).

$$S(u) \stackrel{\mathrm{d}}{=} \min\{n \geq 1 : X_{n+1} > u\} \quad \text{given } X_1 > u \tag{3.20.7}$$

$$P\{S(u) > n\} = P(M_{1,n+1} \leq u | X_1 > u) \text{ for } \quad n \geq 1 \tag{3.20.8}$$

Equation (3.20.6) denotes that $\bar{F}(u)S(u)$ is convergent in distribution ($\xrightarrow{\mathrm{d}}$) to S_θ, where S_θ is a random variable distributed in accordance with the mixture distribution

$$(1 - \theta)\varsigma_0 + \theta\mu_\theta, \tag{3.20.9}$$

where μ_θ represents an exponential distribution with mean θ^{-1} and ς_0 is the degenerate probability distribution at 0. Hence, the extremal index θ is both the reciprocal of the mean of the non-zero interexceedance times, and the proportion of non-zero interexceedance times (Ferro, 2003; Ferro and Segers, 2003). These characteristics can be seen in the (Ferro, 2003) quantile-quantile plot of Figure 3.1 with extremal index $\theta = 0.5$.

The plot (in Figure 3.1) shows the "normalised inter-exceedance times" on the $y-$axis plotted against the "standard exponential quantiles" on the $x-$axis. The mixture distribution in equation (3.20.9) determines that plots of this type should be piecewise-linear with a breakpoint at the $(1 - \theta)$-quantile, $-\log\theta$. The sloping line has gradient θ^{-1}, and the vertical line denotes the $(1 - \theta)$-quantile. Furthermore,

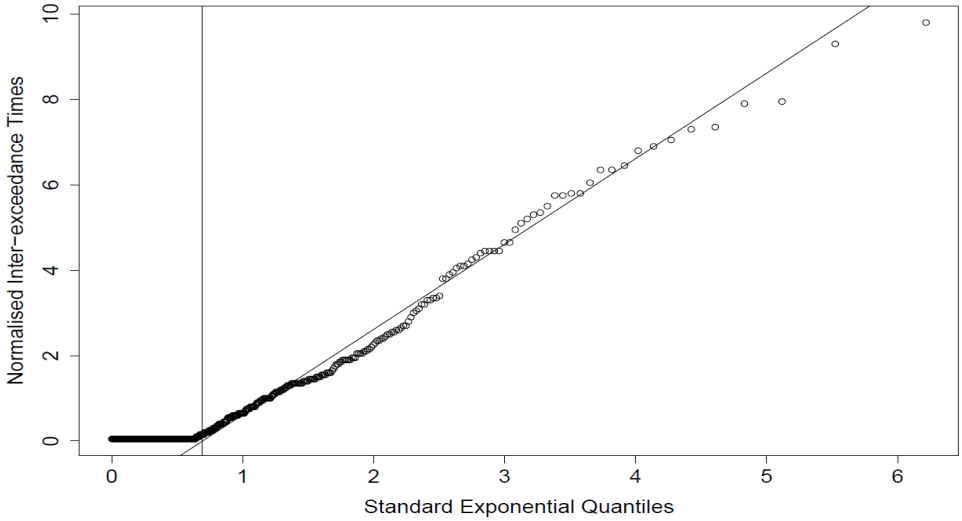


Figure 3.1: Quantile-quantile plot of normalised inter-exceedance times against standard exponential quantiles (Ferro, 2003).

the left segment of the plots should be constant with value 0, and the right segment should have gradient θ^{-1} (Ferro, 2003). It is shown from the work of Ferro 2003 as displayed in Figure 3.1 that an extremal index $\theta = 0.5$ approximately aligns to this piecewise-linearity requirement. In this study, declustering of each of the five BRICS returns exceedances will be done following the Ferro's (2003) (approximate) extremal index θ of 0.5.

3.21 Return level estimation

Estimates of return levels are used to measure the next extreme event of an extreme variable. The return level is the level that an event is expected to be exceeded on average once in the next "m" number of years (Coles, 2001). This can be described

as the g-observation return level ($\ddot{R}_g$) stated as:

$$\ddot{R}_g = \begin{cases} u + \frac{\sigma}{\xi}\left[(g\eta_u)^{\xi} - 1\right], & \text{for } \xi \neq 0, \\ u + \sigma\log(g\eta_u), & \text{for } \xi = 0, \end{cases} \tag{3.21.1}$$

provided g is large enough so that $\ddot{R}_g > u$ (Coles, 2001). Suppose the level is described on an annual basis, the return level can be restated as the level expected to be surpassed once every "m" years. The annual scale expression is like the g-observation return level if there are m_x observations per year, so $g = m \times m_x$. The m-year return level is therefore,

$$\ddot{R}_m = \begin{cases} u + \frac{\sigma}{\xi}\left[(mm_x\eta_u)^{\xi} - 1\right], & \text{for } \xi \neq 0, \\ u + \sigma\log(mm_x\eta_u), & \text{for } \xi = 0. \end{cases} \tag{3.21.2}$$

The return level estimate $\ddot{R}_g$ can be obtained by substituting the maximum likelihood parameter estimates ($\hat{\sigma}$ and $\hat{\xi}$) into the equation (in 3.21.2). We can obtain the probability that an individual observation can exceed the threshold as

$$\hat{\eta}_u = \frac{w}{n} \tag{3.21.3}$$

This is the sample proportion of observations surpassing u (Coles, 2001), where w and n denote the number of observations exceeding u and the total set of observations respectively.

The delta method (in equation 3.21.4) or profile likelihood intervals approach can be used to obtain the standard error needed to compute the return-level's confidence interval (Coles, 2001).

$$\text{Var}(\hat{\ddot{R}}_g) \approx \nabla\ddot{R}_g^T\ddot{V}_c\nabla\ddot{R}_g, \tag{3.21.4}$$

where

$$\nabla \ddot{R}_g^T = \left[\frac{\partial \ddot{R}_g}{\partial \eta_u}, \frac{\partial \ddot{R}_g}{\partial \sigma}, \frac{\partial \ddot{R}_g}{\partial \xi}\right]$$
$$= [\sigma g^{\xi} \eta_u^{\xi-1}, \xi^{-1}\{(g\eta_u)^{\xi} - 1\}, -\sigma\xi^{-2}\{(g\eta_u)^{\xi} - 1\} + \sigma\xi^{-1}(n\eta_u)^{\xi}\log(n\eta_u)], \tag{3.21.5}$$

evaluated at $(\hat{\eta_u}, \hat{\sigma}, \hat{\xi})$. The complete variance-covariance matrix, $\ddot{V}_c$, for $(\hat{\eta_u}, \hat{\sigma}, \hat{\xi})$ is approximated as

$$\ddot{V}_c = \begin{bmatrix} \hat{\eta}_u(1-\hat{\eta}_u)/m & 0 & 0 \\ 0 & \ddot{v}_{1,1} & \ddot{v}_{1,2} \\ 0 & \ddot{v}_{2,1} & \ddot{v}_{2,2} \end{bmatrix}, \tag{3.21.6}$$

where $\text{Var}(\hat{\eta}_u) \approx \hat{\eta}_u(1-\hat{\eta}_u)/m$ from the standard properties of binomial distribution, and $\ddot{v}_{i,j}$ are the (i,j) variance-covariance matrix term of $\hat{\sigma}$ and $\hat{\xi}$ (Coles, 2001).

3.22 Poisson point process

3.22.1 Overview

Like the GPD approximation to excesses above high thresholds, the point process can equally be used to describe exceedances over a sufficiently high threshold. A region is described above the selected threshold such that points in the region signify the extreme events or risk to model. The point process approach incorporates other EVT models including the r largest order statistics, the block maxima and the threshold excess models. The development of these EVT models is a result of the representation of the point process, which forms a good reason for considering the approach (Coles, 2001).

The technique of a point process on a set $\mathbb{K}$ is a stochastic approach for the existence and position of events represented as points that are randomly distributed

in space. This process can be used to describe the behaviour of extreme events concentration at the tails of the markets, where $\mathbb{K}$ denotes a time period.

For each of the selected markets (random variables), a set of non-negative integer values *N(K)* can be defined such that $K \subset \mathbb{K}$. $N(K)$ is the number of points (events) in each set K and it signifies the number of events that can occur within a specified period of time.

Furthermore, let the intensity measure of the process be defined as

$$\Phi(K) = E\{N(K)\}. \tag{3.22.1}$$

This is the average number of events (points) in any given subset $K \subset \mathbb{K}$. Also, the derivative function of the intensity measure, with the assumption that $K = [k_1, x_1] \times \ldots \times [k_p, x_p] \subset R^p$, defines the intensity (density) function of the process as

$$\lambda(x) = \frac{\partial \Phi(K)}{\partial x_1 \cdots \partial x_p}. \tag{3.22.2}$$

That is, the intensity function is the derivative function of the intensity measure of the process. The point process models can be applied statistically by estimating the process using a set of observed events (points) $x_1, \ldots, x_n$ in an interval or a specified region $\mathbb{K}$. Hence, x is an observed point event or simply, events from $N(K)$, such that $N(K)$ is the number of points in the set K, where K denotes a time-period. In this study on extremal dependence modelling, x is an extreme event or realisation like a stock market crash etc.

The one-dimensional homogeneous Poisson process is the canonical point process and it is a process on $\mathbb{K} \subset \mathbb{R}$ with a $\lambda > 0$ parameter (Coles, 2001), where it satisfies the following:

1. $N(K) \sim \text{Poi}(\lambda(h_2 - h_1))\ \forall K = [h_1, h_2] \subset \mathbb{K}$. Here, the number of events (points) in a specified interval $N(K)$ follows a Poisson distribution whose mean λ is proportional to the length of the interval $(h_2 - h_1)$.

2. The number of point events, say $N(K_1)$ and $N(K_2)$ taking place in different non-overlapping or disjoint subsets (intervals) K_1 and K_2 are mutually independent (see Coles, 2001; Beirlant et al., 2004; Resnick, 1987).

This homogeneous Poisson process of parameter λ is a suitable model for points occurring randomly in time at a uniform of λ per unit interval of time. The intensity measure and intensity function of this are $\Phi([h_1, h_2]) = \lambda(h_2 - h_1)$ and $\lambda(h) = \lambda$ respectively.

The homogeneous Poisson process can be extended to the non-homogeneous Poisson process if the uniform time rate is varied at a variable rate $\lambda(h)$ for points that occur randomly in time (Coles 2001). This non-homogeneous process also possesses the same independent counts properties but with the adjusted property that for all $K = [h_1, h_2] \subset \mathbb{K}$, the number of points in the interval K follows a Poisson distribution with intensity measure, i.e. $N(K) \sim \text{Poi}(\Phi(K))$,

for

$$\Phi(K) = \int_{h_1}^{h_2} \lambda(h)dh.$$

For multivariate dependence modelling at the tails, the non-homogenous Poisson process can be generalised to the p-dimensional space for randomly existing points characterization. For the bivariate case, a point process on $K \subset \mathbb{R}^2$ is known to be a 2-dimensional non-homogeneous Poisson process with intensity function $\lambda(.)$ if the property of independent counts on separate time intervals is satisfied. Hence for all $K \subset \mathbb{K}$

$$N(K) \sim \text{Poi}(\Phi(K)),$$

where

$$\Phi(K) = \int_K \lambda(x)dx.$$

The Poisson process has a fundamental property such that the events (points) happen independently of one another. The existence of a point at $x \in K$ location does not influence the existence of other points in a region of x, or elsewhere (Coles, 2001).

3.22.2 The univariate case

The concept of convergence of random variables is required in order to apply point processes representation for extreme values modelling.

Theorem 3.22.1. *Assume a series of independent and identically distributed (i.i.d.) random variables with a common distribution function F is represented by $X_1, X_2, \ldots, X_n$, where $\ddot{M}_n = max\{X_1, \ldots, X_n\}$. The distribution of the normalised maxima, with sequences of constants $\{b_n\}$ and $\{a_n > 0\}$ can be reasonably approximated by a generalised extreme value (GEV) distribution as*

$$Pr\{(\ddot{M}_n - b_n)/a_n \leq c\} \to \ddot{G}(c),$$

with

$$\ddot{G}(c) = exp\left\{-\left[1 + \xi\left(\frac{c-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\}, \tag{3.22.3}$$

for c_+ and c_- denoting the upper and lower endpoints of $\ddot{G}$ in that order, and ξ, μ and σ (for $\sigma > 0$) are the shape, location and scale parameters respectively.

Following this, a sequence of point processes (N_n) is defined as

$$N_n = \{(i/(n+1), (X_i - b_n)/a_n)\} \quad \text{for} \quad i = 1, \ldots, n. \tag{3.22.4}$$

The first ordinate ensures that the time axis consistently maps to $(0, 1)$ while the second ordinate is scaled to sustain stability in the behaviour pattern of extremes as $n \to \infty$ (Coles, 2001).

Now for a sufficiently large value of the threshold u, consider a region of the form $K = [0,1] \times [u, \infty]$ where the point processes sequence N_n converge in distribution to a non-homogeneous Poisson process N for any $u > c_-$, i.e. $N_n \xrightarrow{\text{d}} N$ as $n \to \infty$.

This limit occurs since the random variables X_i are mutually independent and each of the points in $\ddot{M}_n$ has p probability (in equation 3.22.5) of falling in the K region, hence $N_n(K)$ follows a binomial distribution, i.e. $N_n(K) \sim \text{Bin}(n, p)$.

$$p = \Pr\left\{\frac{(X_i - b_n)}{a_n} > u\right\} \approx \frac{1}{n}\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}} \tag{3.22.5}$$

By the standard approximation of a binomial distribution to a Poisson limit, as $n \to \infty$, the limiting distribution of $N_n(K)$ for any region of the type $K = [h_1, h_2] \times (u, \infty)$, where $[h_1, h_2] \subset [0,1]$ is the Poisson distribution having intensity measure $\Phi(K)$ i.e. $\text{Poi}(\Phi(K))$ with

$$\Phi(K) = n_{\tilde{y}}(h_1 - h_2)\left[1 + \xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}, \tag{3.22.6}$$

where the intensity measure is the average number of points that occur in any given subset $K \subset \mathbb{K}$, $n_{\tilde{y}}$ denotes the number of years of observation, and σ, μ, and ξ are the scale, location and shape parameters respectively. Hence, the Poisson process is a realistic approximation of the point processes for large but finite sample behaviour on the specified region where the threshold is sufficiently large (Coles, 2001).

3.22.3 Estimation of parameters

For consistency and because of its adaptable application to financial data and several models, MLE will be used for the parameter estimation of the point process. The parameter estimates are based on extreme data above a sufficiently high threshold, like the threshold excess approach (Coles, 2001). By maximizing the likelihood, the parameters (ξ, μ, σ) estimation of the limiting intensity function of the point process

can be produced, where the likelihood is given as

$$\begin{aligned} L_K(\mu,\sigma,\xi;x_1,...,x_n) &= \exp\{-\Phi(K)\}\prod_{i=1}^{N(K)}\lambda(h_i,x_i) \\ &\propto \exp\left\{-n_{\tilde{y}}\left[1+\xi\left(\frac{u-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}}\right\} \\ &\propto \prod_{i=1}^{N(K)}\sigma^{-1}\left[1+\xi\left(\frac{x_i-\mu}{\sigma}\right)\right]^{-\frac{1}{\xi}-1} \end{aligned} \tag{3.22.7}$$

The likelihood estimator will be used to find the maximum likelihood estimates, standard errors and the approximate confidence intervals of the model's parameters. For non-homogeneous Poisson process models, maximization of the likelihood generally requires numerical techniques (Coles 2001).

3.23 Return level estimation

The point process approach has both non-stationary and stationary versions for return level estimation. While the latter is easy to compute, the former takes the form (Coles, 2001)

$$1-\frac{1}{g}=\Pr\{\max(X_1,...,X_n)\leq R_g\}\approx\prod_{i=1}^{n}q_i \tag{3.23.1}$$

That is, if R_g signifies the g-year return level, and n is the number of observations in a year, then R_g satisfies equation (3.23.1), where

$$q_i = \begin{cases} 1-m^{-1}[\Omega]^{-1/\xi_i}, & \text{if } [\Omega]>0, \\ 1, & \text{otherwise,} \end{cases} \tag{3.23.2}$$

for $\Omega = 1+\xi_i(R_g-\mu_i)/\sigma_i$ and (μ_i,σ_i,ξ_i) are the point process model's parameters for observation i. With the logarithms taken, this becomes

$$\sum_{i=1}^{m} \log q_i = \log(1 - 1/g). \tag{3.23.3}$$

The result in equation (3.23.3) can easily be solved to obtain the return level estimate (R_g) using standard numerical methods for non-linear equations. However, the standard error for the return level's confidence interval can only be estimated via simulation since neither the delta method nor the profile likelihood can be practically used for this estimation (Coles, 2001).

3.24 Bivariate point process model

Multivariate point processes can be considered as special cases of univariate point processes where a real valued quantity is linked with each point event. A bivariate process of two types of events (e.g., type m and type n) will be the case if the real valued quantity takes only two likely values. The process of one of the events, say, type m event alone is known as a marginal process. A bivariate Poisson process is defined as a bivariate point process of which the marginal processes are Poisson processes (Cox and Lewis, 1972). The Poisson limit is a reasonable approximation to a sequence of point processes on a suitable region (Coles, 2001).

The point process characterization for the bivariate modelling of the extremal dependence can be described as stated in Theorem (3.24.1).

Theorem 3.24.1. *Let $(x_1, y_1), (x_2, y_2)$ ... be a sequence of realizations (or observed values) that are independent bivariate from a distribution having standard Fréchet margins, and satisfying the convergence for componentwise maxima.*

$$Pr\{\ddot{M}^*_{x,n} \leq x, \ddot{M}^*_{y,n} \leq y\} \to G(x, y), \tag{3.24.1}$$

where $\ddot{M}_{x,n}$ *and* $\ddot{M}_{y,n}$ *are the maxima of sequences of two separate independent, univariate random variables* X_i *and* Y_i *with standard Fréchet marginal distributions.*

That is,

$$\ddot{M}_{x,n} = \max\{X_i\} \text{ and } M_{y,n} = \max\{Y_i\}, \quad \text{for } i = 1, ..., n. \tag{3.24.2}$$

Rescaled or standardised as

$$M^*_{x,n} = \max\{X_i\}/n \text{ and } M_{y,n} = \max\{Y_i\}/n, \quad \text{for } i = 1, ..., n, \tag{3.24.3}$$

where G is a distribution function that is non-degenerate, and it takes the form

$$G(x, y) = \exp\{-V(x, y)\}, \ x > 0, \ y > 0 \tag{3.24.4}$$

with $V(x, y)$ as specified in equations (3.24.5) and (3.24.6) respectively (Coles, 2001).

$$V(x, y) = 2\int_0^1 \max\left(\frac{\ell}{x}, \frac{1-\ell}{y}\right) dH(\ell). \tag{3.24.5}$$

H indicates a distribution function on [0, 1], and it satisfies the mean constraint

$$\int_0^1 \ell \ dH(\ell) = 1/2. \tag{3.24.6}$$

Now, let $\{N_n\}$ represent a sequence of point processes defined by

$$N_n = \left\{\left(\frac{x_1}{n}, \frac{y_1}{n}\right), ..., \left(\frac{x_n}{n}, \frac{y_n}{n}\right)\right\}. \tag{3.24.7}$$

Then, as $n \to \infty$, N_n can be reasonably approximated by a non-homogeneous Poisson process (N), as the limit distribution, on $(0, \infty) \times (0, \infty)$, such that

$$N_n \xrightarrow{\text{d}} N, \tag{3.24.8}$$

on a region of the type K in equation (3.24.9), bounded from the origin (0, 0).

$$K = \{(0, \infty) \times (0, \infty)\} \backslash \{(0, x) \times (0, y)\} \tag{3.24.9}$$

The intensity function of the Poisson process (or the limiting process) N is stated in equation (3.24.10) and it indicates that the intensity factorises across angular and radial components, where H determines the angular spread of points in the process.

$$\lambda(\hbar, \ell) = 2\frac{dH(\ell)}{\hbar^2} \tag{3.24.10}$$

for

$$\hbar = x + y \text{ and } \ell = \frac{x}{x + y} \tag{3.24.11}$$

The choice of a sufficiently large threshold with application of the bivariate point process entails the same consideration as that used by the threshold excess model. For both threshold models, the values of the selected thresholds intersect at the same points on the Cartesian x and y axes, and this can be used in comparing the outcomes of the two models. The bivariate point process has an added advantage that it can be transformed to pseudo-polar coordinates from Cartesian, i.e. $(x, y) \rightarrow (\hbar, \ell)$, where the distance from the origin is measured in $\hbar$ units, while ℓ measures the angle on a scale of [0, 1]. The value of $\ell = 0$ corresponds to $x = 0$ axis and $\ell = 1$ corresponds $y = 0$ axis.

Like in the univariate case, all representations of the multivariate types can be obtained as special cases of the representation of the point process. This can be illustrated with derivation of the componentwise block maxima's limit distribution as (Coles, 2001)

$$\Pr\{\ddot{M}^*_{x,n} \leq x, \ddot{M}^*_{y,n} \leq y\} = \Pr\{N_n(K) = 0\} = \exp\{-\Phi(K)\}, \tag{3.24.12}$$

where the point process as defined in equation (3.24.7) is denoted by N_n and K is the region defined in equation (3.24.9). Hence the limit of the point process is stated as

$$\Pr\{\ddot{M}^*_{x,n} \leq x, \ddot{M}^*_{y,n} \leq y\} \to \Pr\{N(K) = 0\} = \exp\{-\Phi(K)\}, \tag{3.24.13}$$

for the intensity measure

$$\begin{aligned}\Phi(K) &= \int_K 2\frac{d\hbar}{\hbar^2} dH(\ell) \\ &= \int_{\ell=0}^{1} \int_{\hbar=\min\{x/\ell, y/(1-\ell)\}}^{\infty} 2\frac{d\hbar}{\hbar^2} dH(\ell) \\ &= 2\int_{\ell=0}^{1} \max\left(\frac{\ell}{x}, \frac{1-\ell}{y}\right) dH(\ell)\end{aligned} \tag{3.24.14}$$

The Poisson limit is a good approximation to the point process and convergence of the points is definite on regions bounded from the origin. For sufficiently large $\hbar$, convergence can be made simple if a region of the type $K = \{(x, y) : x(n^{-1}) + y(n^{-1}) > \hbar_\circ\}$ is chosen, since the intensity measure $\Phi(K)$ is then stated as

$$\Phi(K) = 2\int_K 2\frac{d\hbar}{\ell^2} dH(\ell) = 2\int_{\hbar=\hbar_\circ}^{\infty} \frac{d\hbar}{\hbar^2} \int_{\ell=0}^{1} dH(\ell) = \frac{2}{\hbar_\circ}, \tag{3.24.15}$$

and it is constant with respect to H parameters (Coles, 2001).

3.24.1 Dependence structure

The different dependence structures faced in general datasets modelling can be considered using both symmetric and asymmetric models (Coles and Tawn, 1991; Lipika 2018). Six parametric dependence models associated with the point process bivariate dependence modelling as indicated in the package "evd" (see Stephenson, 2018) are used in this study. The models are the logistic, negative logistic, Husler-Reiss, Bilogistic, negative bilogistic, and Coles-Tawn (or Dirichlet).

a) Logistic model

The (bivariate) logistic model in equation (3.24.16) was introduced by Gumbel (1960) as a distribution function with dependence parameter $\alpha \in (0, 1)$, precisely $0 < \alpha \leq 1$.

$$G(x, y) = \exp\left[-(x^{-1/\alpha} + y^{-1/\alpha})^{\alpha}\right], x > 0, y > 0 \tag{3.24.16}$$

As $\alpha \to 1$ in equation (3.24.16),

$$G(x, y) \to \exp\left\{-(x^{-1} + y^{-1})\right\}, \tag{3.24.17}$$

denotes independence variables, while as $\alpha \to 0$,

$$G(x, y) \to \exp\left\{-\max(x^{-1}, y^{-1})\right\}, \tag{3.24.18}$$

indicates perfectly (or completely) dependence variables (Coles, 2001; Stephenson, 2018). Hence, the logistic family covers all stages of bivariate dependence from perfect dependence to independence. Equation (3.24.16) is gotten by making H in equation (3.24.5) have the density function

$$h(\ell) = \frac{1}{2}(\alpha^{-1} - 1)\{\ell(1 - \ell)\}^{-1-\frac{1}{\alpha}}\{\ell^{-\frac{1}{\alpha}} + (1 - \ell)^{-\frac{1}{\alpha}}\}^{\alpha-2} \tag{3.24.19}$$

on $0 < \ell < 1$.

b) Negative logistic model

Galambos (1975) introduced the negative logistic model where the distribution function with the dependence parameter α is stated as

$$G(x, y) = \exp\left\{-x - y + [x^{-\alpha} + y^{-\alpha}]^{-1/\alpha}\right\} \tag{3.24.20}$$

for $\alpha > 0$. Complete dependence is attained as $\alpha \to \infty$, while independence is obtained as $\alpha \to 0$. The density function (Lipika 2018) obtainable via H is

$$h(\ell) = (1+\alpha)\{\ell(1-\ell)^{-\alpha-2}\}\{\ell^{-\alpha} + (1-\ell)^{-\alpha}\}^{-\frac{1}{\alpha}-2} \tag{3.24.21}$$

c) Husler-Reiss model

The Husler-Reiss model was introduced by Husler and Reiss (1989) with dependence parameter α and distribution function (see Stephenson, 2018) stated as

$$G(x,y) = \exp\left(-x\Phi\{\alpha^{-1} + \frac{1}{2}\alpha[\log(x/y)]\} - y\Phi\{\alpha^{-1} + \frac{1}{2}\alpha[\log(y/x)]\}\right), \tag{3.24.22}$$

where $\alpha > 0$ and $\Phi(\cdot)$ denotes the standard normal distribution function. As α tends to ∞, complete dependence is attained, whereas as α approaches 0, independence is attained (Stephenson, 2018).

d) Bilogistic model

The bilogistic model as derived by Joe et al. (1992) is the logistic model generalised, such that asymmetry is allowed in the dependence structure (Coles, 2001).

$$G(x,y) = \exp\left\{-xp^{1-\alpha} - y(1-p)^{1-\beta}\right\}. \tag{3.24.23}$$

The model is obtained by allowing H have the density function

$$h(\ell) = \frac{1}{2}(1-\alpha)(1-\ell)^{-1}\ell^{-2}(1-u)u^{1-\alpha}\{\alpha(1-u) + \beta u\}^{-1} \tag{3.24.24}$$

on $0 < \ell < 1$, for parameters β: $0 < \beta < 1$ and α: $0 < \alpha < 1$. In addition, $u = u(\ell, \alpha, \beta)$ is the root or solution of the equation

$$(1-\alpha)(1-\ell)(1-u)^{\beta} - (1-\beta)\ell u^{\alpha} = 0. \tag{3.24.25}$$

The logistic model is the equivalent of the bilogistic model when the dependence parameter $\alpha = \beta$ (Coles, 2001; Stephenson, 2018). Furthermore, the degree of asymmetry in the dependence structure can be determined by the value of $\alpha - \beta$ (Coles, 2001). As $\alpha = \beta$ tends to 0, complete dependence is reached, while independence is attained as $\alpha = \beta$ approaches 1 (Stephenson, 2018).

e) Negative bilogistic model

The negative bilogistic model was introduced by Coles and Tawn (1994) with dependence parameters α and β in the distribution function:

$$G(x, y) = \exp\left\{-x - y + xp^{1+\alpha} + y(1-p)^{1+\beta}\right\}, \tag{3.24.26}$$

where $p = p(x, y; \alpha, \beta)$ denotes the root of the equation

$$(1+\alpha)xp^{\alpha} - (1+\beta)y(1-p)^{\beta} = 0, \tag{3.24.27}$$

for $\beta > 0$ and $\alpha > 0$. The negative bilogistic model reduces to the negative logistic model when $\beta = \alpha$ with dependence parameter $1/\alpha = 1/\beta$. As $\alpha = \beta$ approaches ∞, independence is attained, while complete dependence is obtained when the limit $\alpha = \beta$ tends to 0 (Stephenson, 2018). The density function is stated as

$$h(\ell) = -\frac{(1-\alpha)(1-p)p^{1-\alpha}}{(1-\ell)\ell^2\{(1-p)\alpha + p\beta\}}, \tag{3.24.28}$$

where $p = p(\ell; \alpha, \beta)$ is in line with equation (3.24.25) (Lipika, 2018).

f) Coles-Tawn model

Another asymmetric model, introduced by Coles and Tawn (1991) is the Dirichlet or Coles-Tawn (ct) model. The distribution function has dependence parameters $\alpha > 0$ and $\beta > 0$ and it is defined as:

Table 3.1: The strength of dependence.

Model	Independence	Dependence
Logistic	$\alpha \to 1$	$\alpha \to 0$
Negative logistic	$\alpha \to 0$	$\alpha \to \infty$
Husler-Reiss	$\alpha \to 0$	$\alpha \to \infty$
Bilogistic	$\alpha = \beta \to 1$	$\alpha = \beta \to 0$
Negative bilogistic	$\alpha = \beta \to \infty$	$\alpha = \beta \to 0$
Coles-Tawn (or Dirichlet)	$\alpha = \beta \to 0$	$\alpha = \beta \to \infty$

$$G(x,y) = \exp[-x\{1 - \text{Be}(p;\alpha+1,\beta)\} - y\text{Be}(p;\alpha,\beta+1)], \tag{3.24.29}$$

where $\text{Be}(p;\alpha,\beta)$ represents the beta function estimated at p with shape1 and shape2 equal to α and β respectively, and $p = \alpha y/(\alpha y + \beta x)$. As $\alpha = \beta$ approaches ∞, complete dependence is achieved in the limit, while independence in the variables is obtained as $\alpha = \beta$ tends to 0 (Stephenson, 2018).

The density function corresponds to

$$h(\ell) = \frac{\alpha\beta\Gamma(\alpha+\beta+1)(\alpha\ell)^{\alpha-1}\{\beta(1-\ell)\}^{\beta-1}}{2\Gamma(\alpha)\Gamma(\beta)\{\alpha\ell + \beta(1-\ell)\}^{\alpha+\beta+1}} \tag{3.24.30}$$

on $0 < \ell < 1$, with parameters satisfying $\alpha > 0$ and $\beta > 0$. The model is symmetric when $\alpha = \beta$ (Coles, 2001).

3.24.2 Summary of the dependence structure

The strength of asymptotic dependence in each of the six stated models are given in Table 3.1:

3.25 Diagnostics: Model checking

To evaluate the quality of the fitted generalised Pareto model and point process on the threshold exceedances, suitable diagnostic model checking plots can be used.

Classical diagnostic plots include the density plots, probability plots, return level plots and quantile plots. If the Poisson approximation of a point process fit and the GPD fit are reasonable models for modelling peaks or excesses above a threshold u, then both the probability and quantile plots, for example should display points that are approximately linear, i.e. points that lie close to the unit diagonal. Significant departures from linearity will signify a failure in the validity of these models for the data (Coles, 2010).

Also, both the Cramer-von Mises and Anderson-Darling goodness-of-fit tests will be used to assess the validity of the point process and GPD parameter estimates.

Cramér-von Mises and Anderson-Darling statistics belong to the class of quadratic EDF (Empirical Distribution Function) statistics (Stephens, 1986) that is defined as

$$z \int_{-\infty}^{\infty} (F_z(x) - F(x))^2 w_f(x)\, dF(x), \tag{3.25.1}$$

where $F(x)$ denotes the theoretical cdf (cumulative distribution function), with a random sample of size z and $w_f(x)$ represents a weighting function.

3.25.1 Cramér-von Mises (CVM) test

Equation (3.25.1) indicates z times the CVM (Cramér-von Mises Test) statistic for $w_f(x) = 1$. Calculation of the statistic can be done with the use of the sum of squared differences between the EDF (empirical distribution function) and the theoretical CDF as stated in equation (3.25.2) (Anderson and Darling, 1954)

$$\text{CVM} = \frac{1}{12z} + \sum_{j=1}^{z} \left(F(x_j, \theta) - \frac{2j-1}{2z} \right)^2, \tag{3.25.2}$$

with $F(x_j, \theta)$ denoting the theoretical distribution function (Singla et al., 2016).

3.25.2 Anderson-Darling (AD) test

The Anderson-Darling (AD) test was developed as a modification of the CVM test, and it gives more weight to the tail of the distribution (Singla et al., 2016). When taking $w_f(x) = [F(x)(1-F(x))]^{-1}$ in equation (3.25.1), the Anderson-Darling statistic is stated as (Anderson and Darling, 1954):

$$AD = n \int_{-\infty}^{\infty} \frac{(F_z(x) - F(x))^2}{F(x)(1 - F(x))} dF(x). \tag{3.25.3}$$

The formula can be stated alternatively as

$$AD = -z - 2\sum_{j=1}^{z} \left\{ \frac{2j-1}{2z} \ln F(x_j, \theta) + \left(1 - \frac{2_j - 1}{2z}\right) \ln(1 - F(x_j, \theta)) \right\}. \tag{3.25.4}$$

Decision can be made to reject the null hypothesis if the AD test statistic value is greater than the critical point (that is, the p-value). For a fully identified distribution F, the distribution theory of the EDF statistics is well developed and also, tables which give the significant levels are available (Singla et al., 2016 and Bere, 2016). But, for a situation where the scale or location parameter is not stated by the null hypothesis, the exact distributions of EDF statistics are hard to find (Bere, 2016). The Anderson-Darling statistic belongs to a class of quadratic statistics with known asymptotic distributions.

A modified version of the Anderson-Darling statistic (D'Agostino and Stephens, 1986) in which values of the statistic for finite sample size having asymptotic significance points can be compared is given as

$$AD^{**} = AD\left(1.0 + \frac{0.75}{z} + \frac{2.25}{z^2}\right). \tag{3.25.5}$$

The formulae for approximating the p-values that correspond to this modified version are given in Table 4.9 (page 127) of D'Agostino and Stephens (1986) (Bere, 2016).

Chapter 4

Volatility of the BRICS equity markets

4.1 Introduction

This chapter deals with modelling volatility using the GARCH model under the assumption of seven selected error distributions to ascertain the best distribution that will be most suitable for modelling returns volatility of each of the BRICS stock markets when the true distribution is unknown. It is believed that volatility modelling using the GARCH model with the right distribution can yield more economic results (Bollerslev, 1987), hence the chapter experimentally investigates the best innovation among the selected seven, that will be most appropriate for the GARCH model to model the markets' return volatility. The chapter further investigates the level of volatility in each of the BRICS markets and arrange them from the highest to the lowest, since volatility level can define the level of risk in a market (Samiev, 2012).

4.1.1 Overview

The chapter begins with the BRICS markets' data description and the exploratory data analysis to detect the presence of outliers in the data series. After this, explanation of the basic descriptive statistics of the markets' daily price and return series are given. The descriptive statistics focus mainly on measures of central tendency, and measures of dispersion.

Following this, three different tests involving stationarity, serial (lagged or auto) correlation, and the presence of ARCH effects tests are carried out on the BRICS return series. Linear dependence (autocorrelation) and heteroscedasticity are filtered out by fitting ARMA-GARCH models on the return series. ARMA-GARCH models are empirically analysed under seven assumptions of a normal, skew-normal, Student's t, skew-Student's t, GED, skew-GED and GHYP distributions to ascertain the distribution that best describes each of the five BRICS markets. The best error distributions characterizing the markets are selected using four information criteria: AIC (Akaike information criterion), BIC (Bayesian information criterion), SIC (Shibata information criterion) and HQIC (Hannan-Quinn information criterion).

4.2 Data description

The raw price data used for this study includes the daily closing equity indices of the Brazilian, Russian, Indian, Chinese and South African stock markets. The data was obtained from Thomson Reuters Datastream and is for the period 5^{th} January 2010 to 6^{th} August 2018 with 2126 observations. That is, the data for each of the BRICS indices is recorded for 260 days per year, which is 5 trading days in a week.

The BRICS markets' indices are the IBOV (or Bovespa) index of Brazil Sau Paulo stock exchange, the IMOEX (Moscow Exchange) index of Russia, the Indian NIFTY (or NIFTY 50) index is the national stock exchange of India. Next is the SHCOMP

(i.e. the Shanghai Stock Exchange Composite) index of China, and the JALSH (JSE Africa All Share) index of South Africa.

4.3 Missing values

For the period under study, it was observed that six daily closing price values were missing in the South African JALSH index on 24^{th} September 2010, 9^{th} August 2010, 16^{th} June 2010, 27^{th} April 2010, 22^{nd} March 2010 and 27^{th} December 2011. Another missing data was observed in the Russian IMOEX index for 20^{th} February 2016. Since these were single day each, adjustment was made by using the average value of the day prior to and after each missing day as the values of the closing price (see Lipika 2018).

The daily closing prices of financial time series like that of the stock market usually exhibit non-stationarity, and empirical studies on such series have shown that approximation to stationarity can be obtained via log-daily returns. By this, logarithms of ratios of successive realizations are taken, and this can generate reasonable transformation to stationarity (from price to return) as seen in the time series plots in Figures 4.1 and 4.2. The stationary returns in Figures 4.1 and 4.2 display volatility clustering and periods of high volatility. For convenience of presentation, the generated return series is further re-scaled by multiplying by 100 as follows

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) \times 100 \tag{4.3.1}$$

where t is the time period in days, P_t signifies the closing stock price index at time t, and previous day's closing market price index is P_{t-1}. The natural logarithm is denoted by ln, while r_t represents the current returns.

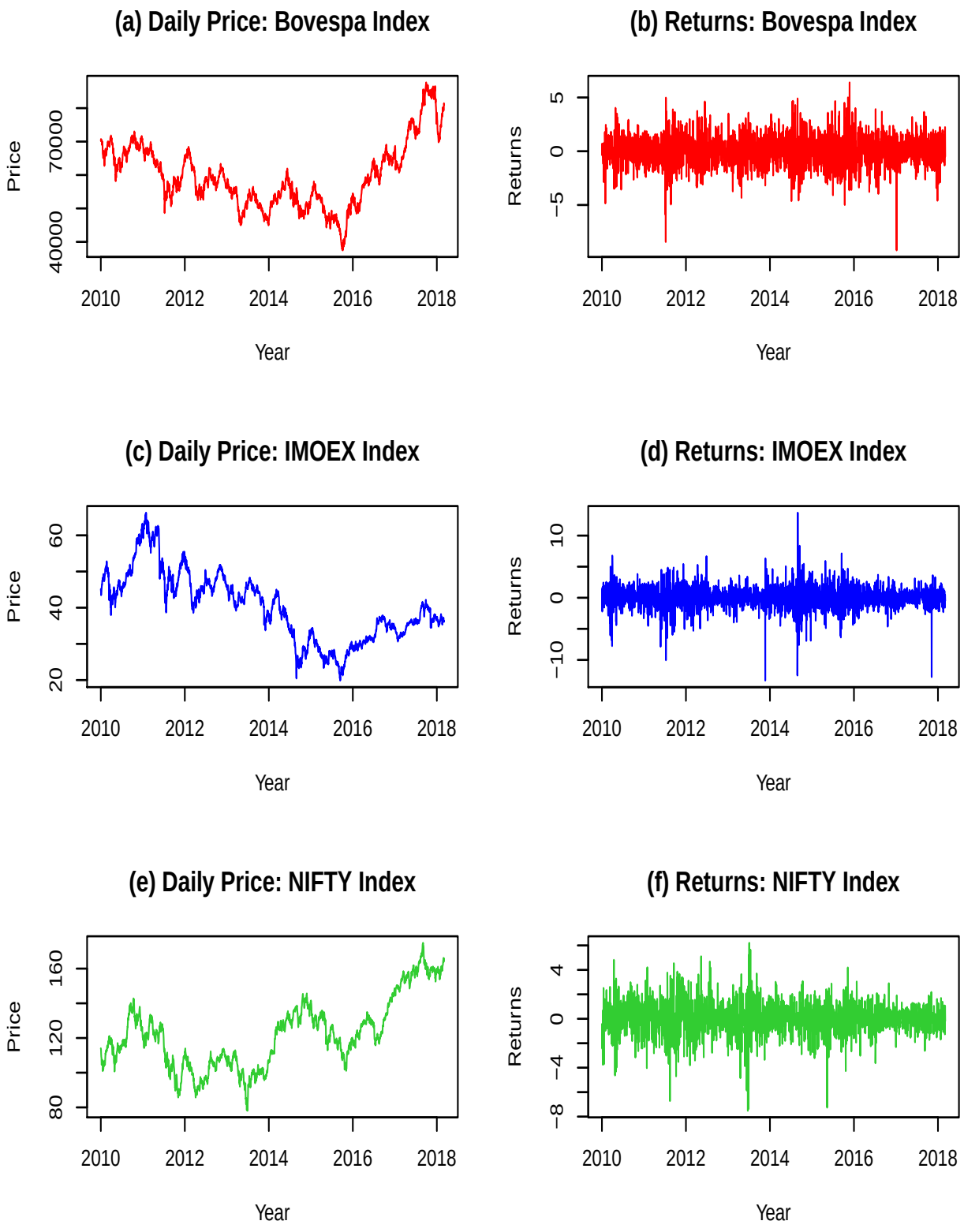


Figure 4.1: Daily price and return for IBOV (Brazil), IMOEX (Russia) and NIFTY (India) indices.

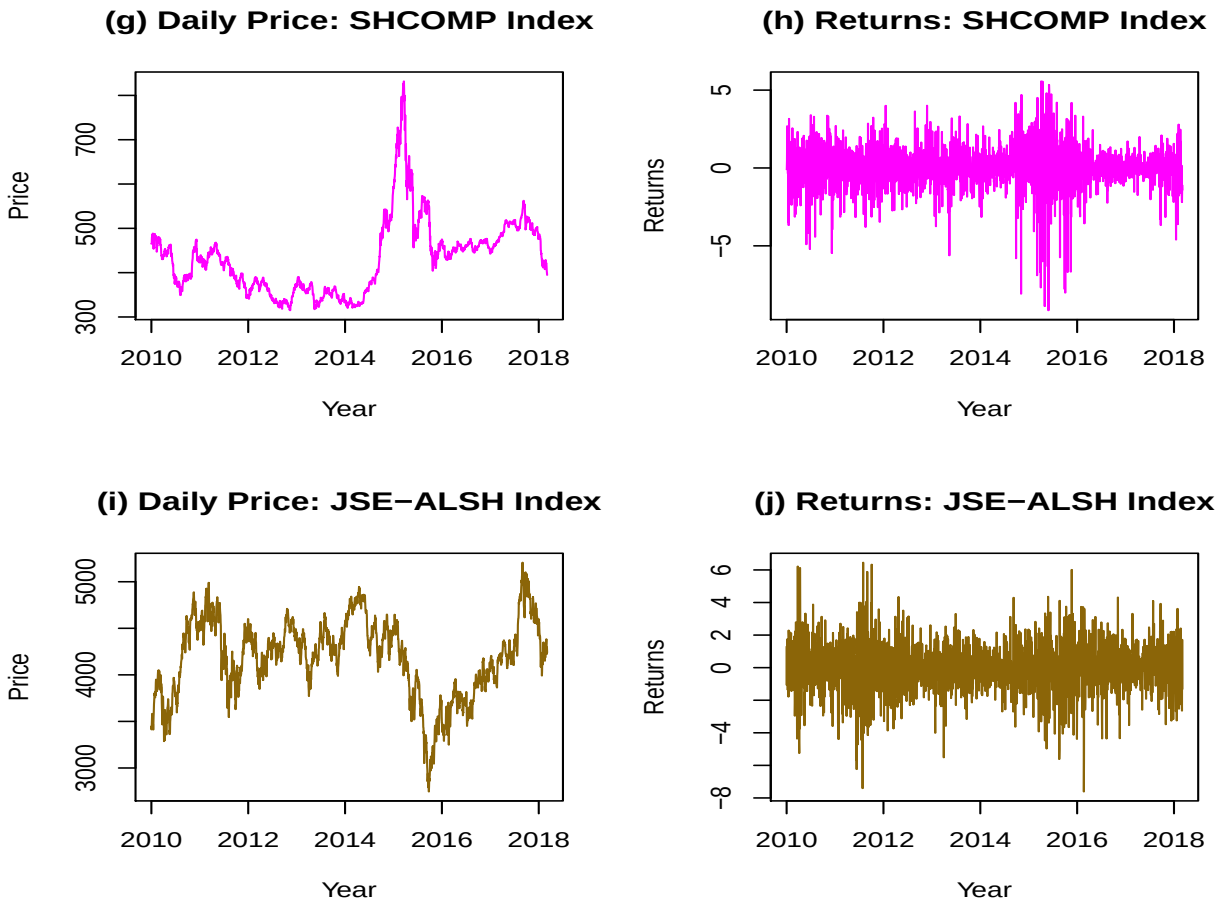


Figure 4.2: Daily price and return for SHCOMP (China) and JALSH (South Africa) indices.

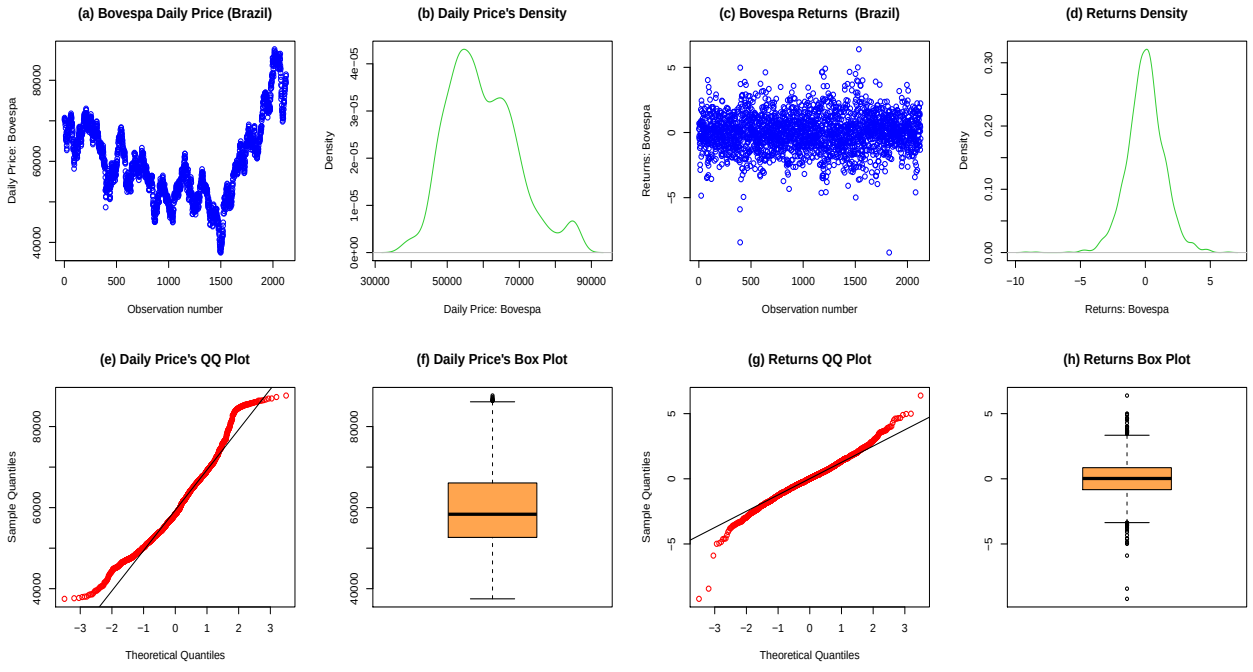


Figure 4.3: Brazil: EDA of price (panels a, b, e, f) and return (panels c, d, g, h).

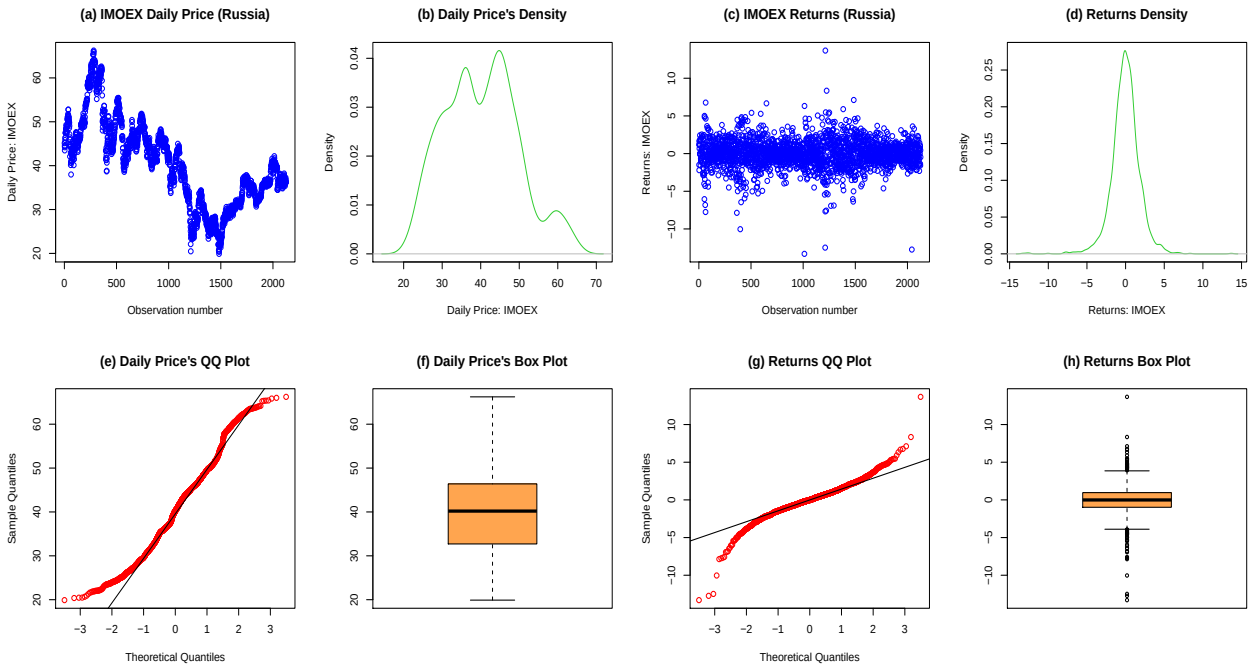


Figure 4.4: Russia: EDA of price (panels a, b, e, f) and return (panels c, d, g, h).

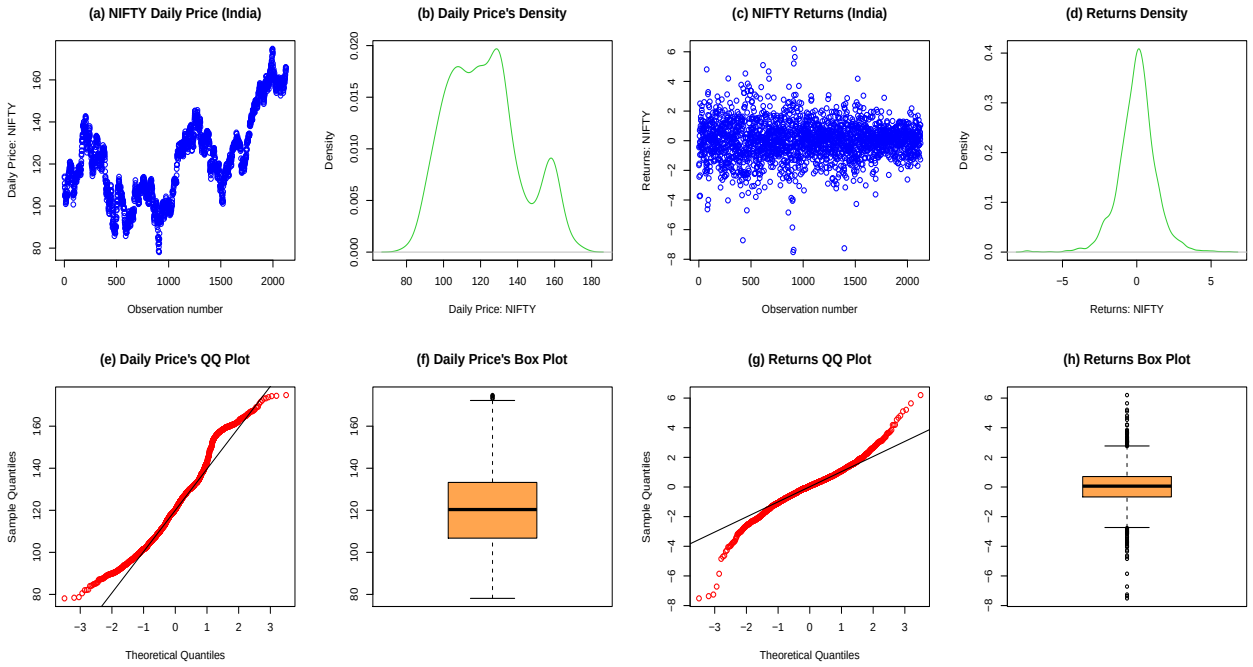


Figure 4.5: India: EDA of price (panels a, b, e, f) and return (panels c, d, g, h).

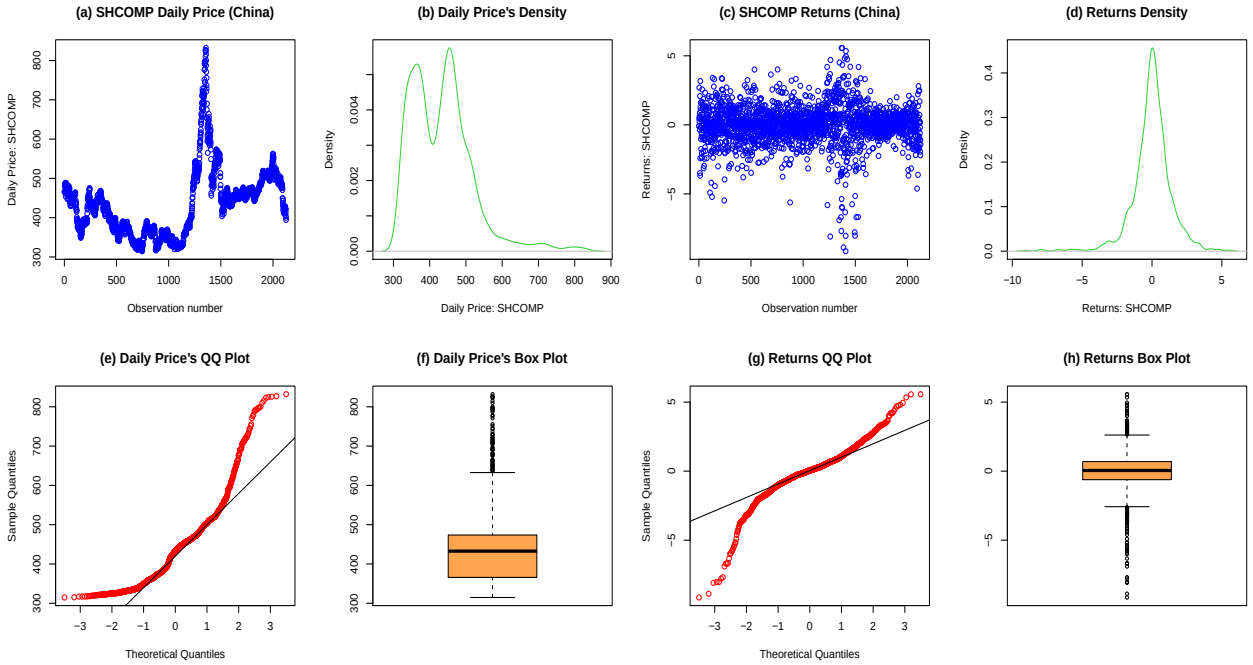


Figure 4.6: China: EDA of price (panels a, b, e, f) and return (panels c, d, g, h).

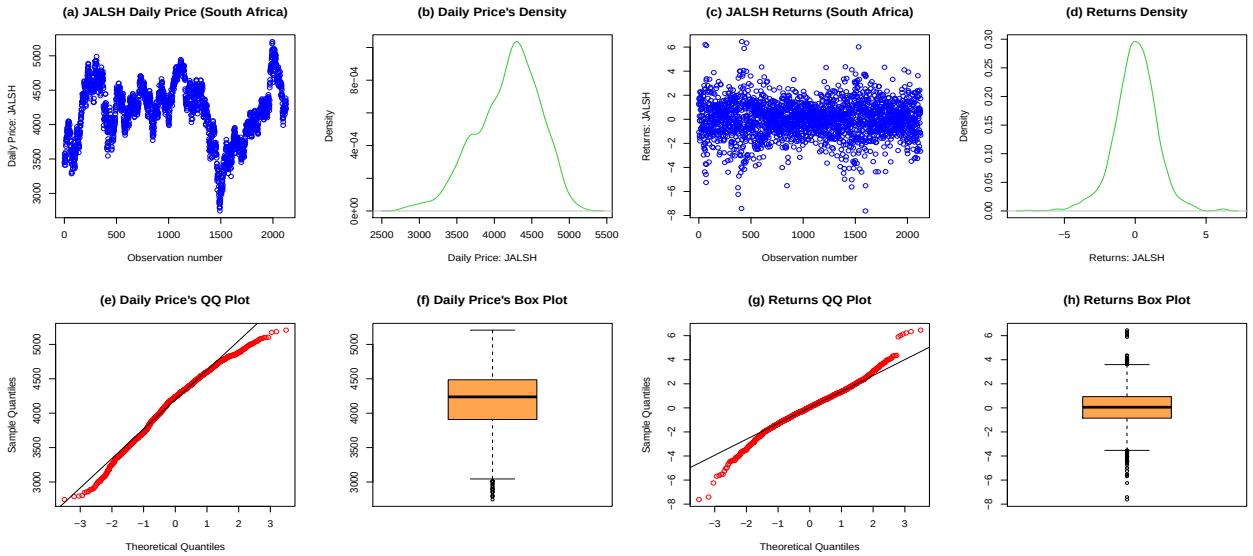


Figure 4.7: South Africa: EDA of price (panels a, b, e, f) and return (panels c, d, g, h).

Table 4.1: Descriptive Statistics.

	Brazil (Bovespa)	Russia (IMOEX)	India (NIFTY)	China (SHCOMP)	South Africa (JALSH)
PRICES					
Mean	59766.72	40.21	122.21	433.65	4182.60
Median	58369.76	40.20	120.35	432.64	4237.20
Maximum	87652.64	66.23	174.78	832.07	5207.14
Minimum	37497.48	19.90	78.12	314.68	2744.87
Std. Dev.	9724.44	9.59	19.97	85.15	422.84
Skewness	0.57	0.28	0.41	1.32	-0.48
Kurtosis	3.10	2.58	2.48	5.95	2.99
Jarque-Bera	115.95	43.65	85.01	1388.51	82.76
Probability	0.00	0.00	0.00	0.00	0.00
Observations	2126	2126	2126	2126	2126
RETURNS					
Mean	0.01	-0.01	0.02	-0.01	0.01
Median	0.02	-0.01	0.06	0.05	0.05
Maximum	6.39	13.67	6.20	5.57	6.46
Minimum	-9.21	-13.30	-7.51	-9.15	-7.62
Std. Dev.	1.43	1.84	1.27	1.41	1.51
Skewness	-0.17	-0.46	-0.30	-0.99	-0.24
Kurtosis	5.01	9.24	6.28	8.98	4.81
Jarque-Bera	368.01	3525.34	983.30	3514.65	310.12
Probability	0.00	0.00	0.00	0.00	0.00
Observations	2126	2126	2126	2126	2126

4.4 Exploratory data analysis

The first requirement in the analysis of any statistical dataset is the exploratory data analysis (EDA) which gives first hand examination of the content of the dataset with regards to detecting outliers and anomalies in the data. Exploratory analysis helps to determine whether the data satisfy basic distributional assumptions and it can suggest useful normalizing transformation.

Figures 4.3, 4.4, 4.5, 4.6 and 4.7 show visual inspection of the EDA of the BRICS equity markets' price and return series. For each of the markets, it can be observed that plots of the daily equity prices are not stationary as displayed in panels a, b, e and f. These panels show the non-stationary trend of the raw price series plot, the density plot, QQ (quantile-quantile) plot and the box plot. Furthermore, non-stationarity of the markets' prices are also displayed in Figures 4.1 and 4.2, panels a, c, e, g and i. On the other hand, panels c, d, g and h of Figures 4.3, 4.4, 4.5, 4.6 and 4.7 display an approximate stationarity for the return series plot, the density plot, QQ plot and the box plot. Stationarity display of the return series can also be seen in Figures 4.1 and 4.2, panels b, d, f, h and j. Kernel density estimation is used to estimate the density (Silverman, 1986). The observations that are not aligned on the unit diagonal of the QQ plot at the extreme sides indicate extreme observations, hence the distribution is fat-tailed.

Stationarity of the returns is also confirmed by the results from the Augmented-Dickey Fuller (ADF), Phillips-Perron (PP), and Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) tests for unit root in Section (4.6.1). The tests denote that the BRICS markets' returns are stationary for the period of study.

4.5 Descriptive statistics

The preliminary descriptive statistics for the BRICS equity prices and returns for the period under study are presented in Table 4.1. The table gives information on the sample mean, median, standard deviation, maximum, minimum, kurtosis, skewness, Jarque-Bera (JB) test and the corresponding p-values for the daily price and returns series.

To begin with, the table shows that the Indian NIFTY index has the highest positive mean returns of 0.02, followed by the Brazilian Bovespa and South African JALSH, both of which have an approximate mean return of 0.01. Russian IMOEX and Chinese SHCOMP markets have an approximate negative expected return of -0.01, indicating financial loss or negative return during the period under study.

The table also shows that the five markets return series are leptokurtic, i.e. their kurtosis are greater than 3. This indicates that extreme price changes take place more frequently during the sampled period (Sigauke et al., 2010). Kurtosis relates to the tails of a distribution, where the value 3 is associated with the kurtosis of a normal distribution. In addition, the table shows negative skewness for the five returns, implying that the market indices have long left tail distribution. Negative skewness also denotes higher probability of large decreases in equity returns for the period sampled. The BRIC's (excluding the South African index) daily prices on the other hand show positive skewness that signifies long right tail distributions. The skewness further indicates that the markets' distributions are non-normal. The non-normality is supported by the significance of the Jarque-Bera (JB) test statistic at

Table 4.2: Unit root tests for price index.

INDEX	ADF test	PP test	KPSS test
Brazil	Test statistic: -1.4881 1% CV: -3.96 5% CV: -3.41 10% CV: -3.12	Test statistic: -1.5875 1% CV: -3.9677 5% CV: -3.4145 10% CV: -3.1290	Test statistic: 4.2339 1% CV: 0.216 5% CV: 0.146 10% CV: 0.119
Russia	Test statistic: -2.4693 1% CV: -3.96 5% CV: -3.41 10% CV: -3.12	Test statistic: -2.5288 1% CV: -3.9677 5% CV: -3.4145 10% CV: -3.1290	Test statistic: 2.2499 1% CV: 0.216 5% CV: 0.146 10% CV: 0.119
India	Test statistic: -2.1081 1% CV: -3.96 5% CV: -3.41 10% CV: -3.12	Test statistic: -2.1962 1% CV: -3.9677 5% CV: -3.4145 10% CV: -3.1290	Test statistic: 2.7531 1% CV: 0.216 5% CV: 0.146 10% CV: 0.119
China	Test statistic: -2.6021 1% CV: -3.96 5% CV: -3.41 10% CV: -3.12	Test statistic: -2.2564 1% CV: -3.9677 5% CV: -3.4145 10% CV: -3.1290	Test statistic: 1.5238 1% CV: 0.216 5% CV: 0.146 10% CV: 0.119
South Africa	Test statistic: -3.2738 1% CV: -3.96 5% CV: -3.41 10% CV: -3.12	Test statistic: -3.5103 1% CV: -3.9677 5% CV: -3.4145 10% CV: -3.1290	Test statistic: 1.2572 1% CV: 0.216 5% CV: 0.146 10% CV: 0.119

Table 4.3: Unit root tests for the returns.

RETURNS	ADF test	PP test	KPSS test
Brazil	Test statistic: -15.8791 1% CV: -3.96 5% CV: -3.41 10% CV: -3.12	Test statistic: -46.2952 1% CV: -3.9677 5% CV: -3.4145 10% CV: -3.1290	Test statistic: 0.0261 1% CV: 0.216 5% CV: 0.146 10% CV: 0.119
Russia	Test statistic: -15.6789 1% CV: -3.96 5% CV: -3.41 10% CV: -3.12	Test statistic: -43.0720 1% CV: -3.9677 5% CV: -3.4145 10% CV: -3.1290	Test statistic: 0.0589 1% CV: 0.216 5% CV: 0.146 10% CV: 0.119
India	Test statistic: -14.7871 1% CV: -3.96 5% CV: -3.41 10% CV: -3.12	Test statistic: -43.4383 1% CV: -3.9677 5% CV: -3.4145 10% CV: -3.1290	Test statistic: 0.0272 1% CV: 0.216 5% CV: 0.146 10% CV: 0.119
China	Test statistic: -13.8347 1% CV: -3.96 5% CV: -3.41 10% CV: -3.12	Test statistic: -44.3166 1% CV: -3.9677 5% CV: -3.4145 10% CV: -3.1290	Test statistic: 0.0923 1% CV: 0.216 5% CV: 0.146 10% CV: 0.119
South Africa	Test statistic: -16.7423 1% CV: -3.96 5% CV: -3.41 10% CV: -3.12	Test statistic: -44.0049 1% CV: -3.9677 5% CV: -3.4145 10% CV: -3.1290	Test statistic: 0.0405 1% CV: 0.216 5% CV: 0.146 10% CV: 0.119

CV is the critical value with levels of significance 1%, 5% and 10%, at lag 8 for each test.

1% level for the prices and returns in the five indices.

4.6 Testing for stationarity, serial correlation and ARCH effects

4.6.1 Test for stationarity

The test for stationarity is first implemented on the daily price of the raw data as displayed in Table 4.2. The Augmented-Dickey Fuller (ADF) and Phillips-Perron (PP) tests in the table showed that the null hypothesis of non-stationarity, which indicates the presence of a unit root in the price series, cannot be rejected because the (negative) values of the test statistic are greater than the (negative) critical values at 1%, 5% and 10%. However, the only exception to this is the South African market where the ADF test is slightly less than the critical values at 10%, and the PP test statistic is also less than the critical values at 5% and 10%. Also, the Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test with the null hypothesis of stationarity is rejected since the (positive) values of the test statistic are greater than the (positive) critical values at 1%, 5% and 10%. These results denote that the price series are non-stationary.

Next, the raw price data are transformed to the returns by taking the first difference of logarithms of the price. After this, it is observed that in the ADF, PP and KPSS tests, the test statistics are less than the critical values at 1%, 5% and 10% levels of significance for the five BRICS markets as displayed in Table 4.3. These results indicate that the return series are stationary, since the null hypothesis of a unit root in the series is rejected under the ADF and PP tests, and the null hypothesis of stationarity cannot be rejected under the KPSS test at the three levels. Based on this outcome, the time series ARMA (p, q)-GARCH (k, v) joint model can be used to

Table 4.4: Weighted Ljung-Box Test of Standardised Residuals.

(**A**) **Brazil** Price residuals	Lag	Statistic	P-value	(**B**) **Brazil** Return residuals with ARMA(1, 1)	Statistic	P-value
	1	1877	0.0000		0.6367	0.4249
	5	2767	0.0000		1.6665	0.9932
	9	5309	0.0000		3.9328	0.7067
(**A**) **Russia** Price residuals	Lag	Statistic	P-value	(**B**) **Russia** Return residuals with ARMA(1, 1)	Statistic	P-value
	1	1891	0.0000		0.4741	0.4911
	5	2780	0.0000		0.5891	1.0000
	9	5350	0.0000		1.2519	0.9991
(**A**) **India** Price residuals	Lag	Statistic	P-value	(**B**) **India** Return residuals with ARMA(1, 1)	Statistic	P-value
	1	2034	0.0000		1.0270	0.3108
	5	3027	0.0000		1.8920	0.9749
	9	5936	0.0000		2.4840	0.9531
(**A**) **China** Price residuals	Lag	Statistic	P-value	(**B**) **China** Return residuals with ARMA(1, 1)	Statistic	P-value
	1	2044	0.0000		0.5674	0.4513
	5	3047	0.0000		2.3102	0.8677
	9	5992	0.0000		4.9254	0.4707
(**A**) **SouthAfrica** Price residuals	Lag	Statistic	P-value	(**B**) **SouthAfrica** Return residuals with ARMA(1, 1)	Statistic	P-value
	1	1810	0.0000		2.3370	0.1263
	5	2661	0.0000		3.2770	0.3100
	9	5067	0.0000		4.9960	0.4546

Table 4.5: ARCH-LM Test.

	Brazil (Bovespa)	Russia (IMOEX)	India (NIFTY)	China (SHCOMP)	South Africa (JALSH)
ARCH-LM test statistic $[MR^2]$	74.69051	146.3231	136.7945	275.4920	126.4324
Prob. - $\chi^2(5)$	{0.0000}*	{0.0000}*	{0.0000}*	{0.0000}*	{0.0000}*

*** significant at 1% level. The value in (.) is the lag length and that in {.} is the p-values. MR^2 defines the LM test statistic for which M is the number of observations and R is the sample multiple correlation coefficient.**

model the dynamic timely volatility behaviour of the markets' returns.

4.6.2 Test for serial correlation

Following stationarity of the data, it is required to test for the presence of short-range linear dependence termed serial correlation or autocorrelation in the residuals. Serial correlation may occur as a result of the relationship between a variable and its lagged version over various intervals of time. This test is carried out on both the price and return residuals using the Weighted Ljung-Box test (Fisher and Gallagher, 2012). The results are displayed under panels (A) and (B) in Table 4.4. For the price data, panel (A) of the table shows all the p-values < 0.05 at lags 1, 5 and 9 for the standardised residuals of the BRICS equity markets, which denotes strong autocorrelation in the price data.

The price data were transformed to the log-returns and some candidates ARMA(p, q) models were fitted to the five BRICS markets' return series. From the candidate models, ARMA(1,1) model as stated in equation (4.6.1) is found to be the most suitable to remove linear dependency (autocorrelation) in the markets' returns series.

$$r_t = \alpha_o + \alpha_1 r_{t-1} + \varphi_1 \varepsilon_{t-1} + \varepsilon_t. \tag{4.6.1}$$

Panel (B) of Table 4.4 shows the outcome of the "Weighted Ljung-Box" test that relates to the selected ARMA(1, 1) model for each of the BRICS markets' returns. The p-value of return residuals (at lags 1, 5 and 9) are big (greater than 0.05), hence we fail to reject the null hypothesis of "no serial correlation" in the return residuals.

Figures 7.1, 7.2, 7.3, 7.4, and 7.5 in the Appendix show the plots of autocorrelation function (ACF) and partial autocorrelation function (PACF) of the standardised residuals indicating removal of serial correlation in the series.

4.6.3 ARCH-LM test

Table 4.5 shows that the ARCH-LM test statistic in the five BRICS markets is highly significant with all the p-values very small (lower than 0.05), hence, the null hypothesis of "no ARCH effect" is strongly rejected in the residuals of the returns series. This result is a confirmation of the presence of ARCH effects in the residuals of the five indices, which indicates the existence of volatility clustering as a result of the time varying variances of the return series.

4.6.4 McLeod-Li test

The McLeod-Li test is the second approach used to carry out ARCH effects test in this study. From the plotted p-values in Figures 7.6 and 7.7 (in the Appendix), the result shows that in the BRICS markets, the test is significant at the 5% level when the number of lags of the squared residuals autocorrelations ranges from 1 to 33. This is shown with all the 33 circles enclosed within the dashed lines. The result strongly denotes that heteroscedasticity effects are present in the returns of the BRICS markets. Based on this, the GARCH models can be fitted to remove the ARCH effects in the series.

Before modelling risk and extremal dependence, volatility will need to be filtered using GARCH model, since the presence of ARCH effects have been confirmed in the return series of the five BRICS markets. The modelling and filtration of volatility of the returns are implemented using GARCH model under each of the selected error distributions.

Table 4.6: Estimated result of ARMA-GARCH model.

ARMA(1, 1)-GARCH(1, 1) under "**normal distribution**" assumption					
	Brazil (Bovespa)	Russia (IMOEX)	India (NIFTY)	China (SHCOMP)	S/Africa (JALSH)
Coefficients					
	Mean				
$\hat{\mu}$ (constant)	0.0268***	0.0487***	0.0556*	0.0083***	0.0210***
	Variance				
$\hat{\omega}$	0.0821*	0.0737*	0.0225*	0.0062*	0.0257*
$\hat{\alpha}$	0.0657*	0.0732*	0.0693*	0.0480*	0.0493*
$\hat{\beta}$	0.8943*	0.9063*	0.9176*	0.9502*	0.9398*
$\hat{\alpha} + \hat{\beta}$	0.9600	0.9795	0.9869	0.9982	0.9891
AIC	3.4903	3.8767	3.1663	3.2220	3.5489
BIC	3.5063	3.8927	3.1822	3.2379	3.5649
SIC	3.4903	3.8767	3.1662	3.2219	3.5489
HQIC	3.4962	3.8825	3.1721	3.2278	3.5548
ARCH LM test statistic (5)	0.7027	0.1583	0.4248	1.3510	0.9846
p-value (5)	(0.8226)	(0.9750)	(0.9056)	(0.6325)	(0.7375)
ARMA(1, 1)-GARCH(1, 1) under "**student's t distribution**" assumption					
Coefficients					
	Mean				
$\hat{\mu}$ (constant)	0.0289***	0.0320***	0.0616*	0.0411**	0.0339***
	Variance				
$\hat{\omega}$	0.0623*	0.0335*	0.0202*	0.0106*	0.0251*
$\hat{\alpha}$	0.0571*	0.0600*	0.0626*	0.0523*	0.0496*
$\hat{\beta}$	0.9120*	0.9298*	0.9256*	0.9443*	0.9399*
$\hat{\alpha} + \hat{\beta}$	0.9691	0.9898	0.9882	0.9966	0.9895
AIC	3.4664	3.7844	3.1302	3.1501	3.5298
BIC	3.4851	3.8030	3.1488	3.1688	3.5484
SIC	3.4664	3.7843	3.1302	3.1501	3.5297
HQIC	3.4733	3.7912	3.1370	3.1570	3.5366
ARCH LM test statistic (5)	0.3741	0.1038	0.4484	1.4610	0.9945
p-value (5)	(0.9200)	(0.9861)	(0.8987)	(0.6026)	(0.7346)

Note: "*", "" and "***" are 1%, 5% and 10% levels of significance respectively. ARCH LM test statistic (5) denotes ARCH effects up to the 5^{th} order with p-values in parentheses, and 5% level of significance is used in every case.**

Table 4.7: Estimated result of ARMA-GARCH model.

ARMA(1, 1)-GARCH(1, 1) under "**generalised error distribution" (GED)** assumption					
	Brazil (Bovespa)	Russia (IMOEX)	India (NIFTY)	China (SHCOMP)	S/Africa (JALSH)
Coefficients					
	Mean				
$\hat{\mu}$ (constant)	0.0281***	0.0221***	0.0672*	0.0413*	0.0372***
	Variance				
$\hat{\omega}$	0.0722*	0.0450*	0.0207*	0.0089*	0.0285*
$\hat{\alpha}$	0.0610*	0.0637*	0.0646*	0.0502*	0.0496*
$\hat{\beta}$	0.9036*	0.9233*	0.9229*	0.9458*	0.9378*
$\hat{\alpha} + \hat{\beta}$	0.9646	0.9870	0.9875	0.9960	0.9874
AIC	3.4711	3.8091	3.1303	3.1459	3.5342
BIC	3.4897	3.8278	3.1489	3.1645	3.5529
SIC	3.4710	3.8091	3.1303	3.1459	3.5342
HQIC	3.4779	3.8160	3.1371	3.1527	3.5411
ARCH LM test statistic (5)	0.5283	0.1230	0.3915	1.4830	1.0515
p-value (5)	(0.8752)	(0.9824)	(0.9151)	(0.5968)	(0.7177)
under "**skew − normal distribution"** assumption					
Coefficients					
	Mean				
$\hat{\mu}$ (constant)	0.0225***	0.0218***	0.0478**	-0.0057***	0.0126***
	Variance				
$\hat{\omega}$	0.0775*	0.0667*	0.0210*	0.0076*	0.0243*
$\hat{\alpha}$	0.0645*	0.0675*	0.0657*	0.0484*	0.0495*
$\hat{\beta}$	0.8976*	0.9129*	0.9216*	0.9483*	0.9400*
$\hat{\alpha} + \hat{\beta}$	0.9621	0.9804	0.9873	0.9967	0.9895
AIC	3.4895	3.8665	3.1639	3.2101	3.5410
BIC	3.5081	3.8852	3.1825	3.2287	3.5597
SIC	3.4895	3.8665	3.1639	3.2100	3.5410
HQIC	3.4963	3.8734	3.1707	3.2169	3.5478
ARCH LM test statistic (5)	0.6653	0.1262	0.4315	1.2450	0.9522
p-value (5)	(0.8340)	(0.9818)	(0.9036)	(0.6619)	(0.7472)

Note: "*", "" and "***" are 1%, 5% and 10% levels of significance respectively. ARCH LM test statistic (5) denotes ARCH effects up to the 5^{th} order with p-values in parentheses, and 5% level of significance is used in every case.**

Table 4.8: Estimated result of ARMA-GARCH model.

ARMA(1, 1)-GARCH(1, 1) under "**skew − student's t distribution**" assumption					
	Brazil (Bovespa)	Russia (IMOEX)	India (NIFTY)	China (SHCOMP)	S/Africa (JALSH)
Coefficients					
	Mean				
$\hat{\mu}$ (constant)	0.0246***	0.0241***	0.0430**	0.0147***	0.0116***
	Variance				
$\hat{\omega}$	0.0619*	0.0332*	0.0196*	0.0107*	0.0241*
$\hat{\alpha}$	0.0569*	0.0593*	0.0623*	0.0522*	0.0503*
$\hat{\beta}$	0.9124*	0.9303*	0.9261*	0.9437*	0.9396*
$\hat{\alpha} + \hat{\beta}$	0.9693	0.9896	0.9884	0.9959	0.9899
AIC	3.4672	3.7850	3.1283	3.1476	3.5241
BIC	3.4885	3.8063	3.1496	3.1689	3.5454
SIC	3.4672	3.7850	3.1283	3.1475	3.5241
HQIC	3.4750	3.7928	3.1361	3.1554	3.5319
ARCH LM test statistic (5)	0.3678	0.1021	0.4492	1.2468	0.9660
p-value (5)	(0.9218)	(0.9864)	(0.8985)	(0.6614)	(0.7431)
ARMA(1, 1)-GARCH(1, 1) under "**skew − GED distribution**" assumption					
Coefficients					
	Mean				
$\hat{\mu}$ (constant)	0.0193***	0.0079***	0.0453**	0.00028***	0.0117***
	Variance				
$\hat{\omega}$	0.0688*	0.0446*	0.0202*	0.0098*	0.0262*
$\hat{\alpha}$	0.0599*	0.0624*	0.0633*	0.0505*	0.0509*
$\hat{\beta}$	0.9062*	0.9242*	0.9241*	0.9446*	0.9377*
$\hat{\alpha} + \hat{\beta}$	0.9661	0.9866	0.9874	0.9951	0.9886
AIC	3.4712	3.8094	3.1285	3.1411	3.5278
BIC	3.4925	3.8307	3.1498	3.1624	3.5491
SIC	3.4711	3.8093	3.1285	3.1411	3.5278
HQIC	3.4790	3.8172	3.1363	3.1489	3.5356
ARCH LM test statistic (5)	0.4955	0.1192	0.4189	1.1673	0.9983
p-value (5)	(0.8849)	(0.9832)	(0.9073)	(0.6840)	(0.7334)

Note: "*", "" and "***" are 1%, 5% and 10% levels of significance respectively. ARCH LM test statistic (5) denotes ARCH effects up to the 5^{th} order with p-values in parentheses, and 5% level of significance is used in every case.**

Table 4.9: Estimated result of ARMA-GARCH model.

ARMA(1, 1)-GARCH(1, 1) under "**GHYP distribution**" assumption					
	Brazil (Bovespa)	Russia (IMOEX)	India (NIFTY)	China (SHCOMP)	S/Africa (JALSH)
Coefficients					
	Mean				
$\hat{\mu}$ (constant)	0.0217***	0.0122***	0.0451**	0.0045***	0.0098***
	Variance				
$\hat{\omega}$	0.0622*	0.0332*	0.0196*	0.0102*	0.0240*
$\hat{\alpha}$	0.0568*	0.0582*	0.0626*	0.0507*	0.0496*
$\hat{\beta}$	0.9123*	0.9312*	0.9253*	0.9442*	0.9401*
$\hat{\alpha} + \hat{\beta}$	0.9691	0.9894	0.9879	0.9949	0.9897
AIC	3.4678	3.7847	3.1281	3.1431	3.5252
BIC	3.4917	3.8087	3.1521	3.1670	3.5492
SIC	3.4677	3.7847	3.1281	3.1430	3.5252
HQIC	3.4765	3.7935	3.1369	3.1518	3.5340
ARCH LM test statistic (5)	0.3627	0.0978	0.4318	1.1990	0.9815
p-value (5)	(0.9232)	(0.9872)	(0.9035)	(0.6749)	(0.7384)

Note: "*", "" and "***" are 1%, 5% and 10% levels of significance respectively. ARCH LM test statistic (5) denotes ARCH effects up to the 5^{th} order with p-values in parentheses, and 5% level of significance is used in every case.**

Table 4.10: The BRICS volatility hierarchy.

	Most volatile				Least volatile
Error distributions					
Student's t	Chinese	Russian	South African	Indian	Brazilian
Skew-student's t	Chinese	South African	Russian	Indian	Brazilian
GHYP	Chinese	South African	Russian	Indian	Brazilian
GED	Chinese	Indian	South African	Russian	Brazilian
Skew-GED	Chinese	South African	Indian	Russian	Brazilian

4.7 Empirical outcomes of the ARMA-GARCH models

This section focuses on the volatility dynamics of the five BRICS markets to determine the magnitudes of volatility in each and arrange them from the highest to the lowest for investors' decision making. To begin with, several candidate ARMA(p, q)-GARCH(k, v) models were run to obtain a combined model that can best remove linear dependency and heteroscedasticity in the return series. From the candidate models, ARMA (1, 1) and GARCH (1, 1) are jointly selected as the most adequate for the Brazilian, Russian, Indian, Chinese and South African markets. That is, the parsimonious and effective ARMA (1, 1) - GARCH (1, 1) joint process is used. The fitting of this joint ARMA-GARCH model is carried out under the selected error distributions of a normal, student's t, GED (with their skew versions), and the GHYP.

Tables 4.6, 4.7, 4.8, and 4.9 show the empirical results (parameter estimates) of the ARMA$(1, 1)$-GARCH$(1, 1)$ under each of the stated distributions. The evidence from the daily series in the five indices indicates that all the GARCH parameters (i.e. ω, α, and β) under all the error distributions are statistically significant at 1% level. The AIC, BIC, SIC and HQIC are jointly used for model selection under each of the best error distributions. Model selection is based on the values of the information criteria, where the least values are indicative of the best model, under a given error assumption, to describe a market.

4.7.1 Residual diagnostic test

After fitting the GARCH models to the returns, residual diagnostics were carried out with the use of the weighted ARCH LM tests to determine whether ARCH (heteroskedastic) effects have been filtered out of the residuals or not. The results from the "ARCH LM test statistic (5)" in Tables 4.6, 4.7, 4.8, and 4.9 show that at lag order 5, under all the stated error distributions, the p-values are sufficiently large (above 5%) for the five BRICS markets. Based on this, we fail to reject the null hypothesis of "no ARCH effect" in the residuals, and hence conclude that the variance equations are well specified.

From Tables 4.6, 4.7, 4.8, and 4.9, for the Brazilian market, all the information criteria have their least values under the Student's t innovations of the ARMA(1, 1)-GARCH(1, 1) model. This indicates that the Brazilian Bovespa market can be best described by an ARMA(1, 1)-GARCH(1, 1) model using the Student's t distribution.

For the Russian market also, all the information criteria have their least values under the Student's t innovations of the ARMA-GARCH model (like the Brazilian market). Hence, it is concluded that the Russian IMOEX market can be best characterised by an ARMA(1, 1)-GARCH(1, 1) model under the Student's t distribution.

In the Indian NIFTY market, both the AIC and SIC have their least values under the generalised hyperbolic (GHYP) distribution, while the BIC and HQIC have their least values under the Student's t and skew-student's t distributions respectively. However, since the values of the $\{\text{AIC} = \text{SIC}\} < \{\text{HQIC}\} < \{\text{BIC}\}$ (see Tables 4.6, 4.8 and 4.9), it is concluded that the Indian NIFTY market can be best characterised by an ARMA(1, 1)-GARCH(1, 1) model under the generalised hyperbolic (GHYP) distribution using the AIC-SIC's least values.

For the Chinese SHCOMP market, all the information criteria have their least values under the skew-GED innovations, which therefore indicates that it is the best distribution assumption to characterise the market under the ARMA(1, 1)-GARCH(1,

1) model.

Lastly, for the South African JALSH market, all the information criteria have their least values under the skew-student's t innovations. Hence, the South African market is best described by the ARMA(1, 1)-GARCH(1, 1) model under the skew-student's t distribution.

As a summary, it is concluded that using the ARMA(1, 1)-GARCH(1, 1) model, the Brazilian Bovespa and the Russian IMOEX markets can both be well characterised (or described) by a heavy-tailed Student's t distribution, while the Indian NIFTY market is best characterised by the generalised hyperbolic (GHYP) distribution. Also, the Chinese SHCOMP and South African JALSH markets are best described by the skew-GED and skew-student's t distributions, respectively.

4.7.2 The QQ diagnostic plots of ARMA-GARCH models

Figures 7.8, 7.9, 7.10, 7.11 and 7.12 in the Appendix show the densities of the standardised residuals and diagnostic quantile-quantile (QQ) plots of the ARMA-GARCH models under each of the best describing error distributions. The QQ plots under each innovations display suitable levels of linearity since the points are well aligned on the straight lines, except at the tails (which denotes extreme movement). Furthermore, the plots of the densities of the standardised residuals confirm that each of the best error distributions (i.e., the student's t, GHYP, skew-student's t and skew-GED) is a better fit to the BRICS markets' return distributions than a normal distribution (Sigauke, 2016).

4.7.3 Persistence of volatility

The sum of the coefficients (α, β) in the conditional variance equation of the GARCH process measures the speed of decay of shocks to volatility. This is referred to as the

persistence of the GARCH model, and it indicates how fast large volatilities decline after a shock. Shocks to the conditional variability are highly persistent when the sum is greater than one ($\alpha + \beta > 1$), suggesting that the forecasts of volatility are explosive. This implies the presence of volatility clustering in the series. If the sum of the coefficients equals one (i.e. $\alpha + \beta = 1$), then the persistence of shocks to volatility is felt forever, and the model will be unable to determine the unconditional variance of the process. Engle and Bollerslev (1986) refer to this type of process as "Integrated-GARCH".

Lastly, shocks to volatility displays long persistence into the future if the sum ($\alpha + \beta$) is close to one. This occurs because the variance process reverts very slowly to the mean (normal) state in a process termed "mean reversion". The closer the sum ($\alpha + \beta$) to 1, the longer it takes volatility to revert to the mean state.

4.8 Conclusion: Volatility order of the BRICS equity markets

The pattern of volatility persistence in each of the BRICS markets can be determined by a non-Gaussian distribution like the GED, student's t and GHYP. Financial returns are known to exhibit fat tails or leptokurtosis, and Gaussian distributions like the normal distribution is not suitable for modelling such (Danielsson and de Vries, 1997; McNeil and Frey, 2000).

The values of volatility persistence, i.e. the sum of the coefficients ($\alpha + \beta$), at various magnitudes is close to one under the student's t, skew-student's t, GED, skew-GED and GHYP distribution (see Table 4.6, 4.7, 4.8, and 4.9). The volatility hierarchy of the BRICS equity markets is evaluated based on the relative outcome of the sum of the coefficients, where the highest summation gives the highest volatility

persistence. From the summary in Table 4.10, it can be observed that the persistence of volatility in the BRICS markets does not follow the same hierarchical pattern under the error distributions, except under the skew-student's t and GHYP distributions where the pattern is the same. Under these two assumptions, i.e. the skew-student's t and GHYP, in a descending hierarchical order of magnitudes, volatility with persistence is highest in the Chinese market, followed by the South African market, then the Russian, Indian and Brazilian markets respectively (see Table 4.10).

For the Student's t distribution, the Chinese market has the highest volatility persistence, followed by the Russian, South African, Indian and Brazilian markets in that order. For the GED, the pattern of the volatility persistence in a descending hierarchy is the Chinese, Indian, South African, Russian and Brazilian markets respectively. Lastly, for the skew-GED, the descending hierarchical pattern is the Chinese, South African, Indian, Russian and Brazilian markets respectively. However, under each of the error distributions, the Chinese market is the most volatile, while the Brazilian market is the least volatile.

Since the skew-student's t and GHYP distributions gave the same volatility pattern, we can use one of them to demonstrate the spread of the markets' volatility from the highest to the lowest. Hence, Figures 7.13, 7.14, 7.15, 7.16 and 7.17 in the Appendix show the volatility persistence results under the skew-student's t distribution as displayed by the visual inspection in the "Shock versus Persistence" plots of the BRICS stock markets' volatility persistence diagrams.

From the "Shock versus Persistence" diagrams, volatility of the Chinese market persists for as long as about 25000 in magnitude (Figure 7.13), while that of South African market which is next in the hierarchy is about 14000 in persistence (Figure 7.14). The Russian, Indian and Brazilian markets are about 12000, 6000 and 3000 respectively in magnitudes of persistence (see Figures 7.15, 7.16 and 7.17).

Chapter 5

Risks of the BRICS equity markets

5.1 Introduction: the univariate analysis

This chapter focuses on the univariate aspect of the two models: the conditional extreme value (CEV) and point process. The risk in each of the five BRICS domestic markets is modelled using the univariate versions of the two models. The steps involved in the risk modelling and analysis for each of the markets include: threshold selection and diagnostics, sensitivity analysis for a suitable threshold choice, declustering of threshold exceedances, parameters' estimation and diagnostics, return levels analysis and diagnostics.

5.1.1 The conditional univariate extreme value (CUEV) model

This is the univariate version of the CEV model that fits the GPD to marginal variables. Hence, the GPD is fitted to each of the BRICS return's residuals to obtain the magnitude of the risks in the markets. This fitting is done one after the other as follows: Brazil, Russia, India, China and South Africa.

5.1.2 Large positive and negative residual observations

From the GARCH regression equation of the financial return r_t as stated in Section 3.5 (equations 3.5.1 and 3.5.2),

$$r_t = \mu + \varepsilon_t. \tag{5.1.1}$$

The residual ε_t becomes,

$$\varepsilon_t = r_t - \mu_t \tag{5.1.2}$$

$$\varepsilon_t = r_t - \hat{r}_t \tag{5.1.3}$$

where r_t and $\hat{r}_t$ are the return observations (observed values) and the predicted or estimated values respectively.

When the observed return value r_t is greater than the estimate or predicted value $\hat{r}_t$, the residual is positive $\varepsilon_t > 0$ and that amounts to under-prediction, which corresponds to financial losses. Otherwise, it is over-prediction with a negative residual $\varepsilon_t < 0$ and it relates to gains. That is, large or extreme positive and negative residuals relate to losses and gains respectively (McNeil and Frey, 2000). This study is primarily interested in modelling the unexplained large positive residuals which reflect extreme financial losses in the BRICS return series.

5.2 The Brazilian market: IBOV

The Brazilian market index is called the "Bovespa index", abbreviated - IBOV, and it contains 2126 observations for the periods 5^{th} January 2010 to 6^{th} August 2018.

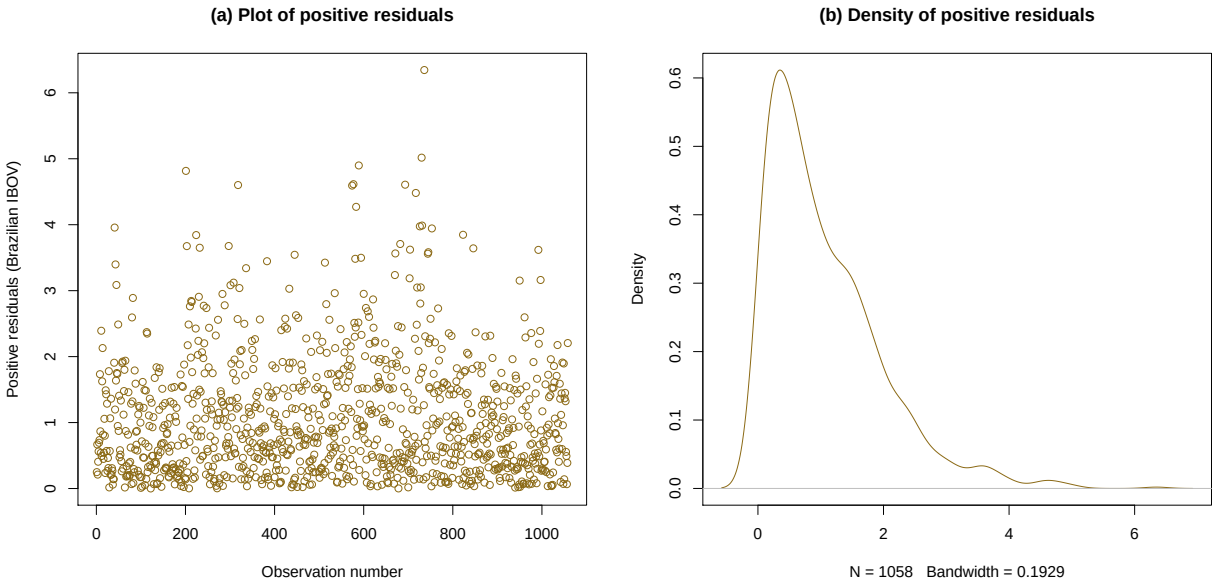


Figure 5.1: IBOV: Positive residuals.

5.2.1 IBOV: Positive residual observations

The modelling of the risk in each of the markets is restricted to the positive residuals of the entire equity observations. From the original equity data with 2126 number of observations, there are 1058 positive residual observations in this market as displayed in Figure 5.1. The positive residual observations and their corresponding density are given in Figure 5.1, panels (a) and (b) respectively.

5.2.2 IBOV: Threshold selection

Two threshold selection models are used in this market and the rest of the BRICS markets for the selection of an appropriate threshold. The first model is the "extreme value mixture models" and the second model is the "shape threshold stability plot". The latter model is used to verify the outcome of the former model.

The extreme value mixture models can be implemented under the parametric,

semi-parametric and non-parametric approaches. This study uses the non-parametric approach over the other two approaches since it is known to provide the best tail estimator if the population distribution (of the market's return) is unknown (Hu and Scarrott, 2018), and furthermore, it is more robust to bulk model than the parametric technique (Yang, 2013). In the non-parametric mixture model, we have the bulk model where a kernel density is fitted and a GPD is fitted to the tail. The resulting mixture model is then called the Kernel-GPD (KenGPD) model. That is, the bulk model under the threshold is the standard kernel density estimator and the tail model is a GPD above the threshold. The threshold can be estimated using either the bulk model based or the parameterised tail fraction approach.

Following the threshold selection steps of Hu and Scarrott (2018), this study combined the plots of the bulk model based tail fraction and the parameterised tail fraction approaches on the same diagram and used their diagnostic plots to assess which of the two gives a better threshold's parameter estimate (see Figure 5.2). The figure shows the output of the bulk model based tail fraction (blue solid line), where the threshold $u = 2.1731$. For the parameterised tail fraction (red solid line) approach, $u = 2.1725$ is obtained. These threshold values are approximately the same, hence they overlapped, where the red colour of the density and threshold line of the parameterised tail fraction almost completely overshadow the blue colour of that of the bulk model based tail fraction.

5.2.3 Diagnostic plots of the bulk and parameterised approaches

There is no obvious difference between the diagnostic plots of the two tail fraction approaches since the threshold values are roughly the same (see Figures 5.3 and 5.4). But, from a closer scrutiny, the diagnostic plots of the fitted parameterised tail

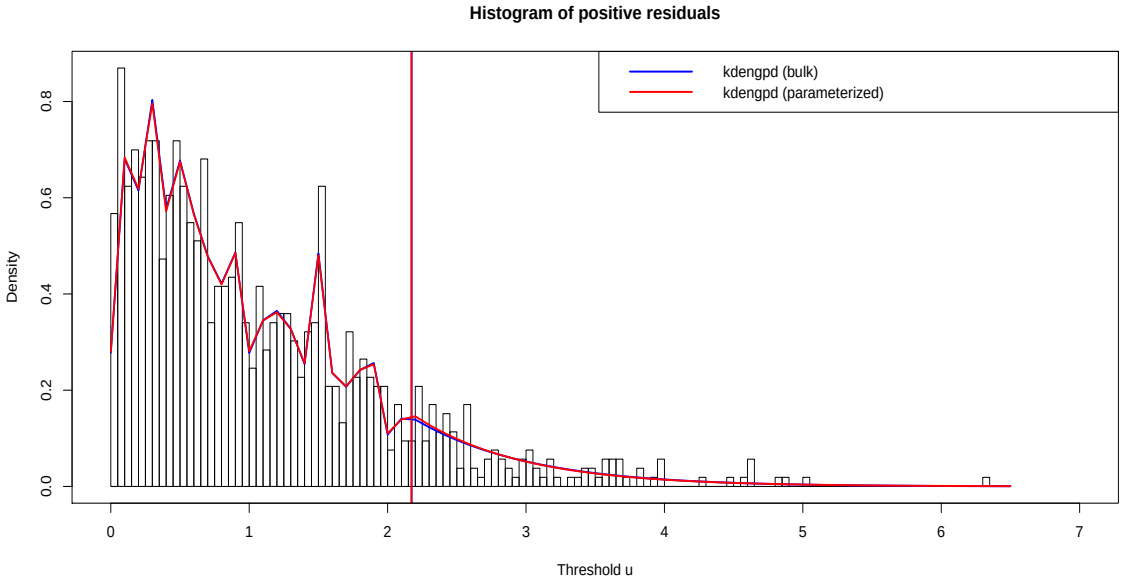


Figure 5.2: IBOV: Threshold selection.

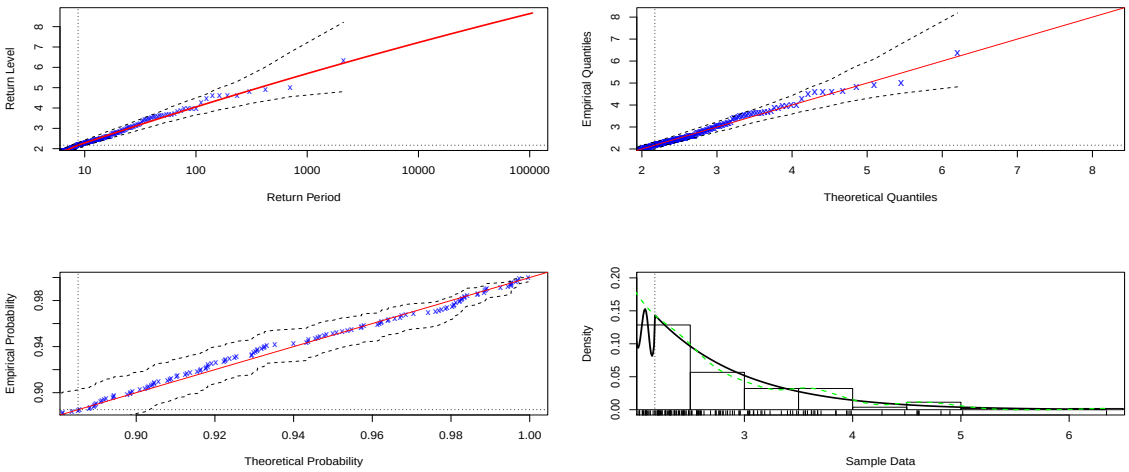


Figure 5.3: IBOV: Diagnostic plots of the bulk model based tail fraction.

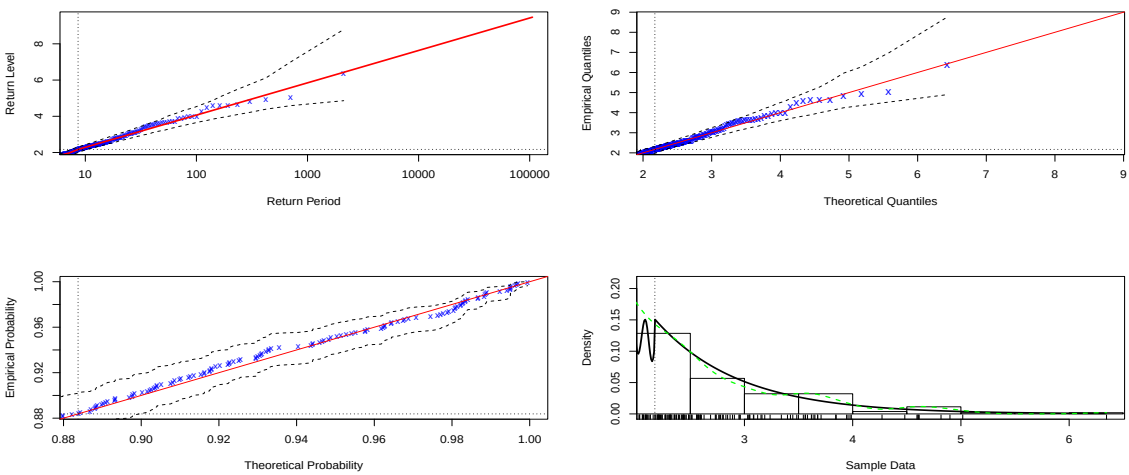


Figure 5.4: IBOV: Diagnostic plots of the parameterised tail fraction.

fraction show a slightly higher levels of accuracy in linearity of the data points when compared to the bulk model based. The data points are more closely aligned on the straight lines of the "return level", "quantiles", and "probability" plots than they are observed in the bulk model based approach.

However, the diagnostic plots of the fitted bulk and parameterised models for the threshold selection indicate that the fit is adequate. Each set of plotted points is close to linear and the data points are all enclosed within the dashed confidence limits. In addition, the density estimate seems consistent with the data's histogram. Hence, the quartet diagnostic plots indicate that the fit of the non-parametric Kernel-GPD models is reasonably satisfactory.

5.2.4 Extremal index for final threshold choice

Following Ferro (2003) and Ferro and Seger (2003), a sensible choice of an appropriate threshold is where the extremal index (θ) is greater than or equal to 0.5, hence a minimum of $\theta = 0.5$ is used in this study for final threshold selection, required for declustering of the cluster exceedances. In order to determine this appropriate threshold, it is required that the plot of "normalised inter-exceedance times" against "standard exponential quantiles" should be piecewise-linear with a breakpoint at the $(1-\theta)$-quantile, $-\log\theta$ (see Figures 3.1 and 5.5). The sloping line has gradient θ^{-1}, and the vertical line is indicated by the $(1-\theta)$-quantile. Based on this, a sensitivity analysis is carried out to ascertain the plot of a suitable threshold that is best piecewise-linear with extremal index θ of (a minimum of) 0.5.

As explained in Section 3.20, the extremal index θ is both the reciprocal of the mean of the non-zero interexceedance times, and the proportion of non-zero interexceedance times (Ferro, 2003; Ferro and Segers, 2003). The findings of Ferro (2003) showed that an extremal index $\theta = 0.5$ approximately aligns to this piecewise-linearity requirement (see Figures 3.1).

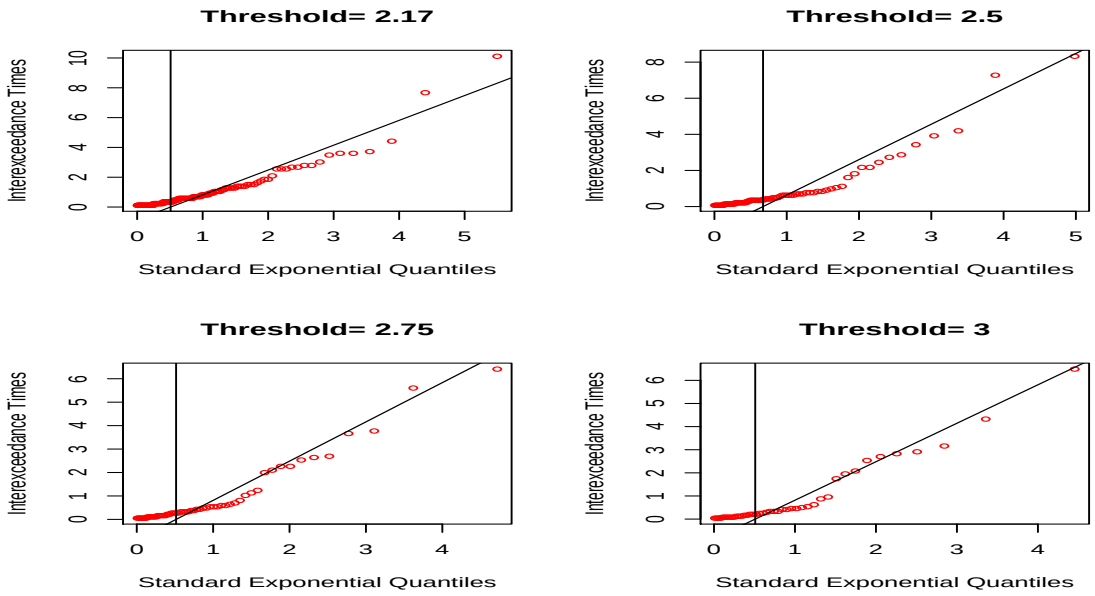


Figure 5.5: IBOV: Sensitivity analysis plots for threshold selection.

5.2.5 Sensitivity analysis

In order to obtain a suitable threshold value where the exceedances are piecewise-linear with a minimum extremal index of 0.50, a sensitivity analysis was carried out as displayed in Figure 5.5. The sensitivity analysis is initiated with the approximate threshold value of $u = 2.17$ from the parameterised tail fraction approach since its diagnostic plots display slightly higher levels of linearity. Out of the four sensitivity plots with different thresholds in the figure, threshold $u = 2.17$ is chosen because it is the most appropriate on the sensitivity analysis plot in terms of piecewise-linearity with extremal index $\hat{\theta}$ of 0.6002. This means that exceedances occur in groups of $\frac{1}{0.6002} = 1.666 \approx 2$.

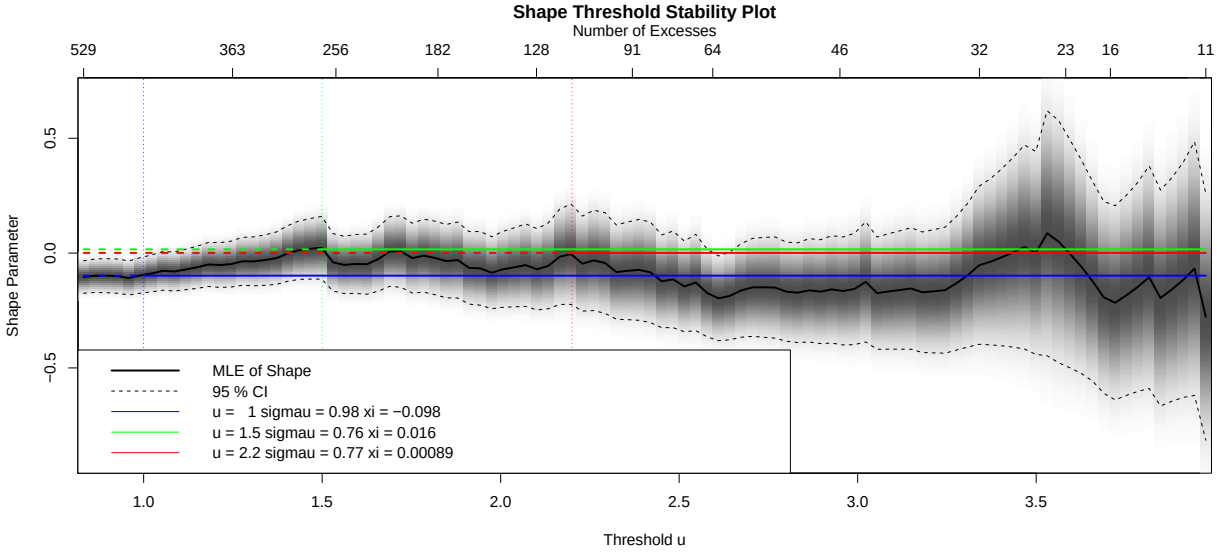


Figure 5.6: IBOV: Shape threshold stability plot.

5.2.6 Shape threshold stability plot

The traditional shape threshold stability plot is used in this study to verify and give some levels of credence to the threshold choice from the mixture models. As displayed in Figure 5.6, three potential thresholds where the plot shows significant departures from linearity are identified around $u = 1.0$, $u = 1.5$ and $u = 2.2$. Therefore, it can be seen that the choice of $u = 2.2$ from the shape threshold stability plot is approximately consistent with the threshold estimate ($u = 2.17$) from the Kernel density mixture models.

5.2.7 Declustering

From the estimated threshold $u = 2.17$, 123 threshold exceedances are generated. After declustering at this threshold, 71 cluster-maxima are obtained as shown in Figure 5.7. The cluster-maxima are "exceedance residuals" (McNeil and Frey, 2000).

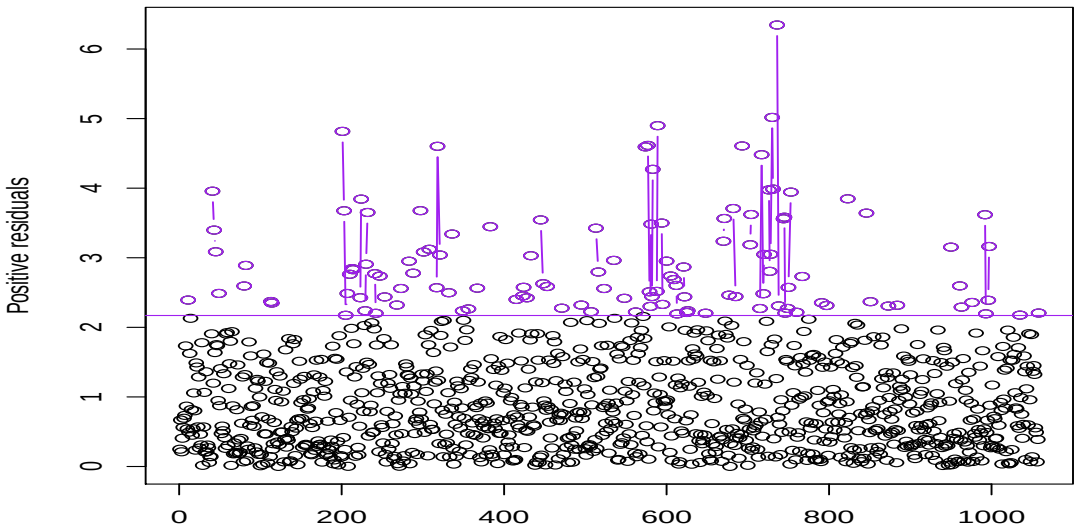


Figure 5.7: IBOV: Declustered exceedances at $u = 2.17$ (cluster-maxima).

They are the maxima of the clusters of exceedances over a high threshold, and are the independent extreme tail observations on which the GPD is fitted.

5.2.8 CEV model: GPD fit to cluster-maxima

The univariate version of the conditional extreme value (CEV) model fits the GPD to marginal variables (Southworth, Heffernan and Metcalfe, 2016). Hence, the GPD will now be fitted to the 71 cluster-maxima that are generated after declustering the cluster exceendances. These cluster-maxima are collectively classified as the risk to be modelled in the Brazilian IBOV market. This risk indicates extreme (financial) losses in the market and it is modelled by fitting the GPD to the 71 observations. The result of the GPD fit is shown in Table 5.1, where the standard errors of the GPD estimates are enclosed in parentheses and the confidence intervals are in the square brackets. The table also includes the sample proportion, η_u of points which are the

Table 5.1: IBOV: GPD parameter estimates.

	CI	u	n	w	$\eta_u = \frac{w}{n}$	$\hat{\xi}$	$\hat{\sigma}$	Log. lik
	90%	2.17	1058	71	0.0671	-0.05 (0.14)	0.94 (0.18)	-62.57
						[-0.28; 0.18]	[0.63; 1.24]	
IBOV	95%	2.17	1058	71	0.0671	-0.05 (0.14)	0.94 (0.18)	-62.57
						[-0.32; 0.22]	[0.58; 1.29]	
	99%	2.17	1058	71	0.0671	-0.05 (0.14)	0.94 (0.18)	-62.57
						[-0.41; 0.31]	[0.46; 1.41]	

Table 5.2: IBOV: GPD parameter estimates and profile likelihood intervals.

	CI	u	n	w	$\eta_u = \frac{w}{n}$	$\hat{\xi}$	$\hat{\sigma}$	Log. lik
	90%	2.17	1058	71	0.0671	-0.05 (0.14)	0.94 (0.18)	-62.57
						[-0.22; 0.23]	[0.68; 1.24]	
IBOV	95%	2.17	1058	71	0.0671	-0.05 (0.14)	0.94 (0.18)	-62.57
						[-0.22; 0.30]	[0.64; 1.31]	
	99%	2.17	1058	71	0.0671	-0.05 (0.14)	0.94 (0.18)	-62.57
						[-0.22; 0.45]	[0.56; 1.44]	

cluster-maxima exceeding the threshold at $u = 2.17$. The sample proportion η_u is the ratio of the number of cluster-maxima w to the total positive residual observations n. The negative estimate ($\hat{\xi} = -0.05$) of the shape parameter as given in Table 5.1 indicates a short tailed distribution and it is a reflection of convexity (Coles, 2001).

Confidence intervals can alternatively be obtained for the shape parameter (ξ) with a better accuracy by the use of profile likelihood (Coles 2001). The profile confidence intervals in this study are computed using the package "POT" developed by Mathieu and Dutang (2016). Table 5.2 shows the results of the GPD parameters estimation and their corresponding profile log-likelihood confidence intervals. However, since the values of the confidence intervals of the estimated shape parameter at the three confidence levels in the table are all from negative to positive, i.e. they include zero, a formal hypothesis test is carried out in Section 5.2.9 using the likelihood ratio (LR) and the modified likelihood ratio (LR_{**}) tests to determine if $\xi = 0$ or otherwise.

Table 5.3: Likelihood ratio test for GPD's ξ estimate.

IBOV index	w	$\hat{\xi}$	LR	CV	LR_{**}	CV
	71	-0.05	-125.14	10%: 2.706 5%: 3.841 1%: 6.635	-120.21	10%: 2.706 5%: 3.841 1%: 6.635

5.2.9 Formal hypothesis test

The likelihood ratio (LR) and the modified likelihood ratio (LR_{**}) tests as detailed in Sections 3.13.3 and 3.13.4 are used with the null hypothesis that $\xi = 0$ against the alternative hypothesis that $\xi \neq 0$. From the outcome in Table 5.3, at the 1%, 5% and 10% levels of significance, we fail to reject the null hypothesis, implying that the shape parameter $\xi = 0$. This suggest a Gumbel domain of attraction for the data, hence the risk in the Brazilian IBOV market can be characterised by the Gumbel class of distributions.

5.2.10 Diagnostics: Model checking

Statistical models are fit to data so as to make reasonable conclusions about the population where the data is drawn. Model diagnostics are used to ascertain the validity of a model. The accuracy of a model can be checked or judged using diagnostics to know whether it agrees with the data used to estimate it.

Four diagnostic plots are displayed in Figure 5.8 to assess the performance of the GPD fit to the 71 cluster-maxima. Each set of plotted points of the quantile plot, probability plot and return level plot is near-linear, and the corresponding density estimate appears consistent with the histogram of the cluster-maxima. Consequently, the quartet diagnostic plots give support to the validity of the CEV model's GPD fit.

The goodness of fit tests of Anderson-Darling and Cramér-von Mises are further used to ascertain how well the GPD model fits the cluster-maxima observations in

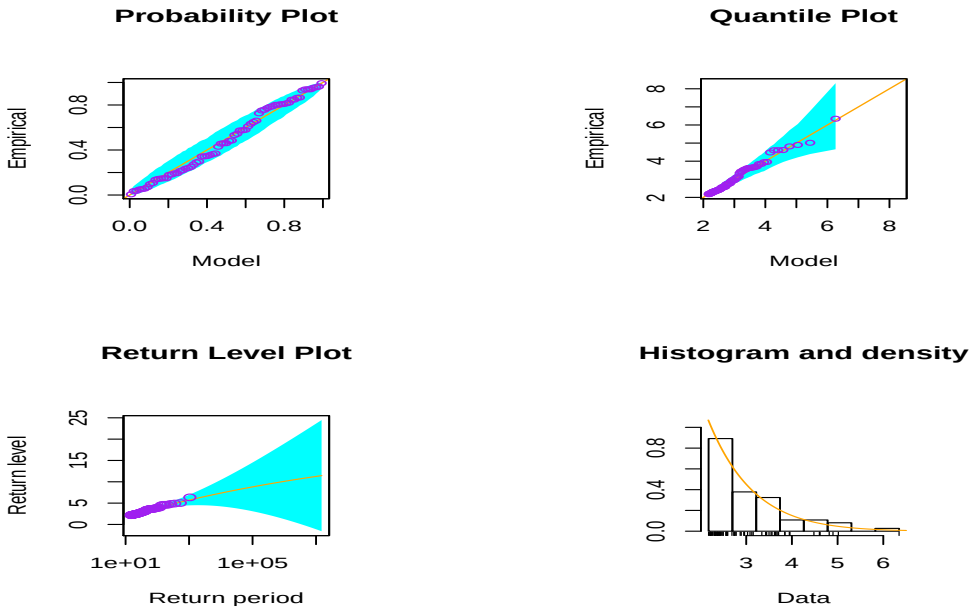


Figure 5.8: IBOV: GPD diagnostic plots.

the IBOV market. The tested hypothesis are:

$$H_0 : \text{GPD fits the cluster-maxima well}$$
$$H_1 : \text{GPD does not fit the cluster-maxima well}$$

Table 5.4 displays the outcomes of the test statistic and p-values of the tested hypothesis of the Anderson-Darling and Cramér-von Mises tests. For both goodness of fit tests, it is observed that the null hypothesis is not rejected since the p-value is large (greater than 0.05). This indicates that the GPD fits the generated cluster-maxima observations well.

Table 5.5 shows the return level ($\ddot{R}_g$) estimates and their corresponding confidence intervals (in parentheses) for the Brazilian IBOV index. The confidence intervals at 90%, 95%, and 99% are included to measure the uncertainty surrounding the $\ddot{R}_g$

Table 5.4: IBOV: Goodness of fit test.

Test	GPD values	Statistic values	P-value
Anderson-Darling	σ: 0.94 ξ: -0.05	0.4209	0.4788
Cramér-von Mises	σ: 0.94 ξ: -0.05	0.0559	0.5126

Table 5.5: IBOV: GPD return level ($\ddot{R}_g$) estimates.

$u = 2.17$	2-year	5-year	10-year	20-year	50-year
$\widehat{\ddot{R}_g}$	7.41	8.00	8.44	8.86	9.39
90% CI	(4.73; 10.09)	(4.51; 11.50)	(4.27; 12.61)	(3.97; 13.75)	(3.49; 15.29)
95% CI	(4.21; 10.60)	(3.84; 12.17)	(3.47; 13.41)	(3.03; 14.68)	(2.36; 16.42)
99% CI	(3.21; 11.60)	(2.53; 13.48)	(1.91; 14.97)	(1.21; 16.51)	(0.16; 18.63)

estimates. The table shows that a maximum loss of 7.41% is expected once every 2 years, 8% once every 5 years, 8.44% once every 10 years, 8.86% once every 20 years, and 9.39% once every 50 years.

5.2.11 Univariate analysis: Point process

For the univariate analysis of the point process, the same procedure as that of the GPD fit by the CEV model is followed. To enable appropriate comparison between these two models, the same threshold is used. Hence, for the parameter estimation of the point process model, the Poisson process is fitted to the same declustered exceedances (i.e. the cluster-maxima) used for the GPD fit.

As it is under the GPD parameters estimation, the 71 cluster-maxima are different magnitudes or levels of risk in the Brazilian equity market at threshold $u = 2.17$. The point process is fitted to these 71 cluster-maxima and the results in Table 5.6 are obtained. The standard errors of the estimates are enclosed in parentheses while the confidence intervals are in the brackets. The results obtained for the shape parameter ξ are very similar to that of the GPD estimation with negligible differences. The shape

Table 5.6: IBOV: Univariate point process parameter estimates.

IBOV	CI	u	μ	$\hat{\xi}$	$\hat{\sigma}$	Log. lik
	90%	2.17	6.93 (1.27) [4.84; 9.03]	-0.05 (0.14) [-0.28; 0.17]	0.69 (0.47) [-0.08; 1.46]	285.32
	95%	2.17	6.93 (1.27) [4.44; 9.43]	-0.05 (0.14) [-0.32; 0.22]	0.69 (0.47) [-0.22; 1.61]	285.32
	99%	2.17	6.93 (1.27) [3.66; 10.21]	-0.05 (0.14) [-0.40; 0.30]	0.69 (0.47) [-0.51; 1.90]	285.32

Table 5.7: Likelihood ratio test for PP's ξ estimate.

IBOV index	w	$\hat{\xi}$	LR	CV	LR_{**}	CV
	71	-0.05	570.64	10%: 2.706 5%: 3.841 1%: 6.635	548.14	10%: 2.706 5%: 3.841 1%: 6.635

parameter estimate is negative ($\xi < 0$), and that reflects convexity (Coles, 2001), indicating a short tailed distribution. However, since the range of the confidence intervals of the shape parameter ξ in Table 5.6 at the three confidence levels is from negative to positive values including zero, a formal hypothesis test is carried out in Section 5.2.12 using the likelihood ratio (LR) and the modified likelihood ratio (LR_{**}) tests to determine if $\xi = 0$ or otherwise.

5.2.12 Formal hypothesis test

From the outcome in Table 5.7, at the 1%, 5% and 10% levels of significance, the values of the likelihood ratio (LR) and the modified likelihood ratio (LR_{**}) tests are greater than the critical values. Hence, the null hypothesis is rejected, inferring that the shape parameter $\xi < 0$. This suggests a Weibull domain of attraction for the data, hence the risk in the Brazilian IBOV market using the point process approach can be modelled by the Weibull class of distributions, and the data is bounded from above.

Table 5.8: IBOV: Point process return level (R_g) estimates.

$u = 2.17$	2-year	5-year	10-year	20-year	50-year
$\widehat{R_g}$	7.19	7.93	8.41	8.84	9.38
90% CI	(4.79; 9.58)	(4.56; 11.30)	(4.32; 12.49)	(4.02; 13.67)	(3.54; 15.23)
95% CI	(4.33; 10.04)	(3.92; 11.95)	(3.54; 13.28)	(3.10; 14.59)	(2.42; 16.35)
99% CI	(3.43; 10.94)	(2.66; 13.21)	(2.01; 14.80)	(1.29; 16.39)	(0.24; 18.53)

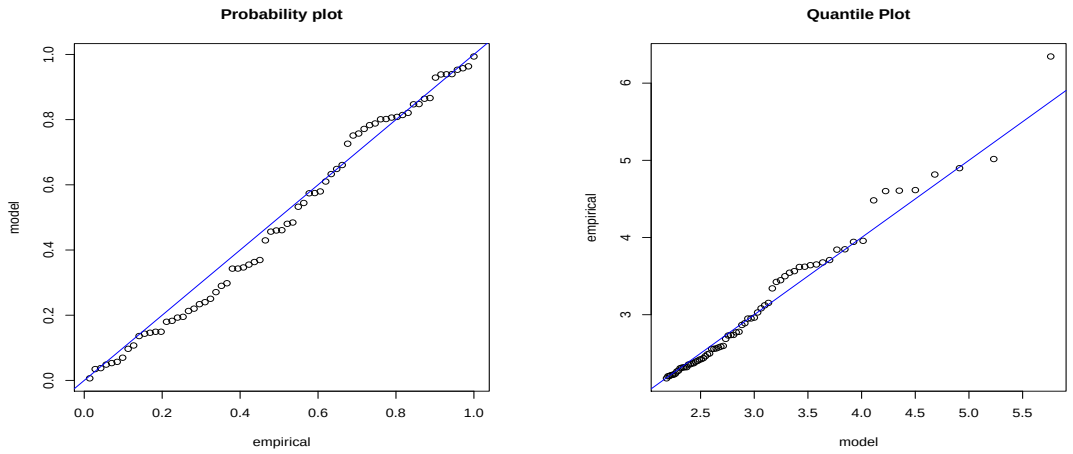


Figure 5.9: IBOV: Point process diagnostic plots.

Table 5.9: PINAW values for Avg heuristic.

IBOV index	GPD	PP	Avg combined PIs
Confidence levels	PINAW	PINAW	PINAW
90%	1.3127	1.1896	1.2490
95%	1.3134	1.1898	1.2495
99%	1.3129	1.1918	1.2503

Table 5.10: PINAW values for Env heuristic.

IBOV index	GPD	PP	Env combined PIs
Confidence levels	PINAW	PINAW	PINAW
90%	1.3127	1.1896	1.3127
95%	1.3134	1.1898	1.3134
99%	1.3129	1.1918	1.3129

As displayed in Figure 5.9, the data points in the diagnostic probability plot and quantile plot of the point process are close to the 45° lines, except for the outlier at the extreme of the quantile plot. Hence, the point process fit is satisfactory. The outlier occurred on the 17th March 2016; it is an adverse market reaction that was triggered by a political event that took place as a result of the 2016 impeachment of the then Brazilian president, Ms. Dilma Rousseff, which lead to an abnormal returns in the Brazilian market (Marques, 2016; Batista et al., 2018).

Table 5.8 shows the point process return level estimates and their corresponding confidence intervals (in parentheses) for the IBOV index. The table shows that a maximum loss of 7.19% is expected once every 2 years, 7.93% once every 5 years, 8.41% once every 10 years, 8.84% once every 20 years, and 9.38% once every 50 years. The return level plot is displayed in panel (a) of Figure 7.18 (see Appendix). The results of the return level estimates and the interval forecasts for the point process are similar to that of the GPD in Table 5.5.

5.2.13 Combining the GPD and PP return levels' interval forecasts

This section compares the prediction intervals (PIs), which is also known as the interval forecasts, of the two individual models (the GDP and point process) with PIs from the combined interval forecast using the PINAW metric as the evaluator. The PIs with the lowest or narrowest PINAW from any of the three, i.e. the GDP, point process and combined PIs, is taken as the best interval forecast.

The PINAW tables for the simple average (Ave) and the envelop (Env) are shown in Tables 5.9 and 5.10. The outcome of the PINAW for the Median (Med) is the same as that of the simple average (Ave) since there are only two models or forecasters, hence it is not displayed. For both the simple average and envelop heuristics in tables, the point process model outperforms the combined PIs approach and the GPD model with the narrowest PINAW values at the three confidence levels. Hence, the point process model is the best model in constructing PIs for the Brazilian IBOV market's return levels forecast.

5.3 The Russian market: IMOEX

The Russian market index is called the "Moscow Exchange index", abbreviated - IMOEX, and it contains 2126 observations for the periods 5^{th} January 2010 to 6^{th} August 2018.

5.3.1 IMOEX: Positive residual observations

The modelling of the risk in each of the markets is restricted to the positive residuals of the entire equity observations. From the original equity data with 2126 number of observations, there are 1034 positive residual observations in this market as displayed

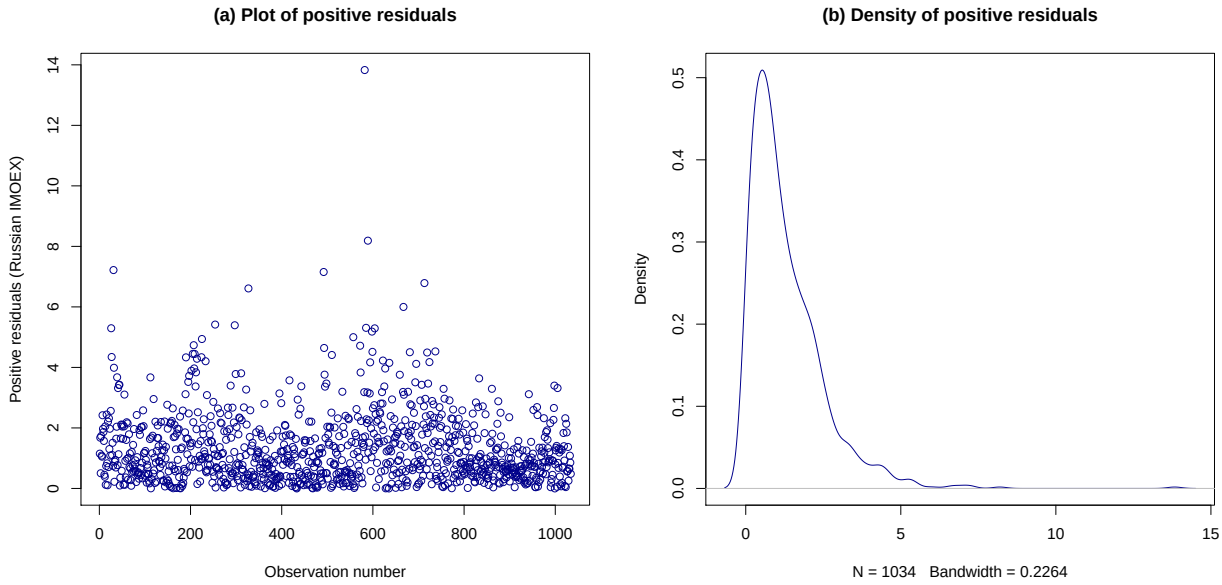


Figure 5.10: IMOEX: Positive residuals.

in Figure 5.10. The positive residual observations and their corresponding density are given in Figure 5.10, panels (a) and (b) respectively.

5.3.2 IMOEX: Threshold selection

The same threshold selection procedure used for the Brazilian IBOV market is applied for the Russian IMOEX market, i.e. the non-parametric Kernel-GPD approach. The bulk model under the threshold is the standard kernel density estimator and the tail model is a GPD above the threshold.

Following the threshold selection steps of Hu and Scarrott (2018), the combined plots of the bulk model based (blue solid line) and the parameterised tail fraction (red solid line) approaches are displayed in Figure 5.11. The figure shows the output of the bulk model based tail fraction with threshold $u = 2.6890$, and that of the parameterised tail fraction approach at $u = 3.0847$.

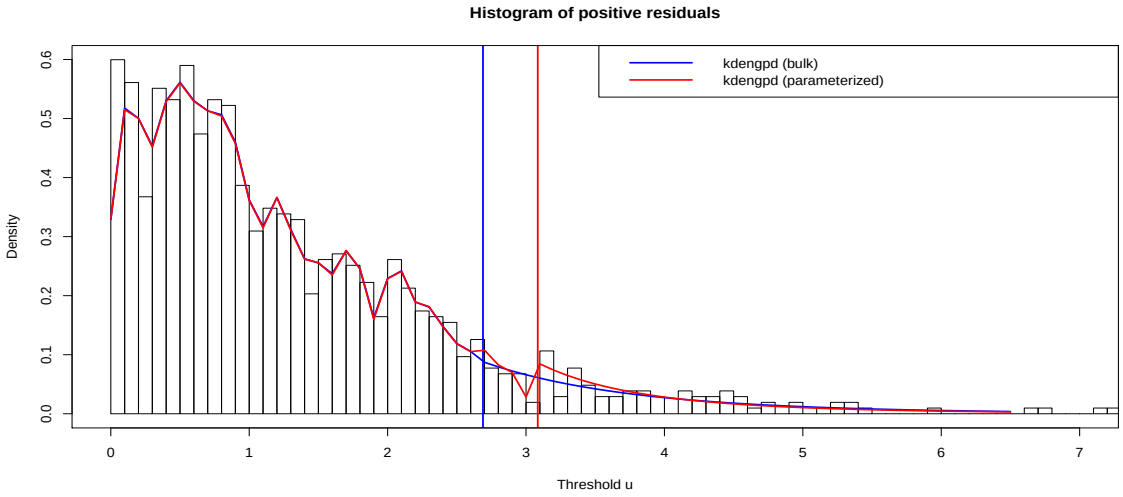


Figure 5.11: IMOEX: Threshold selection.

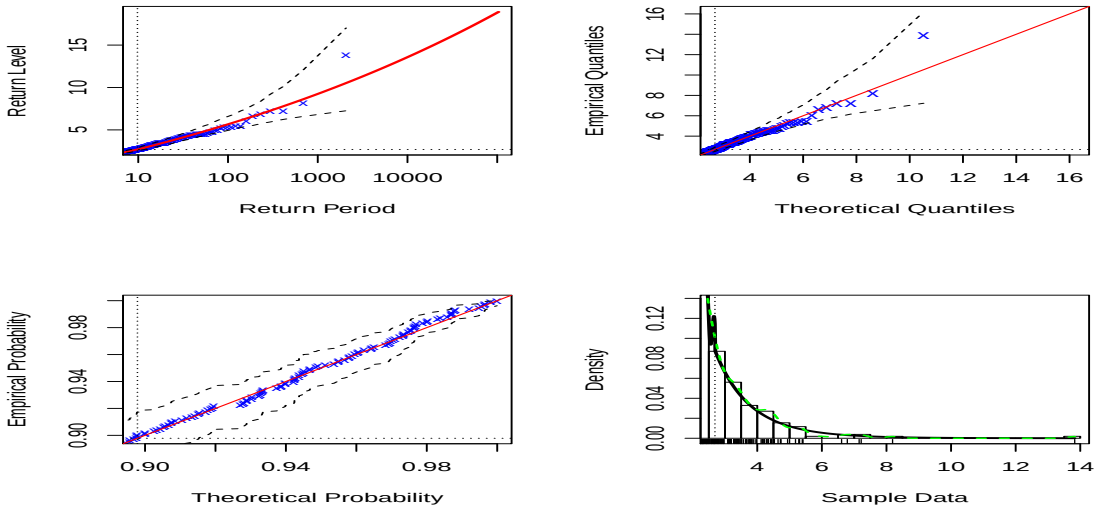


Figure 5.12: IMOEX: Diagnostic plots of the bulk model based tail fraction.

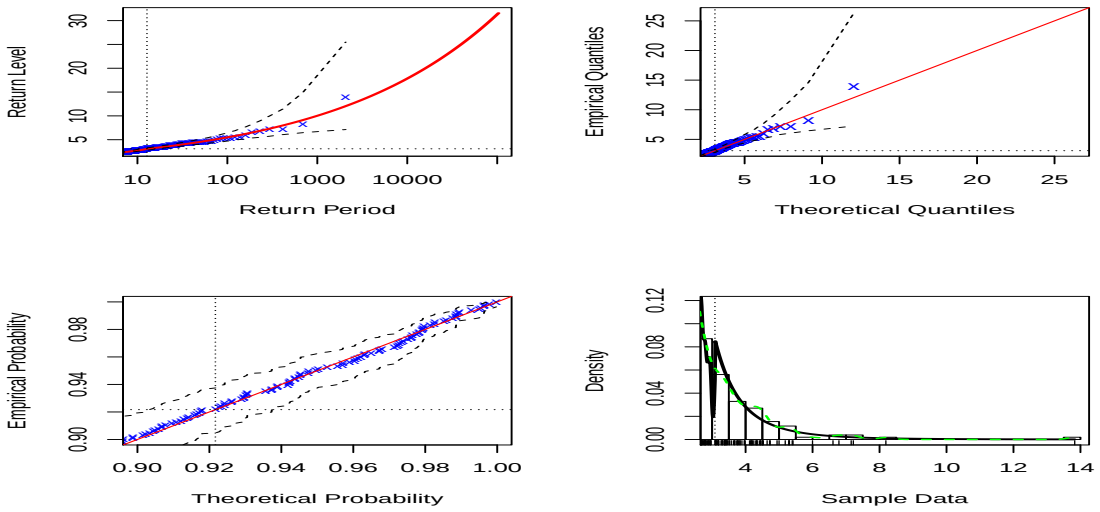


Figure 5.13: IMOEX: Diagnostic plots of the parameterised tail fraction.

5.3.3 Diagnostic plots of the bulk and parameterised approaches

The various diagnostic plots for assessing the goodness of each of the tail fraction approaches are shown in Figures 5.12 and 5.13. Figure 5.13 shows an improvement in the plots of the parameterised tail fraction when compared to that of the bulk model based in Figure 5.12. The data points are more closely aligned on the diagonal lines of the return level, quantiles and probability plots than they are in the bulk model based approach. The parameterised tail fraction approach makes provision for an extra degree of freedom that is used to re-scale both the bulk and tail components in equation (3.19.2) (Hu and Scarrott, 2018). This enhances and improves the tail fit, since the tail fraction is estimated from the sample fraction of exceedances. Hence, a better fit below an estimated threshold is allowed by the mixture because of the re-scaling of the bulk density. This is not the case with the bulk model based tail fraction because it does not adjust the density below the threshold in equation (3.19.1) (Hu and Scarrott, 2018).

However, none of the diagnostic plots gives any cause to cast doubt on the validity of the fitted bulk model based and parameterised tail fraction approaches for the threshold selection. This indicates that the non-parametric Kernel-GPD mixture models is satisfactory.

5.3.4 Sensitivity analysis

Since the diagnostic plots of the parameterised tail fraction approach show more levels of linearity than the bulk model based, sensitivity analysis is done starting with its threshold value $u = 3.09$ as displayed in Figure 5.14. From the figure, a piecewise-linearity of the points are most observed at the plot with threshold $u = 3.09$ than at the other plots. This threshold gives more asymptotic information about the tail

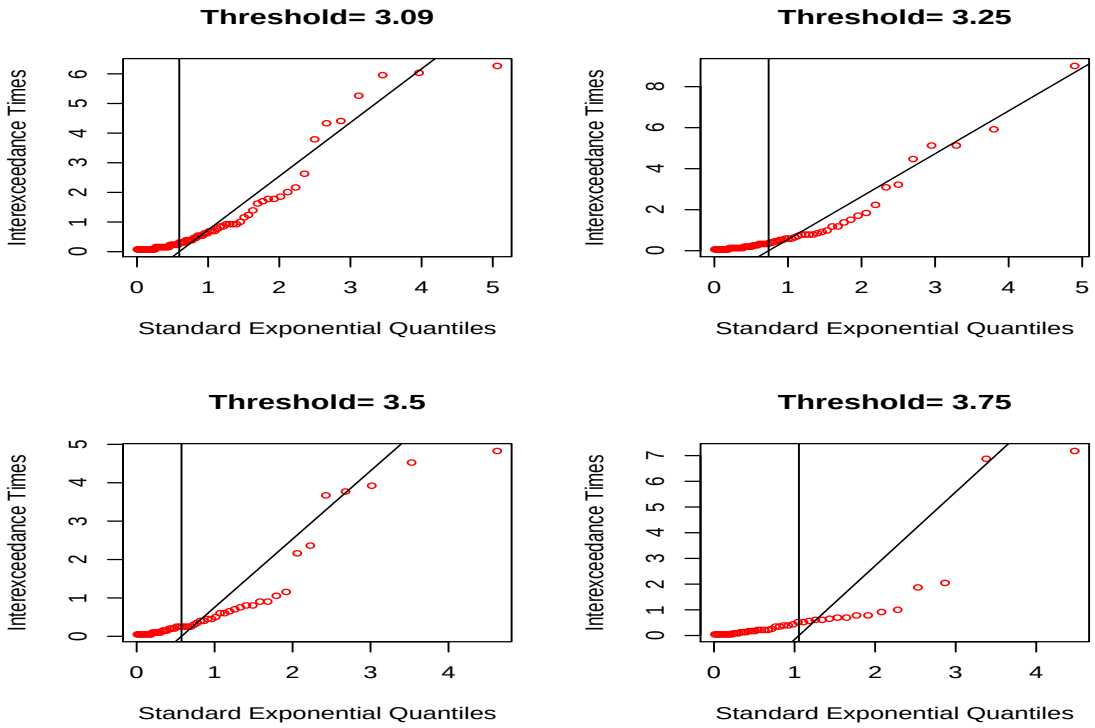


Figure 5.14: IMOEX: Sensitivity analysis plots for threshold selection.

(than the other thresholds) with 80 threshold exceedances and extremal index $\hat{\theta}$ of 0.5521. This means that exceedances occur in groups of $\frac{1}{0.5521} = 1.8113 \approx 2$.

5.3.5 Shape threshold stability plot

The shape threshold stability plot is used in this study to verify the outcome of the threshold choice from the mixture models. As shown in Figure 5.15, three potential thresholds where the plot shows significant departures from linearity are identified around $u = 1.5$, $u = 2$ and $u = 3$. Hence, the threshold estimate ($u = 3.09$) from the Kernel density mixture models is approximately consistent with the choice of $u = 3$ from the shape threshold stability plot.

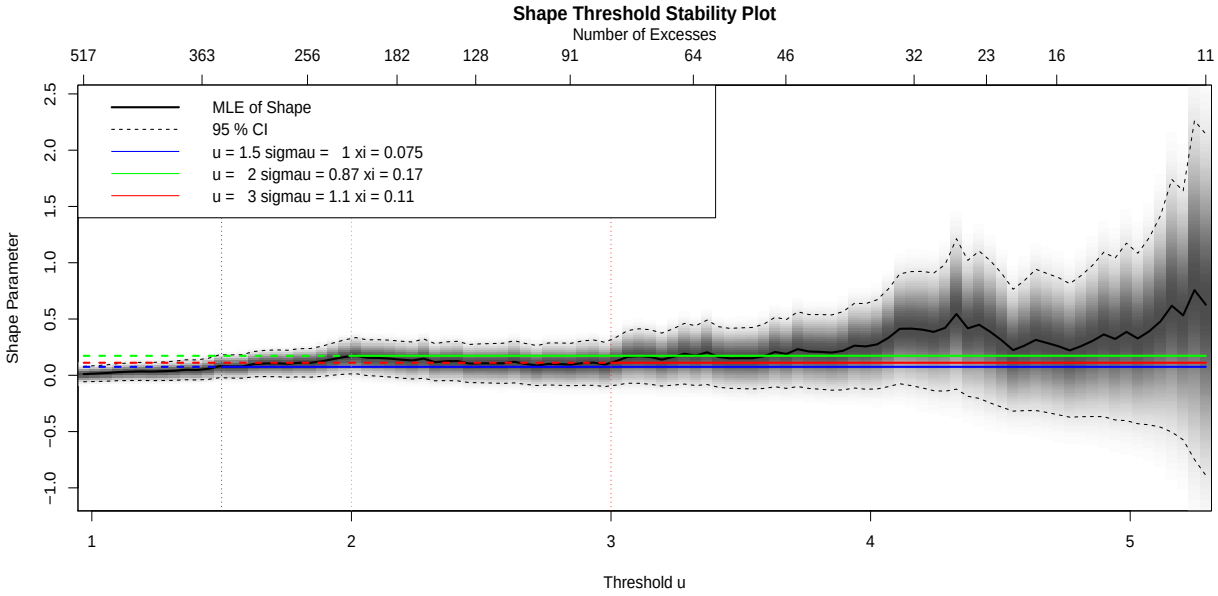


Figure 5.15: IMOEX: Shape threshold stability plot.

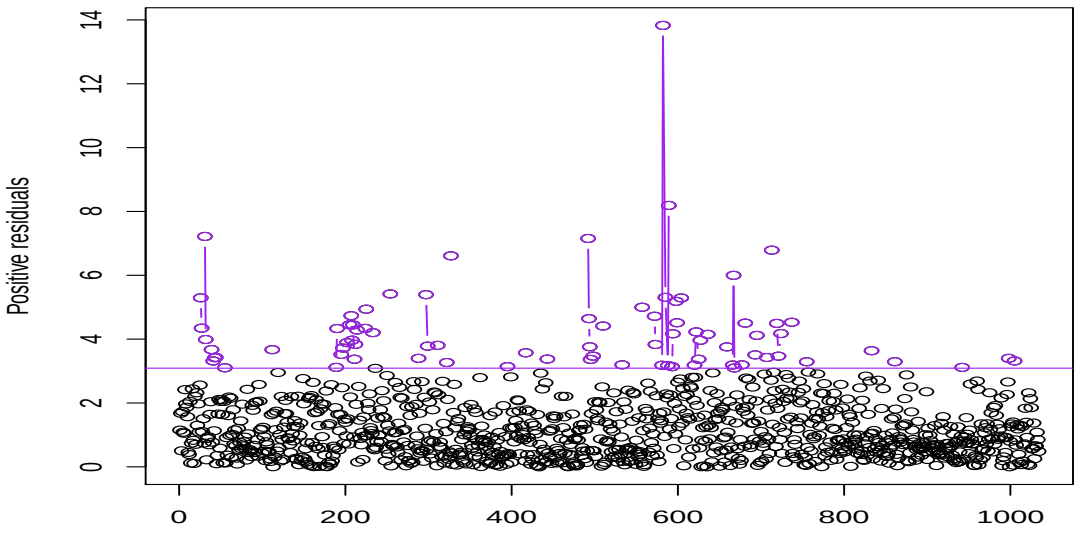


Figure 5.16: IMOEX: Declustered exceedances at $u = 3.09$ (cluster-maxima).

Table 5.11: IMOEX: GPD parameter estimates.

	CI	u	n	w	$\eta_u = \frac{w}{n}$	$\hat{\xi}$	$\hat{\sigma}$	Log. lik
	90%	3.09	1034	45	0.0435	0.14 (0.16)	1.25 (0.22)	-61.63
						[-0.12; 0.40]	[0.90; 1.61]	
IMOEX	95%	3.09	1034	45	0.0435	0.14 (0.16)	1.25 (0.22)	-61.63
						[-0.17; 0.45]	[0.83; 1.68]	
	99%	3.09	1034	45	0.0435	0.14 (0.16)	1.25 (0.22)	-61.63
						[-0.26; 0.55]	[0.70; 1.81]	

Table 5.12: IMOEX: GPD parameter estimates and profile likelihood intervals.

	CI	u	n	w	$\eta_u = \frac{w}{n}$	$\hat{\xi}$	$\hat{\sigma}$	Log. lik
	90%	3.09	1034	45	0.0435	0.14 (0.16)	1.25 (0.22)	-61.63
						[-0.05; 0.50]	[0.86; 1.77]	
IMOEX	95%	3.09	1034	45	0.0435	0.14 (0.16)	1.25 (0.22)	-61.63
						[-0.08; 0.59]	[0.79; 1.89]	
	99%	3.09	1034	45	0.0435	0.14 (0.16)	1.25 (0.22)	-61.63
						[-0.11; 0.78]	[0.68; 2.14]	

5.3.6 Declustering

The estimated threshold $u = 3.09$ generated 80 threshold exceedances. After declustering at this threshold, 45 cluster-maxima are obtained as shown in Figure 5.16. The cluster-maxima are "exceedance residuals" (McNeil and Frey, 2000). They are the maxima of the clusters of exceedances over a high threshold, and are the independent extreme tail observations on which the GPD is fitted.

5.3.7 CEV model: GPD fit to cluster-maxima

The GPD from the conditional extreme value (CEV) model will now be fitted to the 45 cluster-maxima that are generated after declustering the cluster exceendances. These cluster-maxima are collectively classified as the risk, which are the extreme losses, to be modelled in the Russian IMOEX market. The risk is modelled by fitting the GPD on the 45 observations and the result in Table 5.11 is obtained. The positive estimate ($\hat{\xi} = 0.14$) of the shape parameter as shown in the table denotes a heavy or

Table 5.13: Likelihood ratio test for GPD's ξ estimate.

IMOEX index	w	$\hat{\xi}$	LR	CV	LR_{**}	CV
	45	0.14	-123.26	10%: 2.706 5%: 3.841 1%: 6.635	-115.59	10%: 2.706 5%: 3.841 1%: 6.635

fat-tailed distribution and it is a reflection of concavity (Coles, 2001).

An alternative approach to computing the confidence intervals for the shape parameter (ξ) with better accuracy is through the use of the profile likelihood (Coles 2001). Table 5.12 shows the results of the GPD parameters estimation and their corresponding profile log-likelihood confidence intervals. The plots of the scale parameter's returned intervals at 90%, 95% and 99% confidence intervals, and that of the shape parameter at 90% and 95% confidence intervals are displayed in Figures 7.19 and 7.20 respectively (in the Appendix).

Since the values of the confidence intervals of the shape parameter ξ at the three confidence levels in the tables are from negative to positive including zero, a formal hypothesis test is carried out to determine if $\xi = 0$ or otherwise. From the result in Table 5.13, at the 1%, 5% and 10% levels of significance, the values of the likelihood ratio (LR) and the modified likelihood ratio (LR_{**}) tests are less than the critical values. Hence, we fail to reject the null hypothesis and conclude that the shape parameter $\xi = 0$. This suggests Gumbel domain of attraction for the data, hence the risk in the Russian IMOEX market can be modelled by the Gumbel class of distributions.

5.3.8 Diagnostics: Model checking

The correctness of the fitted GPD model is checked by the diagnostic plots in Figure 5.17. The set of points on the probability plot is near-linear, while those on the quantile plot are fairly aligned within the tolerance interval curve. The return level

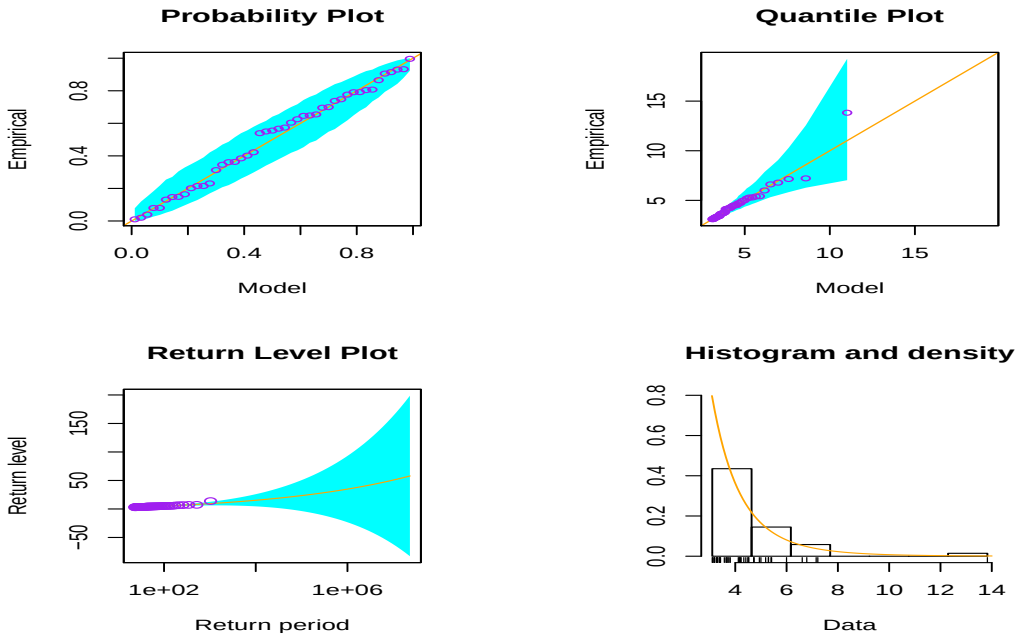


Figure 5.17: IMOEX: GPD diagnostic plots.

plot displays points that are well aligned on the straight line, however it shows very large uncertainties that enlarge at extrapolation of the model to higher levels. The density estimates further shows consistency with the data's histogram. Hence, the quartet model fit diagnostics reasonably support the fitted GPD.

The Anderson-Darling and Cramér-von Mises goodness of fit tests are further used to determine how well the GPD model fits the cluster-maxima observations in the IMOEX market. The tested hypothesis are:

$$H_0 : \text{GPD fits the cluster-maxima well}$$
$$H_1 : \text{GPD does not fit the cluster-maxima well}$$

Table 5.14 displays the outcomes of the test statistic and p-values of the tested hypothesis of the Anderson-Darling and Cramér-von Mises tests. For both goodness

Table 5.14: IMOEX: Goodness of fit test.

Test	GPD values	Statistic values	P-value
Anderson-Darling	σ: 1.25 ξ: 0.14	0.3384	0.6136
Cramér-von Mises	σ: 1.25 ξ: 0.14	0.0420	0.6673

Table 5.15: IMOEX: GPD return level ($\ddot{R}_g$) estimates.

$u = 3.09$	2-year	5-year	10-year	20-year	50-year
$\widehat{\ddot{R}_g}$	16.87	20.03	22.72	25.69	30.10
90% CI	(5.58; 28.15)	(3.60; 36.47)	(1.36; 44.09)	(-1.66; 53.05)	(-7.07; 67.28)
95% CI	(3.42; 30.32)	(0.45; 39.61)	(-2.74; 48.18)	(-6.90; 58.29)	(-14.19; 74.40)
99% CI	(-0.80; 34.54)	(-5.69; 45.76)	(-10.72; 56.17)	(-17.13; 68.51)	(-28.09; 88.29)

of fit tests, it is observed that the null hypothesis is not rejected since the p-value is large (greater than 0.05). This indicates that the GPD fits the generated cluster-maxima observations well.

Table 5.15 displays the estimates of the return levels and their corresponding confidence intervals (in parentheses) for the Russian IMOEX index. The table shows that a maximum loss of 16.87% is expected once every 2 years, 20.03% once every 5 years, 22.72% once every 10 years, 25.69% once every 20 years, and 30.10% once every 50 years.

5.3.9 Univariate analysis: Point process

The procedure used for the GPD fit under the CEV model is also applied to the univariate analysis of the point process. To enable appropriate comparison between the two models, the same threshold is used. Hence, for the parameter estimation of the point process model, the Poisson process is fitted to the same declustered exceedances (i.e. the 45 cluster-maxima) used for the GPD fit.

The 45 cluster-maxima are different levels of the risk in the Russian equity market

Table 5.16: IMOEX: Univariate point process parameter estimates.

IMOEX	CI	u	μ	$\hat{\xi}$	$\hat{\sigma}$	Log. lik
	90%	3.09	14.73 (4.82) [6.80; 22.66]	0.14 (0.15) [-0.10; 0.39]	2.93 (2.27) [-0.81; 6.67]	158.86
	95%	3.09	14.73 (4.82) [5.29; 24.18]	0.14 (0.15) [-0.15; 0.44]	2.93 (2.27) [-1.53; 7.39]	158.86
	99%	3.09	14.73 (4.82) [2.32; 27.15]	0.14 (0.15) [-0.24; 0.53]	2.93 (2.27) [-2.93; 8.78]	158.86

Table 5.17: Likelihood ratio test for PP's ξ estimate.

IMOEX index	w	$\hat{\xi}$	LR	CV	LR_{**}	CV
	45	0.14	317.72	10%: 2.706 5%: 3.841 1%: 6.635	297.95	10%: 2.706 5%: 3.841 1%: 6.635

at threshold $u = 3.09$. The point process is fitted to these cluster-maxima and the results in Table 5.16 are obtained. The standard errors of the estimates are enclosed in parentheses while the confidence intervals are in the brackets. The results obtained for the shape parameter ξ are very similar to that of the GPD estimation with negligible differences. The positive estimate ($\hat{\xi} = 0.14$) of the shape parameter is a reflection of concavity (Coles, 2001) and it indicates a fat-tailed distribution.

The confidence intervals of the shape parameter ξ in Table 5.16 at the three confidence levels take values from negative to positive values including zero, hence a formal hypothesis test is required to determine if $\xi = 0$ or otherwise. From the outcome in Table 5.17, at the 1%, 5% and 10% levels of significance, the values of the likelihood ratio (LR) and the modified likelihood ratio (LR_{**}) tests are greater than the critical values. Hence, the null hypothesis is rejected, implying that the shape parameter $\xi > 0$. This positive shape parameter estimate ($\hat{\xi} = 0.14$) as shown in the table corresponds to an unbounded distribution, hence the risk in the Russian IMOEX market using the point process approach can be described by the Fréchet-Pareto class of distributions (see Beirlant et. al, 2004).

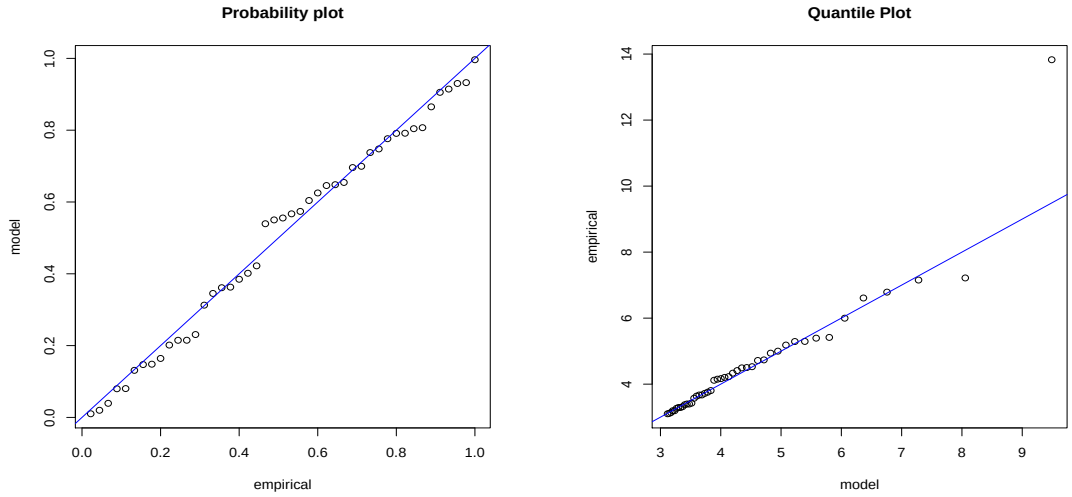


Figure 5.18: IMOEX: Point process diagnostic plots.

As shown in Figure 5.18, the data points in the diagnostic probability plot and quantile plot of the point process are close to the 45° lines, except for the lone outlier at the extreme of the quantile plot. Hence, the point process fit is reasonable. The outlier occurred on the 9th April 2010 and it could be a result of the Russian economic recession that spanned from 2008 to 2009 (Bloomberg, 2009), which likely rippled into early months of 2010. The recession plummeted the Russian markets with over $1 trillion shares losses incurred (Reuters, 2008).

Table 5.18 shows the point process return level estimates and their corresponding confidence intervals (in parentheses) for the IMOEX index. The table shows that a maximum loss of 16.87% is expected once every 2 years, 20.03% once every 5 years, 22.72% once every 10 years, 25.69% once every 20 years, and 30.10% once every 50 years. The return level plot is displayed in panel (b) of Figure 7.18 in the Appendix. The large confidence intervals that are obtained in the table and displayed in the figure for extreme return levels indicates that there is no much information with which to make future forecasts with any degree of certainty (Coles, 2001) in the Russian IMOEX market. When compared with the GPD's return levels results in

Table 5.18: IMOEX: Point process return level (R_g) estimates.

$u = 3.09$	2-year	5-year	10-year	20-year	50-year
$\widehat{R_g}$	15.84	19.65	22.53	25.60	30.08
90% CI	(5.93; 25.75)	(3.72; 35.57)	(1.33; 43.73)	(-1.81; 53.01)	(-7.37; 67.53)
95% CI	(4.04; 27.65)	(0.67; 38.62)	(-2.73; 47.78)	(-7.06; 58.26)	(-14.55; 74.71)
99% CI	(0.33; 31.36)	(-5.29; 44.58)	(-10.65; 55.71)	(-17.31; 68.51)	(-28.55; 88.71)

Table 5.19: PINAW values for Avg heuristic.

IMOEX index	GPD	PP	Avg combined PIs
Confidence levels	PINAW	PINAW	PINAW
90%	0.8777	0.8126	0.8441
95%	0.8778	0.8123	0.8441
99%	0.8778	0.8126	0.8442

Table 5.20: PINAW values for Env heuristic.

IMOEX index	GPD	PP	Env combined PIs
Confidence levels	PINAW	PINAW	PINAW
90%	0.8777	0.8126	0.8712
95%	0.8778	0.8123	0.8710
99%	0.8778	0.8126	0.8710

Table 5.15, both models give similar estimates.

5.3.10 Combining the GPD and PP return levels' interval forecasts

The PINAW tables for the simple average (Ave) and the envelop (Env) are shown in Tables 5.19 and 5.20. For both the simple average and envelop heuristics in tables, the point process model outperforms the combined PIs approach and the GPD model with the narrowest PINAW outcomes at the three confidence levels. Hence, the point process model is the best model in constructing PIs for the Russian IMOEX market's return levels forecast.

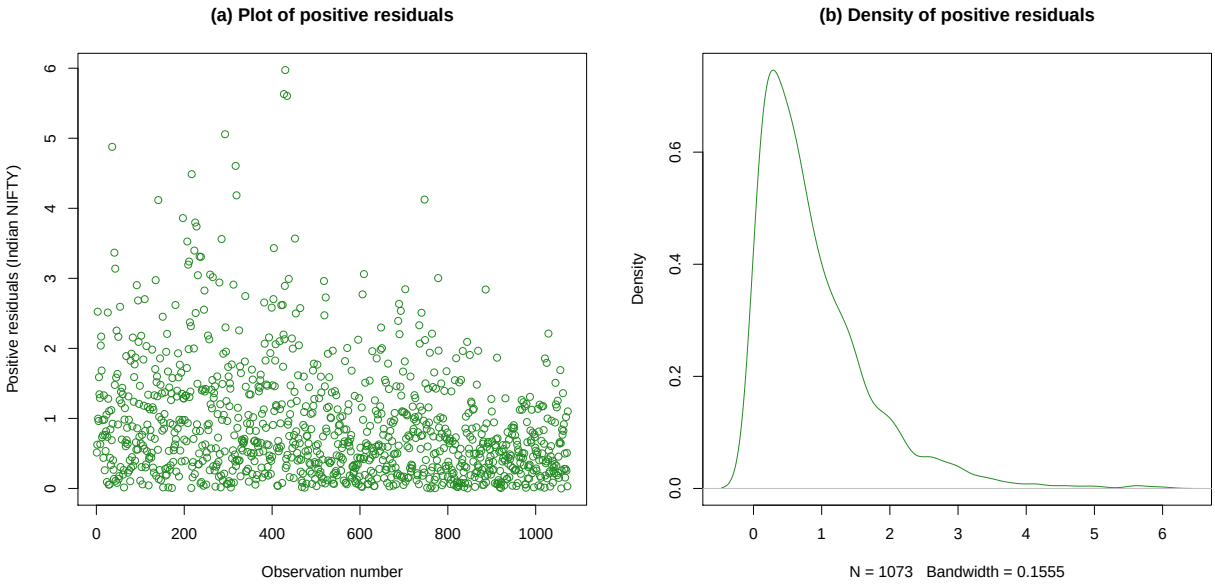


Figure 5.19: NIFTY: Positive residuals.

5.4 The Indian market: NIFTY

The Indian market index is called the "NIFTY 50 index", abbreviated - NIFTY, and it contains 2126 observations for the periods 5^{th} January 2010 to 6^{th} August 2018.

5.4.1 NIFTY: Positive residual observations

The modelling of the risk in each of the markets is restricted to the positive residuals of the entire equity observations. From the original equity data with 2126 number of observations, there are 1073 positive residual observations in this market as shown in Figure 5.19. The positive residual observations and their corresponding density are given in Figure 5.19, panels (a) and (b) respectively.

5.4.2 NIFTY: Threshold selection

Like the two preceding markets, the same threshold selection procedure is used for the Indian NIFTY market, i.e. the non-parametric Kernel GPD approach, and the shape threshold stability plot to verify the outcome of the Kernel-GPD approach.

The combined plots of the bulk model based (blue solid line) and the parameterised tail fraction (red solid line) approaches are displayed in Figure 5.20. From the figure, it can be observed that the red colour of the density and threshold line of the parameterised tail fraction almost completely overlapped the blue colour of that of the bulk model based tail fraction. This is because the threshold estimates of the two tail fraction methods are approximately the same, with threshold value $u = 1.9694$ for the bulk model based and $u = 1.9689$ for the parameterised tail fraction approach.

5.4.3 Diagnostic plots of the bulk and parameterised approaches

The two sets of the four diagnostic plots in Figures 5.21 and 5.22 show very strong similarity between the bulk model based and parameterised tail fraction approaches since the threshold values are approximately the same. However, from a closer scrutiny of each set of plots, the diagnostics of the fitted parameterised tail fraction show a very slightly higher closeness and alignment of the data points on the straight lines especially on the probability plot. This therefore gives the parameterised tail fraction a bit of an edge over the bulk model based for the threshold selection in this market.

The diagnostic plots summarily suggest that the fit of the non-parametric Kernel-GPD models is satisfactory based on the consistency of the density estimate with the histogram of the data, and the linearity of the data points in the return level, quantiles and probability plots.

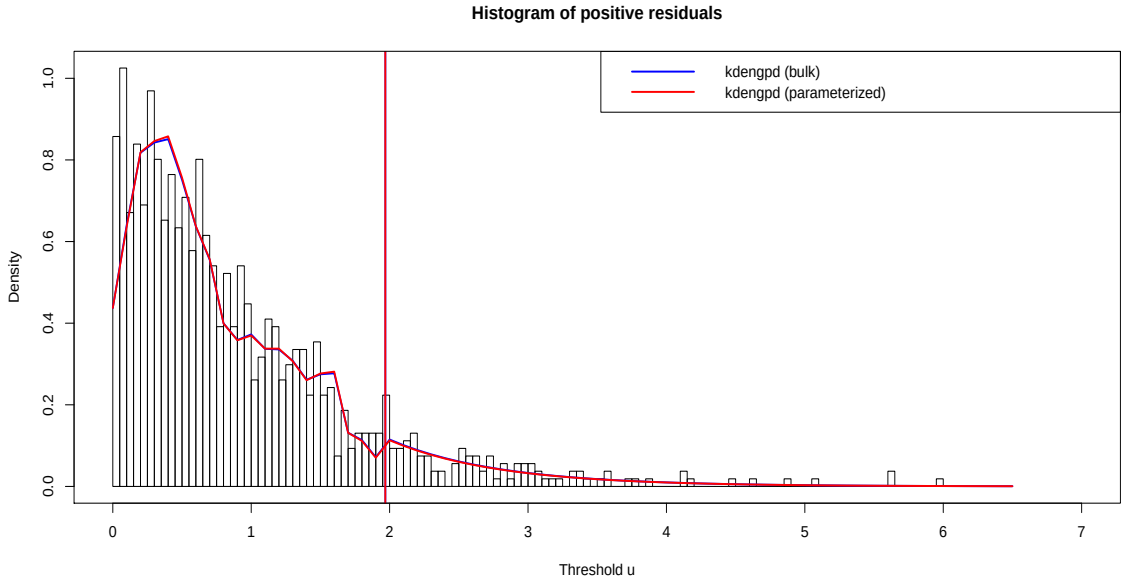


Figure 5.20: NIFTY: Threshold selection.

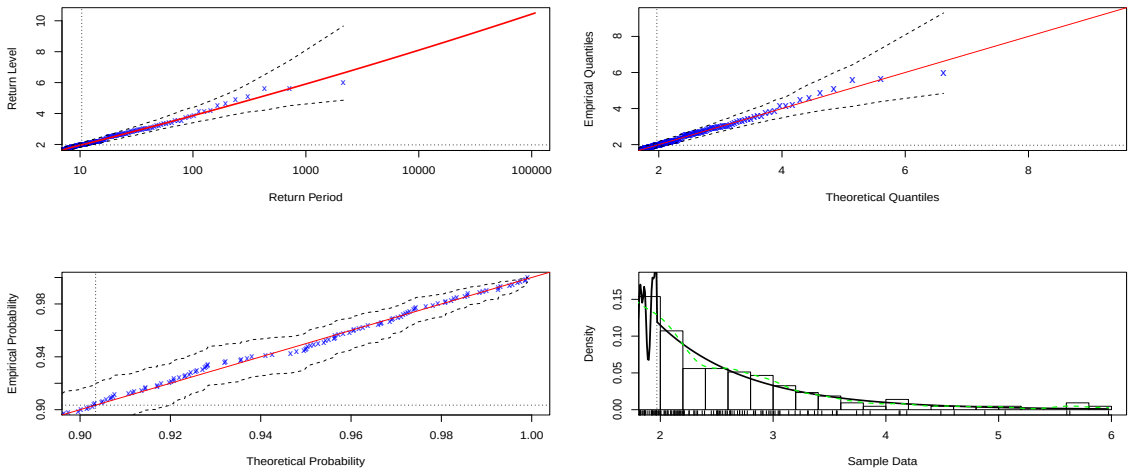


Figure 5.21: NIFTY: Diagnostic plots of the bulk model based tail fraction.

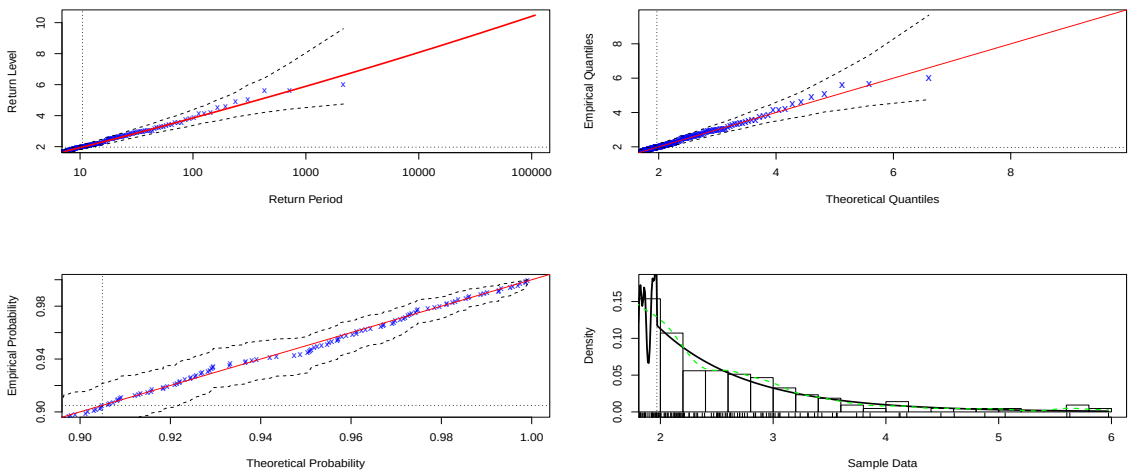


Figure 5.22: NIFTY: Diagnostic plots of the parameterised tail fraction.

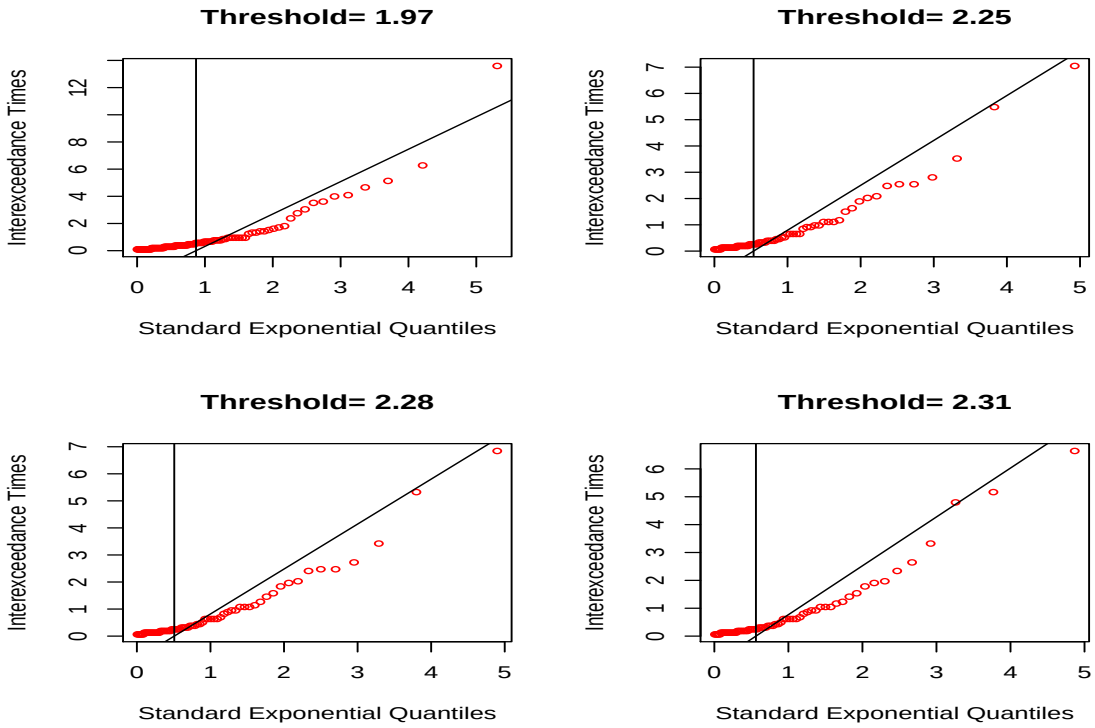


Figure 5.23: NIFTY: Sensitivity analysis plots for threshold selection.

5.4.4 Sensitivity analysis

From the set of sensitivity analysis plots in Figure 5.23, threshold $u = 2.25$ is chosen because it is the best in the set, and it gives more tail information (than the other three plots) with 70 threshold exceedances and extremal index $\hat{\theta} = 0.5852$. It is better than $u = 1.97$ whose extremal index $\hat{\theta} < 0.5$, and it is more informative than thresholds $u = 2.28$ and $u = 2.31$ with 68 and 66 threshold exceedances respectively. An extremal index of 0.5852 means that exceedances occur in groups of $\frac{1}{0.5852} = 1.7088 \approx 2$.

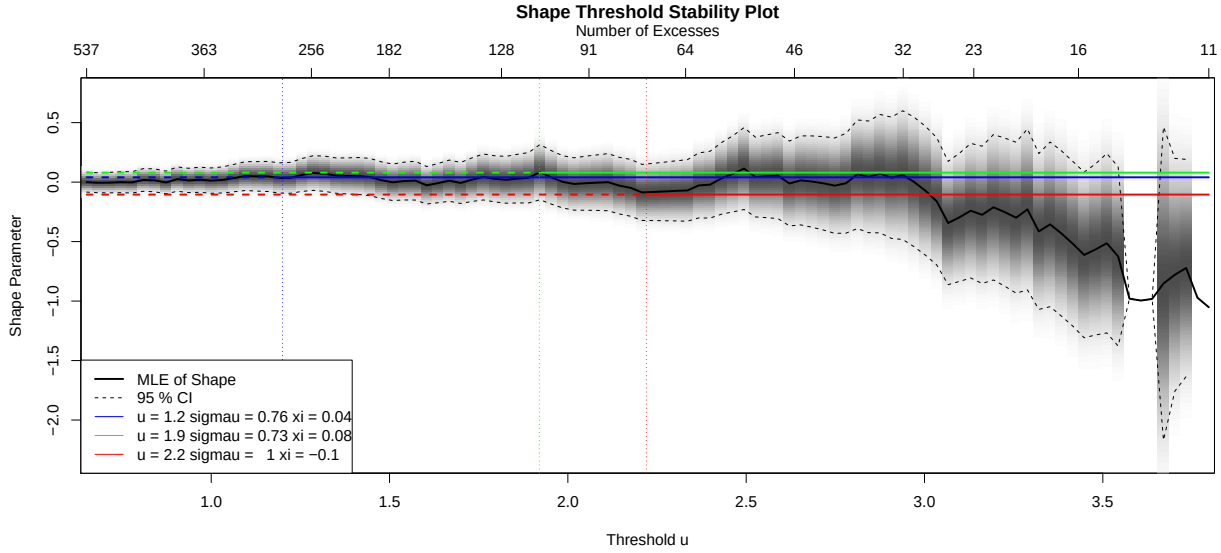


Figure 5.24: NIFTY: Shape threshold stability plot.

5.4.5 Shape threshold stability plot

From the shape threshold stability plot in Figure 5.24, the three potential thresholds identified for significant departures from linearity are $u = 1.2$, $u = 1.9$ and $u = 2.2$. One of these subjective choice of thresholds, i.e. $u = 2.2$ is approximately consistent with the threshold estimate $u = 2.25$ from the Kernel density mixture models. Hence, the objective threshold choice of the Kernel-GPD is a fair tradeoff between variance and bias.

5.4.6 Declustering

The estimated threshold $u = 2.25$ generated 70 threshold exceedances. Upon declustering at this threshold, 39 cluster-maxima are obtained as shown in Figure 5.25. The 39 cluster-maxima are the maxima of the clusters of exceedances over a high threshold ($u = 2.25$ in this case). They are the independent extreme tail observations on which

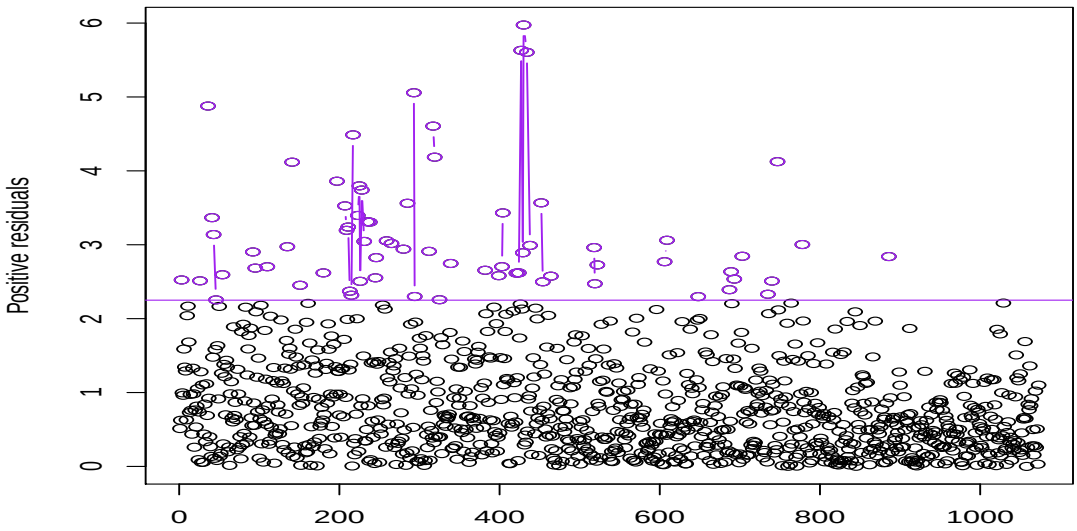


Figure 5.25: NIFTY: Declustered exceedances at $u = 2.25$ (cluster-maxima).

the GPD is fitted, and are also called exceedance residuals (McNeil and Frey, 2000).

5.4.7 CEV model: GPD fit to cluster-maxima

The GPD is now fitted to the 39 declustered exceedances which are the cluster-maxima. The generated cluster-maxima are collectively classified as the risk or extreme losses to be modelled in the Indian NIFTY market. The risk is modelled by fitting the GPD on the 39 observations and the result in Table 5.21 is derived. The shape parameter's negative estimate of $\hat{\xi} = -0.17$ in Table 5.21 indicates a short tailed distribution that reflects convexity (Coles, 2001). This estimate, where $\xi < 0$, corresponds to a bounded distribution.

Better accuracy of confidence intervals can be achieved through the use of profile likelihood (Coles 2001). The approximate profile log-likelihood confidence intervals at 90%, 95% and 99% for the GPD's parameter estimates ($\hat{\xi}$ and $\hat{\sigma}$) are shown in

Table 5.21: NIFTY: GPD parameter estimates.

	CI	u	n	w	$\eta_u = \frac{w}{n}$	$\hat{\xi}$	$\hat{\sigma}$	Log. lik
	90%	2.25	1073	39	0.0364	-0.17 (0.16)	1.14 (0.22)	-37.36
						[-0.43; 0.09]	[0.77; 1.50]	
NIFTY	95%	2.25	1073	39	0.0364	-0.17 (0.16)	1.14 (0.22)	-37.36
						[-0.48; 0.14]	[0.70; 1.57]	
	99%	2.25	1073	39	0.0364	-0.17 (0.16)	1.14 (0.22)	-37.36
						[-0.57; 0.23]	[0.56; 1.71]	

Table 5.22: NIFTY: GPD parameter estimates and profile likelihood intervals.

	CI	u	n	w	$\eta_u = \frac{w}{n}$	$\hat{\xi}$	$\hat{\sigma}$	Log. lik
	90%	2.25	1073	39	0.0364	-0.17 (0.16)	1.14 (0.22)	-37.36
						[-0.30; 0.16]	[0.77; 1.63]	
NIFTY	95%	2.25	1073	39	0.0364	-0.17 (0.16)	1.14 (0.22)	-37.36
						[-0.30; 0.24]	[0.72; 1.74]	
	99%	2.25	1073	39	0.0364	-0.17 (0.16)	1.14 (0.22)	-37.36
						[-0.30; 0.41]	[0.64; 1.98]	

Table 5.23: Likelihood ratio test for GPD's ξ estimate.

NIFTY index	w	$\hat{\xi}$	LR	CV	LR_{**}	CV
	39	-0.17	-74.72	10%: 2.706 5%: 3.841 1%: 6.635	-69.36	10%: 2.706 5%: 3.841 1%: 6.635

Table 5.22. Here, the estimated profile confidence intervals at 90%, 95% and 99% are not very different from the earlier calculation.

The confidence intervals of the shape parameter ξ in the tables at the three confidence levels are from negative to positive values including zero, hence a formal hypothesis test is conducted to ascertain if $\xi = 0$ or otherwise. From the outcome in Table 5.23, at the 1%, 5% and 10% levels of significance, the values of the likelihood ratio (LR) and the modified likelihood ratio (LR_{**}) tests are less than the critical values. Hence, we fail to reject the null hypothesis and conclude that the shape parameter $\xi = 0$. This suggests that the risk in the Indian Nifty market can be modelled by the Gumbel class of distributions.

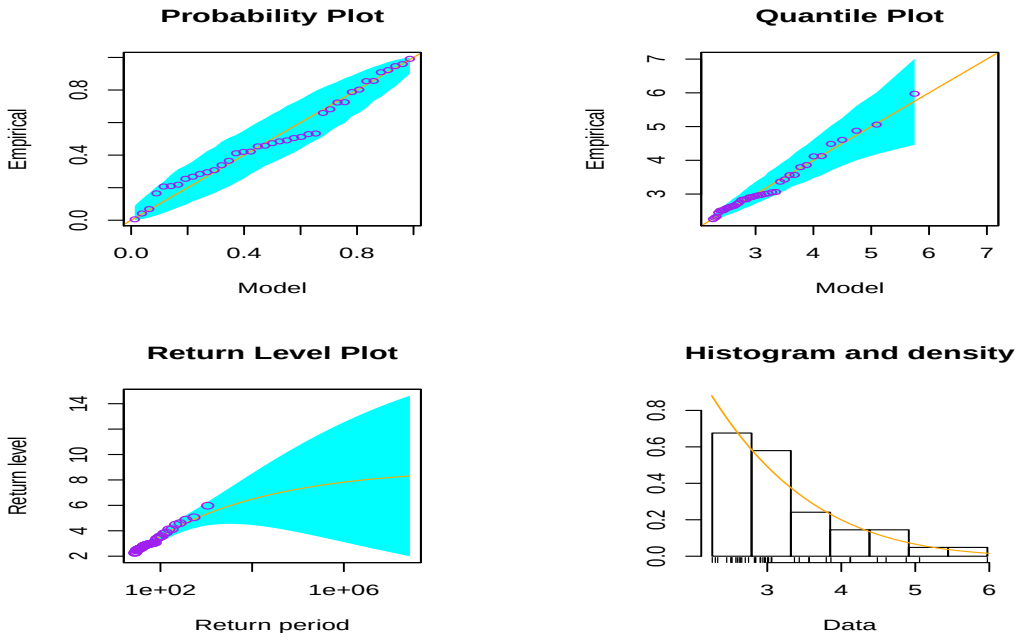


Figure 5.26: NIFTY: GPD diagnostic plots.

The accuracy of the fitted GPD model is judged by the diagnostic plots in Figure 5.26. The set of points on the probability and quantile plots are near-linear, and are aligned within the tolerance interval curves. The return level plot displays points that are well aligned on the straight line, and the density plot also supports the validity of the GPD model's fit.

The Anderson-Darling and Cramér-von Mises goodness of fit tests are further used to know how well the GPD model fits the cluster-maxima exceedances in the NIFTY market. The tested hypothesis are:

$$H_0 : \text{GPD fits the cluster-maxima well}$$
$$H_1 : \text{GPD does not fit the cluster-maxima well}$$

Table 5.24 displays the outcomes of the test statistic and p-values of the tested

Table 5.24: NIFTY: Goodness of fit test.

Test	GPD values	Statistic values	P-value
Anderson-Darling	σ: 1.14 ξ: -0.17	0.3767	0.5964
Cramér-von Mises	σ: 1.14 ξ: -0.17	0.0498	0.6260

Table 5.25: NIFTY: GPD return level ($\ddot{R}_g$) estimates.

$u = 2.25$	2-year	5-year	10-year	20-year	50-year
$\widehat{\ddot{R}_g}$	6.76	7.08	7.29	7.47	7.69
90% CI	(4.71; 8.81)	(4.57; 9.59)	(4.43; 10.14)	(4.27; 10.67)	(4.04; 11.33)
95% CI	(4.32; 9.20)	(4.09; 10.07)	(3.88; 10.69)	(3.66; 11.28)	(3.35; 12.02)
99% CI	(3.56; 9.97)	(3.15; 11.00)	(2.81; 11.76)	(2.46; 12.48)	(1.99; 13.39)

hypothesis of the Anderson-Darling and Cramér-von Mises tests. For both goodness of fit tests, we failed to reject the null hypothesis because the p-value is large (greater than 0.05) in each case. This indicates that the GPD fits the generated cluster-maxima observations well.

Table 5.25 shows the return levels estimates and their corresponding confidence intervals for the Indian NIFTY index. The table shows that a maximum loss of 6.76% is expected once every 2 years, 7.08% once every 5 years, 7.29% once every 10 years, 7.47% once every 20 years, and 7.69% once every 50 years.

5.4.8 Univariate analysis: Point process

The procedure used for the CEV model's GPD fit is also applied to the univariate analysis of the point process. For consistency and comparison between the two models, the same threshold is used. Thus, the point process model will fit the Poisson process to the same 39 cluster-maxima used for the GPD fit.

The 39 cluster-maxima are different risk levels in the Indian equity market at threshold $u = 2.25$. The point process is fitted to these cluster-maxima and the

Table 5.26: NIFTY: Univariate point process parameter estimates.

NIFTY	CI	u	μ	$\hat{\xi}$	$\hat{\sigma}$	Log. lik
	90%	2.25	6.49 (1.04) [4.78; 8.19]	-0.17 (0.16) [-0.43; 0.09]	0.42 (0.32) [-0.10; 0.94]	153.74
	95%	2.25	6.49 (1.04) [4.46; 8.52]	-0.17 (0.16) [-0.48; 0.14]	0.42 (0.32) [-0.20; 1.04]	153.74
	99%	2.25	6.49 (1.04) [3.82; 9.16]	-0.17 (0.16) [-0.57; 0.23]	0.42 (0.32) [-0.39; 1.23]	153.74

Table 5.27: Likelihood ratio test for PP's ξ estimate.

NIFTY index	w	$\hat{\xi}$	LR	CV	LR_{**}	CV
	39	-0.17	307.48	10%: 2.706 5%: 3.841 1%: 6.635	285.40	10%: 2.706 5%: 3.841 1%: 6.635

result in Table 5.26 is obtained. The results obtained for the shape parameter ξ are very similar to that of the GPD estimation with negligible differences. The negative estimate ($\hat{\xi} = -0.17$) of the shape parameter as seen in the table indicates that at this threshold $u = 2.25$, the fitted GPD has a short tail, as $\hat{\xi} < 0$, and it reflects convexity (Coles, 2001).

The values of the confidence intervals of the shape parameter ξ in Table 5.26 at the three confidence levels are from negative to positive values including zero, hence a formal hypothesis test is required to determine if $\xi = 0$ or otherwise. From the outcome in Table 5.27, at the 1%, 5% and 10% levels of significance, the values of the likelihood ratio (LR) and the modified likelihood ratio (LR_{**}) tests are greater than the critical values. Hence, the null hypothesis is rejected, implying that the shape parameter $\xi < 0$. This suggests Weibull domain of attraction for the data, hence the risk in the Indian Nifty market using the point process approach can be modelled by the Weibull class of distributions, and that an upper bound exists.

As shown in Figure 5.27, the data points in the diagnostic probability and quantile plots of the point process are close to the 45° lines. Hence, the point process fit is

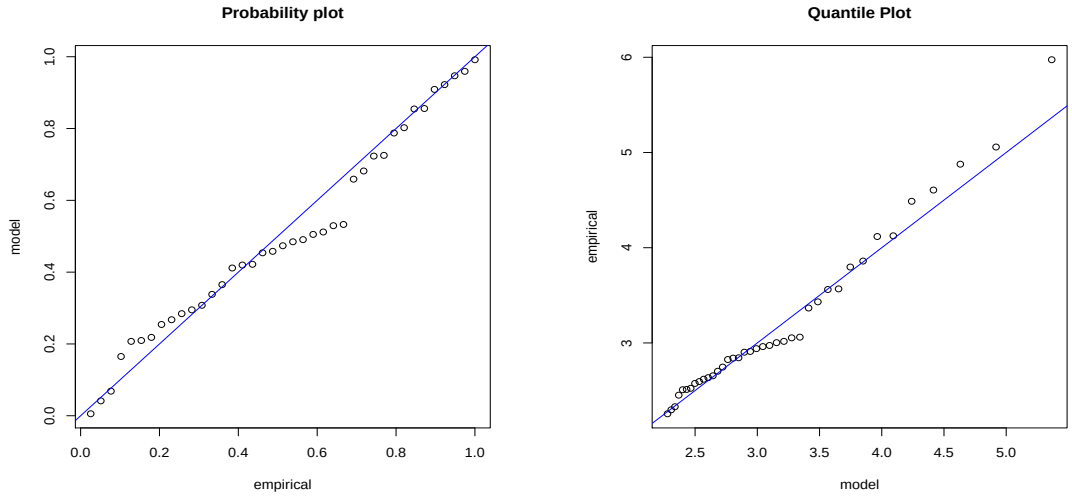


Figure 5.27: NIFTY: Point process diagnostic plots.

Table 5.28: NIFTY: Point process return level (R_g) estimates.

$u = 2.25$	2-year	5-year	10-year	20-year	50-year
$\widehat{R_g}$	6.64	7.04	7.27	7.47	7.68
90% CI	(4.75; 8.52)	(4.60; 9.49)	(4.45; 10.09)	(4.29; 10.64)	(4.06; 11.31)
95% CI	(4.39; 8.88)	(4.13; 9.96)	(3.91; 10.63)	(3.68; 11.25)	(3.36; 12.00)
99% CI	(3.69; 9.59)	(3.21; 10.87)	(2.85; 11.69)	(2.49; 12.44)	(2.01; 13.36)

satisfactory.

Table 5.28 shows the point process return level estimates and their corresponding confidence intervals (in parentheses) for the NIFTY index. The table shows that a maximum loss of 6.64% is expected once every 2 years, 7.04% once every 5 years, 7.27% once every 10 years, 7.47% once every 20 years, and 7.68% once every 50 years. The return level plot is displayed in panel (c) of Figure 7.18 in the Appendix. Comparison of the point process return level estimates with that of the GPD in Table 5.25, shows that both results are similar.

Table 5.29: PINAW values for Avg heuristic.

NIFTY index	GPD	PP	Avg combined PIs
Confidence levels	PINAW	PINAW	PINAW
90%	1.7881	1.6035	1.6918
95%	1.7921	1.6024	1.6930
99%	1.7888	1.6037	1.6921

Table 5.30: PINAW values for Env heuristic.

NIFTY index	GPD	PP	Env combined PIs
Confidence levels	PINAW	PINAW	PINAW
90%	1.7881	1.6035	1.7881
95%	1.7921	1.6024	1.7921
99%	1.7888	1.6037	1.7888

5.4.9 Combining the GPD and PP return levels' interval forecasts

The PINAW tables for the simple average (Ave) and the envelop (Env) are shown in Tables 5.29 and 5.30. For both the simple average and envelop heuristics in tables, the point process model gave the best performance when compared with the combined PIs approach and the GPD model with the narrowest PINAW values at the three confidence levels. Hence, the point process model is the best model in constructing prediction intervals for the Indian Nifty market's return levels forecast.

5.5 The Chinese market: SHCOMP

The Chinese market index is called the "Shanghai Stock Exchange Composite", abbreviated - SHCOMP, and it contains 2126 observations for the periods 5^{th} January 2010 to 6^{th} August 2018.

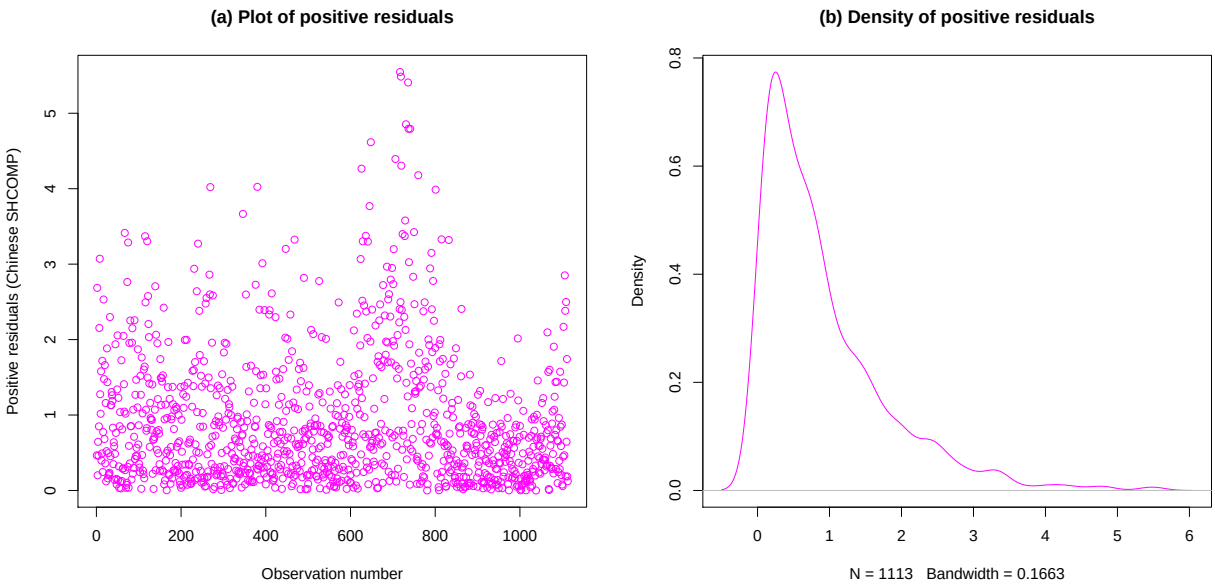


Figure 5.28: SHCOMP: Positive residuals.

5.5.1 SHCOMP: Positive residual observations

The modelling of the risk in each of the markets is restricted to the positive residuals of the entire equity observations. From the original equity data with 2126 number of observations, there are 1113 positive residual observations in this market as represented in Figure 5.28. The positive residual observations and their corresponding density are given in Figure 5.28, panels (a) and (b) respectively.

5.5.2 SHCOMP: Threshold selection

The threshold selection procedure used for the three preceeding markets will also be used for the Chinese SHCOMP market, i.e. the non-parametric Kernel GPD approach, and the shape threshold stability plot to check the result of the Kernel-GPD approach. The combined plots of the bulk model based (blue solid line) and the parameterised tail fraction (red solid line) approaches are displayed in Figure 5.29.

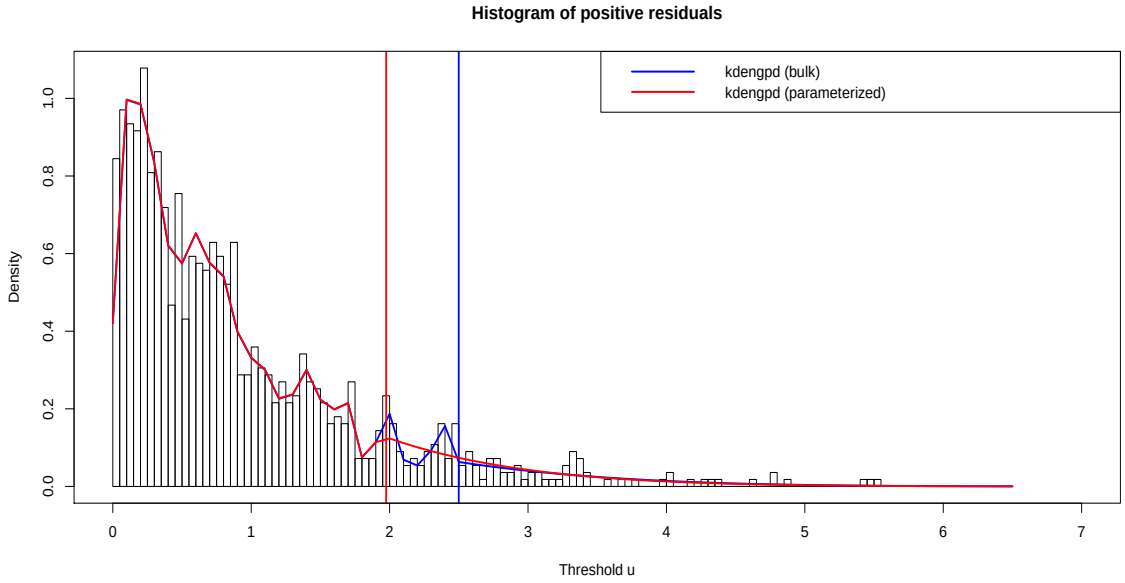


Figure 5.29: SHCOMP: Threshold selection.

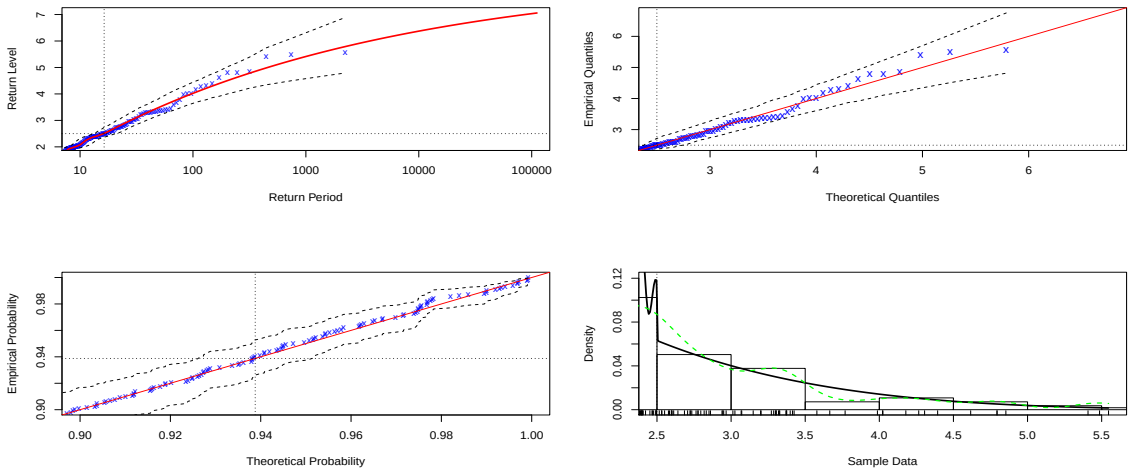


Figure 5.30: SHCOMP: Diagnostic plots of the bulk model based tail fraction.

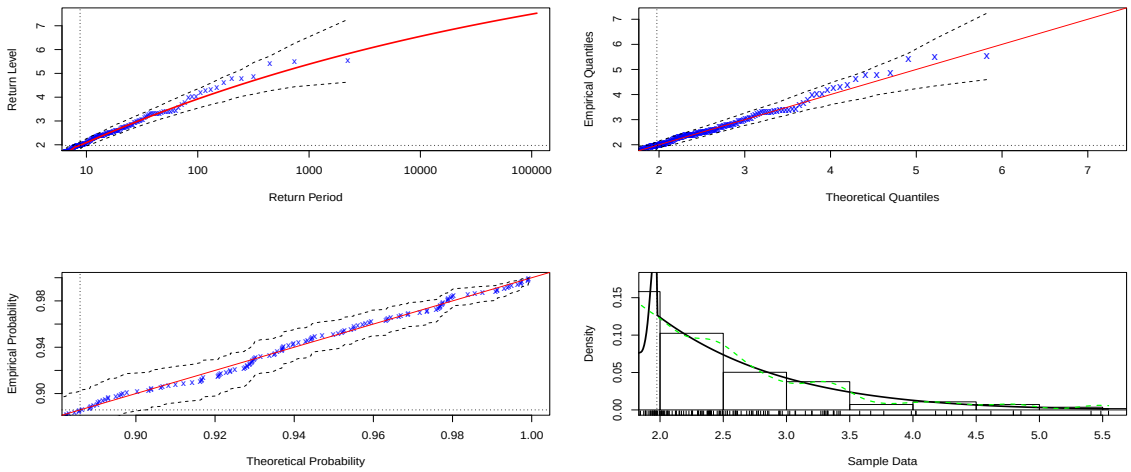


Figure 5.31: SHCOMP: Diagnostic plots of the parameterised tail fraction.

The figure displays a threshold $u = 2.4993$ for the bulk model based tail fraction and $u = 1.9751$ for the parameterised tail fraction approach.

5.5.3 Diagnostic plots of the bulk and parameterised approaches

The two sets of diagnostic plots in Figures 5.30 and 5.31 for the fitted bulk model based and parameterised tail fraction respectively for the threshold selection show reasonable levels of suitability. This suggests that the non-parametric Kernel-GPD mixture models is satisfactory.

However, there is a higher level of linearity in the alignment of the data points on the diagonal lines of the return level, quantiles and probability plots in the parameterised tail fraction than it is observed in the bulk model based. Hu and Scarrott (2018) indicated that the bulk model based tail fraction exposes estimation of the tail to the misspecification of the bulk model. This is a major shortcoming to the use of the bulk model based which the parameterised tail fraction approach corrects by using an extra parameter ϑ_u for the tail fraction. For this reason, the parameterised tail fraction approach reduces the effect of the bulk model's misspecification on the tail estimates.

5.5.4 Sensitivity analysis

Here, the estimated threshold value $u = 1.9751$ from the parameterised tail fraction approach is used as the starting point for the sensitivity analysis (in Figure 5.32) due to its superior diagnostics accuracy. However, the extremal index $\hat{\theta}$ at this threshold value is lower than 0.5, hence it is not chosen.

It is observed that extremal index $\hat{\theta}$ is greater than 0.5 from threshold value $u = 2.86$ and above. At threshold $u = 2.86$, the extremal index is 0.5349 with 43

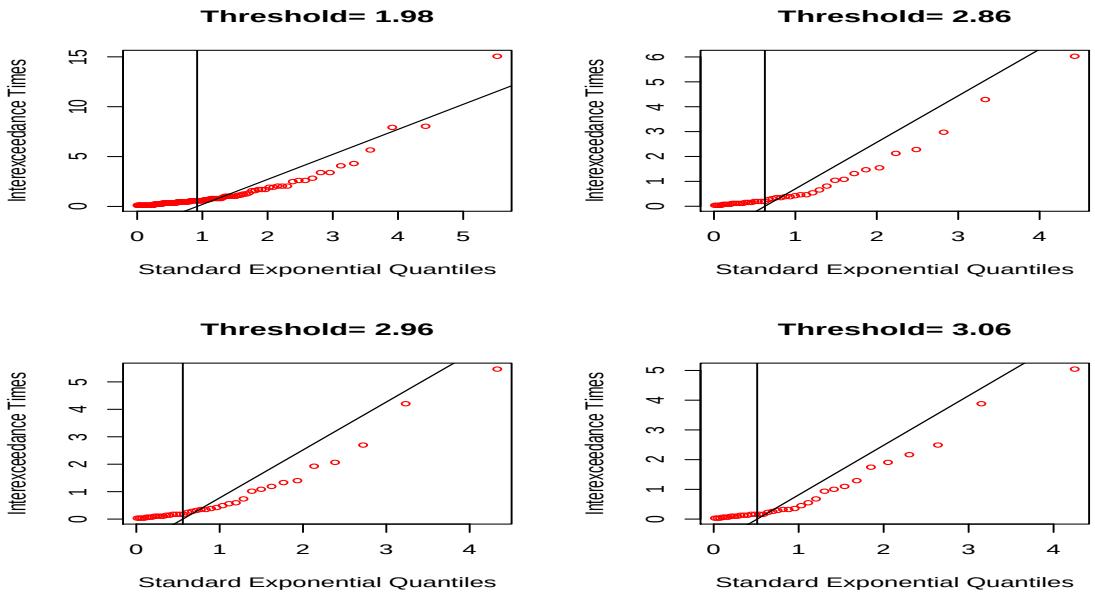


Figure 5.32: SHCOMP: Sensitivity analysis plots for threshold selection.

threshold exceedances, hence exceedances occur in groups of $\frac{1}{0.5349} = 1.8695 \approx 2$. This threshold estimate gives more information about the tail than the other remaining two thresholds (in the figure) at $u = 2.96$ and $u = 3.0$ that give less information with 39 and 36 cluster-maxima exceedances respectively. Based on this, threshold estimate $u = 2.86$ is chosen.

5.5.5 Shape threshold stability plot

The shape threshold stability plot in Figure 5.33 shows three potential thresholds that are identified as significant departures from linearity at $u = 1.3$, $u = 1.9$ and $u = 2.4$. Threshold estimate $u = 1.9$ is close to the $u = 1.9751$ obtained by the parameterised tail fraction, while $u = 2.4$ is not far from the threshold $u = 2.86$ obtained in the sensitivity analysis.

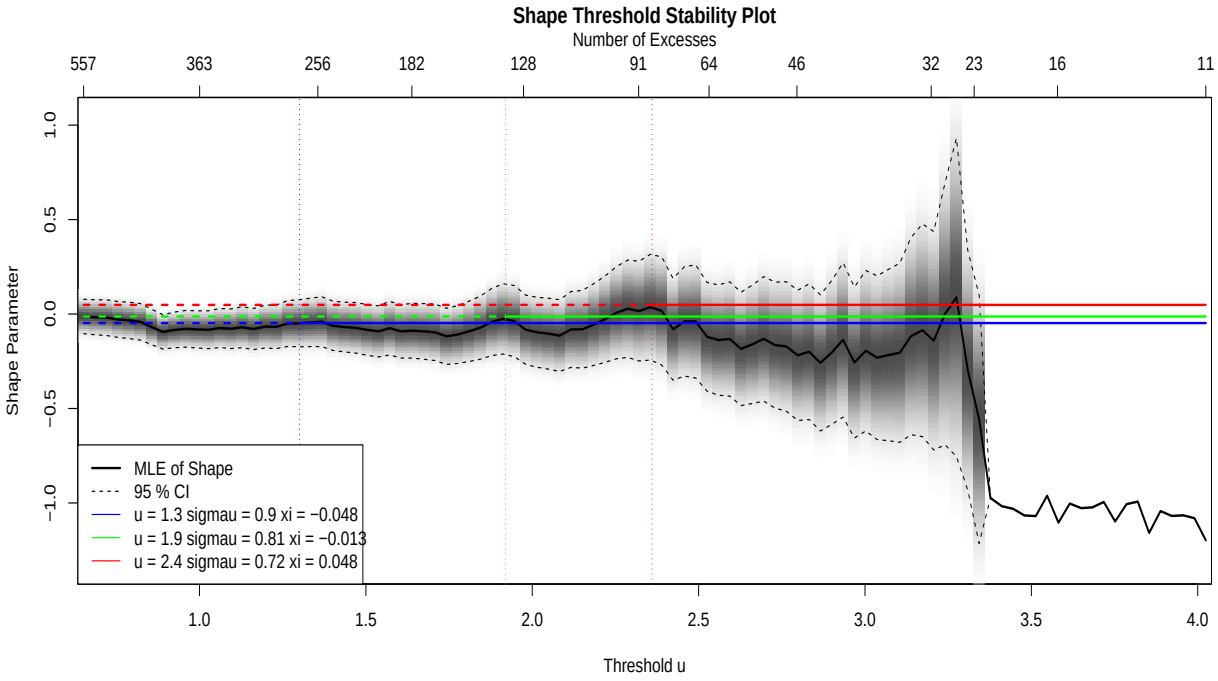


Figure 5.33: SHCOMP: Shape threshold stability plot.

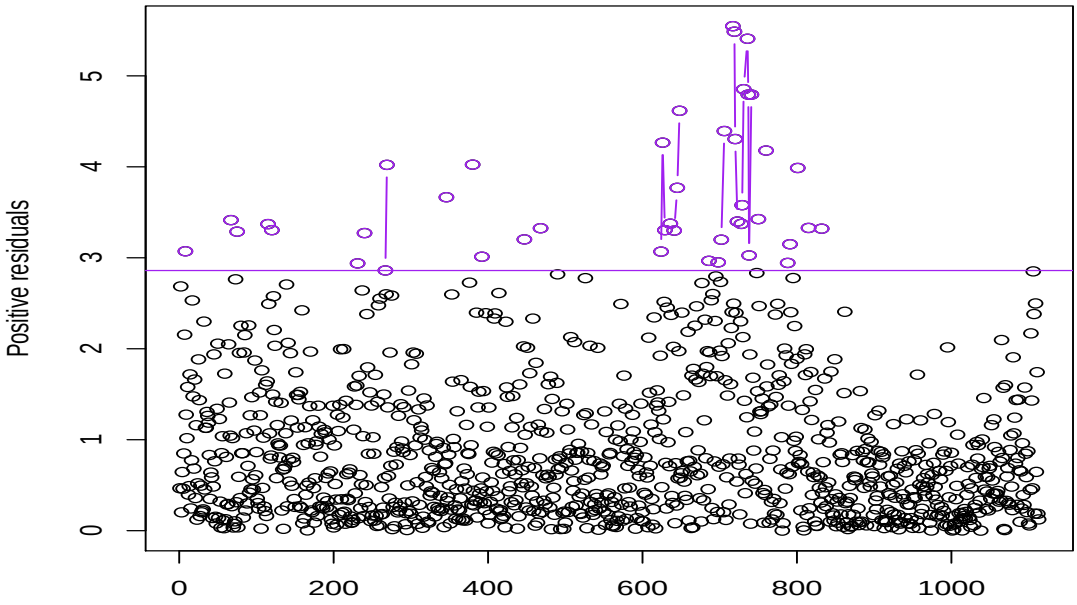


Figure 5.34: SHCOMP: Declustered exceedances at $u = 2.86$ (cluster-maxima).

Table 5.31: SHCOMP: GPD parameter estimates.

	CI	u	n	w	$\eta_u = \frac{w}{n}$	$\hat{\xi}$	$\hat{\sigma}$	Log. lik
SHCOMP	90%	2.86	1113	23	0.0207	-0.25 (0.18) [-0.49; -0.01]	0.98 (0.27) [0.62; 1.34]	-16.75
	95%	2.86	1113	23	0.0207	-0.25 (0.18) [-0.56; 0.07]	0.98 (0.27) [0.51; 1.45]	-16.75
	99%	2.86	1113	23	0.0207	-0.25 (0.18) [-0.71; 0.21]	0.98 (0.27) [0.29; 1.66]	-16.75

5.5.6 Declustering

The estimated threshold $u = 2.86$ generated 43 threshold exceedances. These exceedances were declustered and 23 cluster-maxima are obtained as shown in Figure 5.34. These 23 cluster-maxima are the maxima of the clusters of exceedances over the high threshold $u = 2.86$. They are the independent extreme tail observations on which the GPD is fitted.

5.5.7 CEV model: GPD fit to cluster-maxima

The GPD is fitted to the 23 declustered exceedances (or cluster-maxima). These cluster-maxima are collectively classified as the risk or extreme losses to be modelled in the Chinese SHCOMP market. The result of the GPD fit is shown in Table 5.31. The negative estimate $\hat{\xi} = -0.25$ of the shape parameter shows that at this threshold $u = 2.86$, the fitted GPD has a short tail, as $\hat{\xi} < 0$, and it is a reflection of convexity (Coles, 2001).

Confidence intervals of the GPD parameter estimates ($\hat{\xi}$ and $\hat{\sigma}$) can be achieved with better accuracy using the profile likelihood (Coles 2001). The approximate profile log-likelihood confidence intervals at 90%, 95% and 99% are shown in Table 5.32.

The confidence intervals of the shape parameter ξ in the Table 5.32 at the three

Table 5.32: SHCOMP: GPD parameter estimates and profile likelihood intervals.

	CI	u	n	w	$\eta_u = \frac{w}{n}$	$\hat{\xi}$	$\hat{\sigma}$	Log. lik
SHCOMP	90%	2.86	1113	23	0.0207	-0.25 (0.18) [-0.36; 0.16]	0.98 (0.27) [0.66; 1.53]	-16.75
	95%	2.86	1113	23	0.0207	-0.25 (0.18) [-0.36; 0.27]	0.98 (0.27) [0.67; 1.68]	-16.75
	99%	2.86	1113	23	0.0207	-0.25 (0.18) [-0.36; 0.53]	0.98 (0.27) [0.67; 2.01]	-16.75

Table 5.33: Likelihood ratio test for GPD's ξ estimate.

SHCOMP index	w	$\hat{\xi}$	LR	CV	LR_{**}	CV
	23	-0.25	-33.50	10%: 2.706 5%: 3.841 1%: 6.635	-29.42	10%: 2.706 5%: 3.841 1%: 6.635

confidence levels are from negative to positive values including zero, hence a formal hypothesis test is conducted to ascertain if $\xi = 0$ or otherwise. From the outcome in Table 5.33, at the 1%, 5% and 10% levels of significance, the values of the likelihood ratio (LR) and the modified likelihood ratio (LR_{**}) tests are less than the critical values. Hence, we fail to reject the null hypothesis and conclude that the shape parameter $\xi = 0$. This suggests that the risk in the Chinese SHCOMP market can be modelled by the Gumbel class of distributions.

5.5.8 Diagnostics: Model checking

The validity of the fitted GPD model is verified by the diagnostic plots in Figure 5.35. None of these four plots gives any real cause for concern about the quality of the fitted GPD model.

The result of the Anderson-Darling and Cramér-von Mises goodness of fit tests for knowing how well the GPD model fits the cluster-maxima exceedances in the SHCOMP market are shown in Table 5.34. The tested hypothesis are:

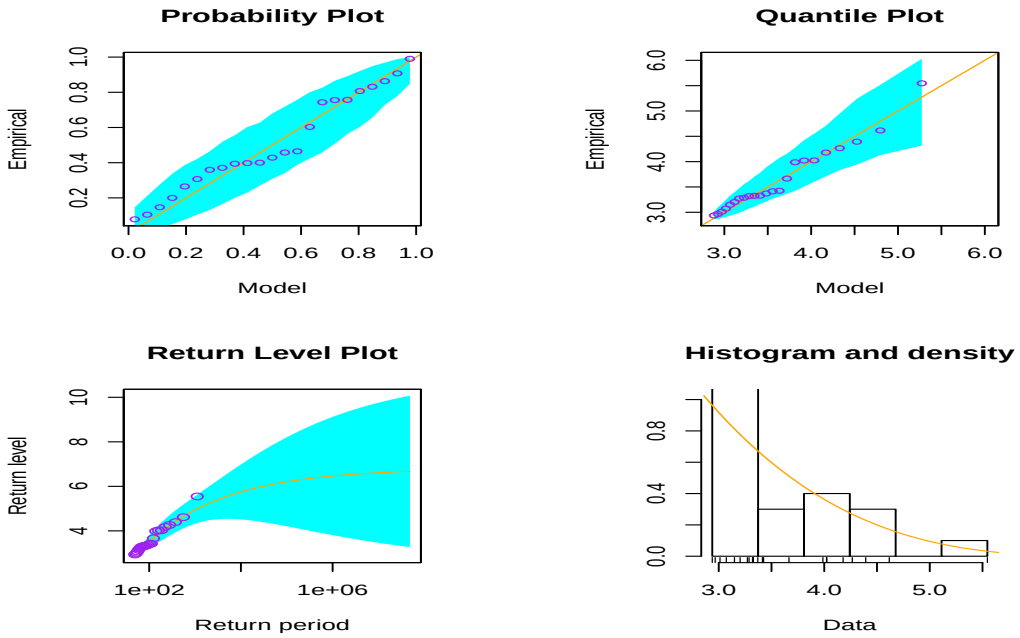


Figure 5.35: SHCOMP: GPD diagnostic plots.

Table 5.34: SHCOMP: Goodness of fit test.

Test	GPD values	Statistic values	P-value
Anderson-Darling	σ: 0.98 ξ: -0.25	0.3958	0.5767
Cramér-von Mises	σ: 0.98 ξ: -0.25	0.0537	0.5950

H_0 : GPD fits the cluster-maxima well

H_1 : GPD does not fit the cluster-maxima well

The table displays the outcomes of the test statistic and p-values of the tested hypothesis of the Anderson-Darling and Cramér-von Mises tests. For both goodness of fit tests, we failed to reject the null hypothesis because the p-value is large (greater than 0.05) in each case. This indicates that the GPD fits the generated cluster-maxima observations well.

Table 5.35: SHCOMP: GPD return level ($\ddot{R}_g$) estimates.

$u = 2.25$	2-year	5-year	10-year	20-year	50-year
$\widehat{\ddot{R}_g}$	6.03	6.19	6.29	6.37	6.46
90% CI	(4.75; 7.31)	(4.68; 7.70)	(4.61; 7.96)	(4.54; 8.20)	(4.43; 8.48)
95% CI	(4.37; 7.69)	(4.23; 8.15)	(4.11; 8.46)	(3.99; 8.75)	(3.83; 9.08)
99% CI	(3.61; 8.46)	(3.32; 9.06)	(3.10; 9.47)	(2.89; 9.85)	(2.62; 10.29)

Table 5.36: SHCOMP: Univariate point process parameter estimates.

SHCOMP	CI	u	μ	$\hat{\xi}$	$\hat{\sigma}$	Log. lik
	90%	2.86	5.89 (0.73) [4.93; 6.85]	-0.25 (0.18) [-0.49; -0.01]	0.23 (0.20) [-0.03; 0.49]	95.95
	95%	2.86	5.89 (0.73) [4.64; 7.13]	-0.25 (0.18) [-0.56; 0.07]	0.23 (0.20) [-0.11; 0.56]	95.95
	99%	2.86	5.89 (0.73) [4.07; 7.71]	-0.25 (0.18) [-0.71; 0.21]	0.23 (0.20) [-0.27; 0.72]	95.95

The return levels estimates and their corresponding confidence intervals for the Chinese SHCOMP index are shown in Table 5.35. The table shows that a maximum loss of 6.03% is expected once every 2 years, 6.19% once every 5 years, 6.29% once every 10 years, 6.37% once every 20 years, and 6.46% once every 50 years.

5.5.9 Univariate analysis: Point process

For consistency and comparison, the univariate analysis of the point process is tailored to follow the same procedure used for the CEV model's GPD fit, at the same threshold. Based on this, the point process model is used to fit the Poisson process to the same 23 cluster-maxima used for the GPD fit.

The 23 cluster-maxima are different magnitudes or levels of risk in the Chinese equity market at threshold $u = 2.86$. The results in Table 5.36 are obtained after fitting the point process to the cluster-maxima. The standard errors of the estimates are enclosed in parentheses while the confidence intervals are in the brackets. The

Table 5.37: Likelihood ratio test for PP's ξ estimate.

SHCOMP index	w	$\hat{\xi}$	LR	CV	LR_{**}	CV
	23	-0.25	191.90	10%: 2.706 5%: 3.841 1%: 6.635	168.54	10%: 2.706 5%: 3.841 1%: 6.635

shape parameter's negative estimate $\hat{\xi} = -0.25$ in the table indicates that at this threshold $u = 2.86$, the fitted GPD has a short tail, as $\hat{\xi} < 0$, and it reflects convexity (Coles, 2001). However, there is no difference in the results of the shape parameter ξ obtained from both the point process and GPD models' fits.

Table 5.36 shows that the confidence interval of the shape parameter ξ at 90% is entirely in the negative domain. No formal hypothesis test is required at this level, indicating that the risk can be described by a bounded (Weibull) distribution. At confidence levels 95% and 99% however, the values of the confidence intervals cover zero, i.e. from negative to positive, hence a formal hypothesis test is required at this instance. From the outcome in Table 5.37, at the 1%, 5% and 10% levels of significance, the values of the likelihood ratio (LR) and the modified likelihood ratio (LR_{**}) tests are greater than the critical values. Hence, the null hypothesis is rejected, implying that the shape parameter $\xi < 0$. This also suggests a Weibull domain of attraction for the data, hence the risk in the Chinese SHCOMP market using the point process approach can be described by the Weibull class of distributions.

As displayed in Figure 5.36, the data points in the diagnostic probability plot and quantile plot of the point process are close to the 45° lines, with the exception of the single outlier at the extreme of the quantile plot. Thus, the point process fit is satisfactory. The outlier occurred on the 14th July 2015 and findings from various studies revealed that it was orchestrated by a turbulence in the Chinese equity market which began on the 12 June 2015 as a result of the bubble that popped in the market (Riley and Yan, 2015). In July 2015, the Shanghai stock market had nosedived by

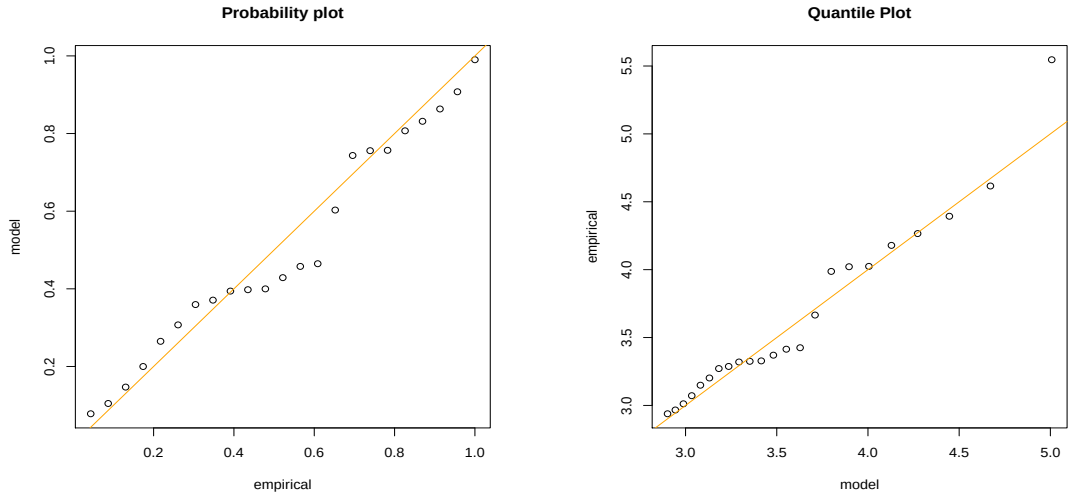


Figure 5.36: SHCOMP: Point process diagnostic plots.

Table 5.38: SHCOMP: Point process return level (R_g) estimates.

$u = 2.86$	2-year	5-year	10-year	20-year	50-year
$\widehat{R_g}$	5.97	6.17	6.28	6.36	6.45
90% CI	(4.78; 7.15)	(4.70; 7.64)	(4.63; 7.93)	(4.55; 8.17)	(4.45; 8.45)
95% CI	(4.43; 7.51)	(4.26; 8.08)	(4.14; 8.42)	(4.01; 8.71)	(3.86; 9.05)
99% CI	(3.72; 8.21)	(3.38; 8.96)	(3.15; 9.41)	(2.93; 9.80)	(2.66; 10.25)

30%, which made about 1,400 companies file for a trading halt so as to avoid further losses (Duggan, 2015; Hunt, 2015; Gough, 2015).

Table 5.38 shows the point process return level estimates and their corresponding confidence intervals in parentheses for the SHCOMP index. The table shows that a maximum loss of 5.97% is expected once every 2 years, 6.17% once every 5 years, 6.28% once every 10 years, 6.36% once every 20 years, and 6.45% once every 50 years. The return level plot is displayed in panel (d) of Figure 7.18 in the Appendix. When compared with the GPD's return levels results in Table 5.15, both models display similar outcomes.

Table 5.39: PINAW values for Avg heuristic.

SHCOMP index	GPD	PP	Avg combined PIs
Confidence levels	PINAW	PINAW	PINAW
90%	2.2336	1.9914	2.1071
95%	2.2383	1.9972	2.1124
99%	2.2404	1.9865	2.1074

Table 5.40: PINAW values for Env heuristic.

SHCOMP index	GPD	PP	Env combined PIs
Confidence levels	PINAW	PINAW	PINAW
90%	2.2336	1.9914	2.2336
95%	2.2383	1.9972	2.2383
99%	2.2404	1.9865	2.2404

5.5.10 Combining the GPD and PP return levels' interval forecasts

The PINAW tables for the simple average (Ave) and the envelop (Env) are shown in Tables 5.39 and 5.40. For both the simple average and envelop heuristics in tables, the point process outperforms the combined PIs approach and the GPD model with the narrowest PINAW values at the three confidence levels. Hence, the point process model is the most efficient probabilistic forecasting approach for the Chinese SHCOMP market's return levels forecast.

5.6 The South African market: JALSH

The South African market index is called the "JSE Africa All Share index", abbreviated - JALSH, and it contains 2126 observations for the periods 5^{th} January 2010 to 6^{th} August 2018.

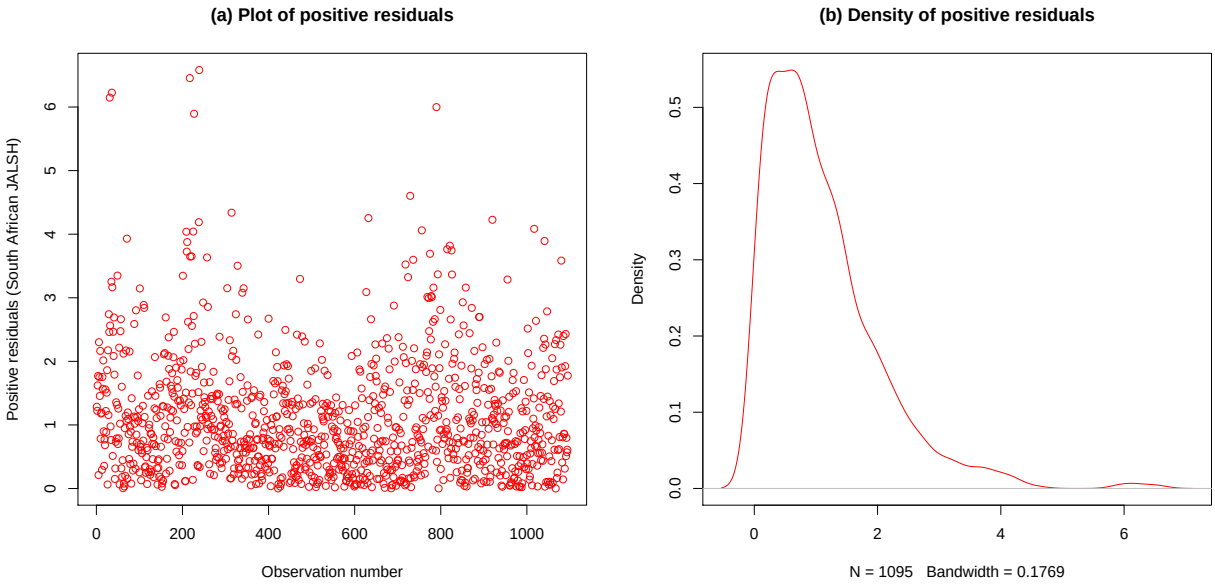


Figure 5.37: JALSH: Positive residuals.

5.6.1 JALSH: Positive residual observations

From the original equity data with 2126 number of realizations, there are 1095 positive residual realizations in this market as shown in Figure 5.37. The positive residual observations and their corresponding density are given in Figure 5.37, panels (a) and (b) respectively.

5.6.2 JALSH: Threshold selection

For the South African JALSH market, the same threshold selection procedure that was used for the other four markets is applied, i.e. the non-parametric Kernel-GPD approach, and the shape threshold stability plot to verify the result of the Kernel-GPD approach.

The combined plots of the bulk model based (blue solid line) and the parameterised tail fraction (red solid line) approaches are shown in Figure 5.38. The figure displays

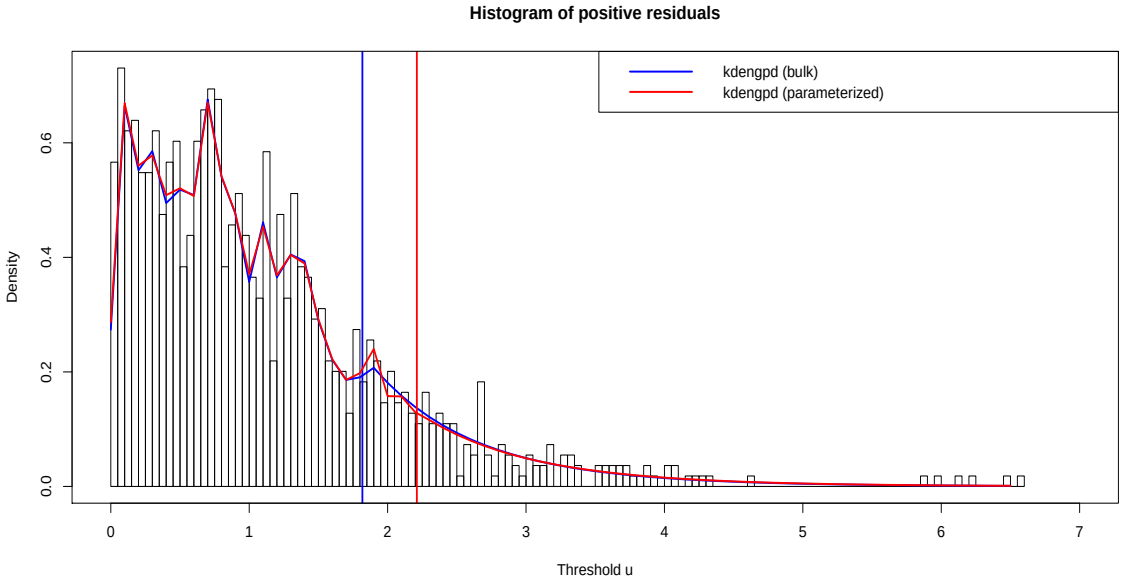


Figure 5.38: JALSH: Threshold selection.

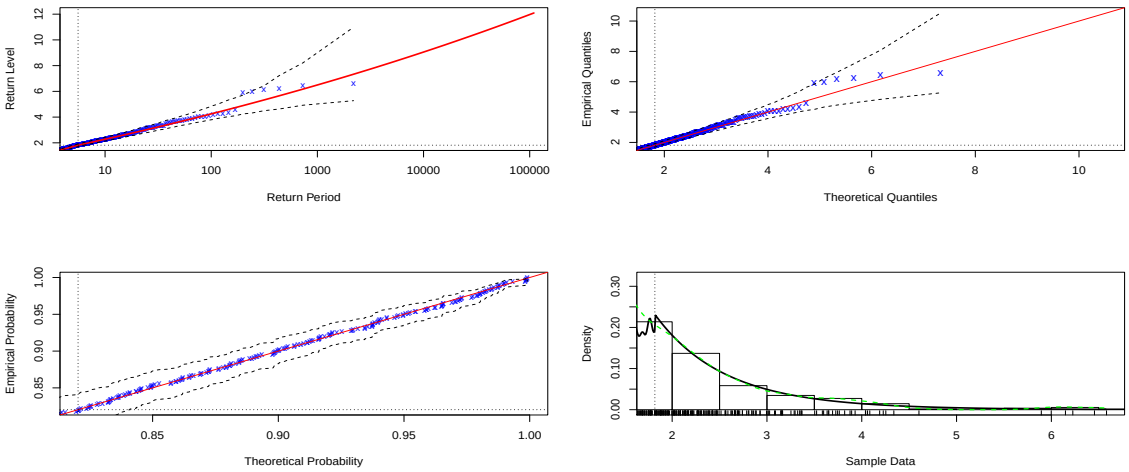


Figure 5.39: JALSH: Diagnostic plots of the bulk model based tail fraction.

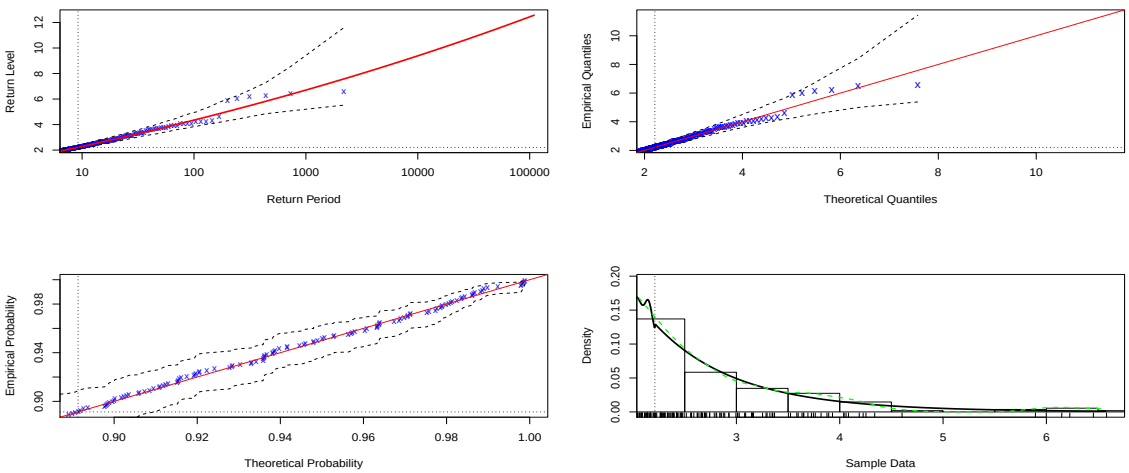


Figure 5.40: JALSH: Diagnostic plot of the parameterised tail fraction.

a threshold $u = 2.2112$ for the parameterised tail fraction approach, and $u = 1.8178$ for the bulk model based tail fraction.

5.6.3 Diagnostic plots of the bulk and parameterised approaches

From each set of the quartet diagnostic plots in Figures 5.39 and 5.40, the alignments of the data points on the daigonal lines of the probability, quantile and return level plots are nearly the same for the bulk model based and the parameterised tail fraction. However, the density estimate seems more consistent with the histogram of the data for the diagnostic plot of the parameterised tail fraction than it is for the bulk model based.

5.6.4 Sensitivity analysis

The threshold value $u = 2.21$ obtained from the parameterised tail fraction approach is used as an initial point for the sensitivity analysis in Figure 5.41. From the plots, threshold value $u = 2.60$ is chosen because it displays the best piecewise linearity of points with extremal index $\hat{\theta} = 0.5117$ than the rest of the plots. This means that exceedances occur in groups of $\frac{1}{0.5117} = 1.9543 \approx 2$. Furthermore, the extremal indices at threshold values $u = 2.30$ and $u = 2.45$ are both less than 0.5.

5.6.5 Shape threshold stability plot

From the shape threshold stability plot in Figure 5.42, thresholds $u = 2.5$, $u = 2.7$ and $u = 3.1$ are identified for significant departures from linearity. The threshold choices $u = 2.5$ and $u = 2.7$ are close to and are approximately consistent with the threshold estimate of $u = 2.60$ from the Kernel density mixture models.

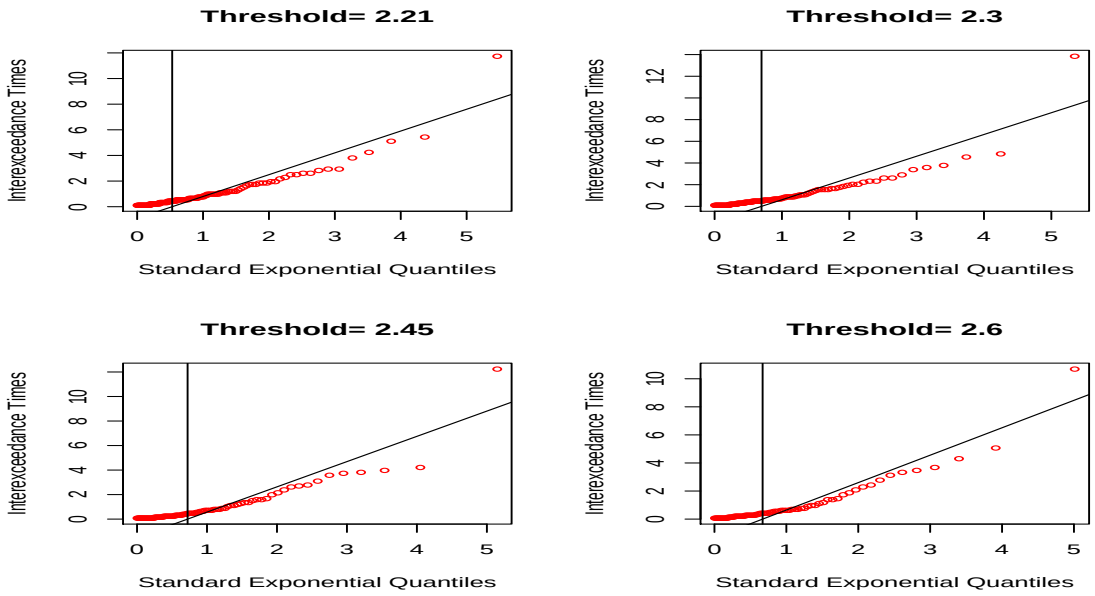


Figure 5.41: JALSH: Sensitivity analysis plots for threshold selection.

5.6.6 Declustering

The estimated threshold $u = 2.60$ generated 76 threshold exceedances. After declustering at this threshold, 36 cluster-maxima are obtained as shown in Figure 5.43. These 36 cluster-maxima are the maxima of the clusters of exceedances over the high threshold $u = 2.60$. They are the independent extreme tail observations on which the GPD is fitted.

5.6.7 CEV model: GPD fit to cluster-maxima

The GPD is now fitted to the 36 cluster-maxima, which collectively represent the risk or extreme losses to be modelled in the South African JALSH market. The result in Table 5.41 is obtained from fitting the GPD to the 36 risk observations.

The table shows a negative shape parameter estimate of $\hat{\xi} = -0.05$, which indicates a short tailed distribution that reflects convexity (Coles, 2001).

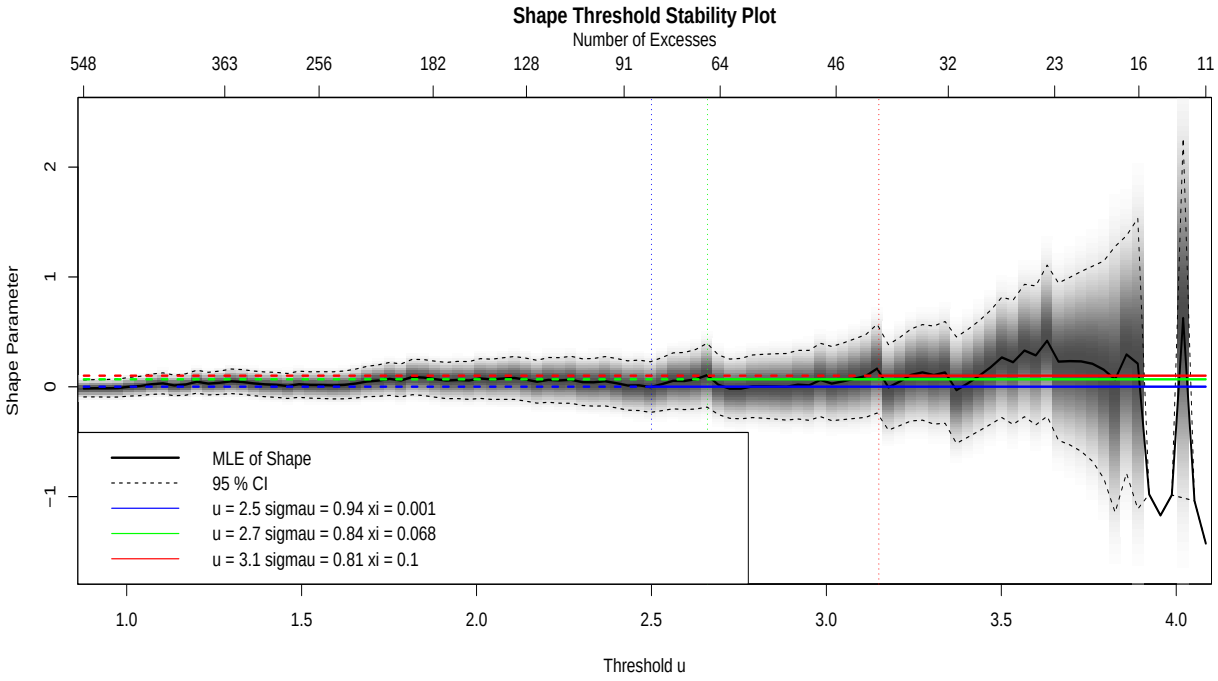


Figure 5.42: JALSH: Shape threshold stability plot.

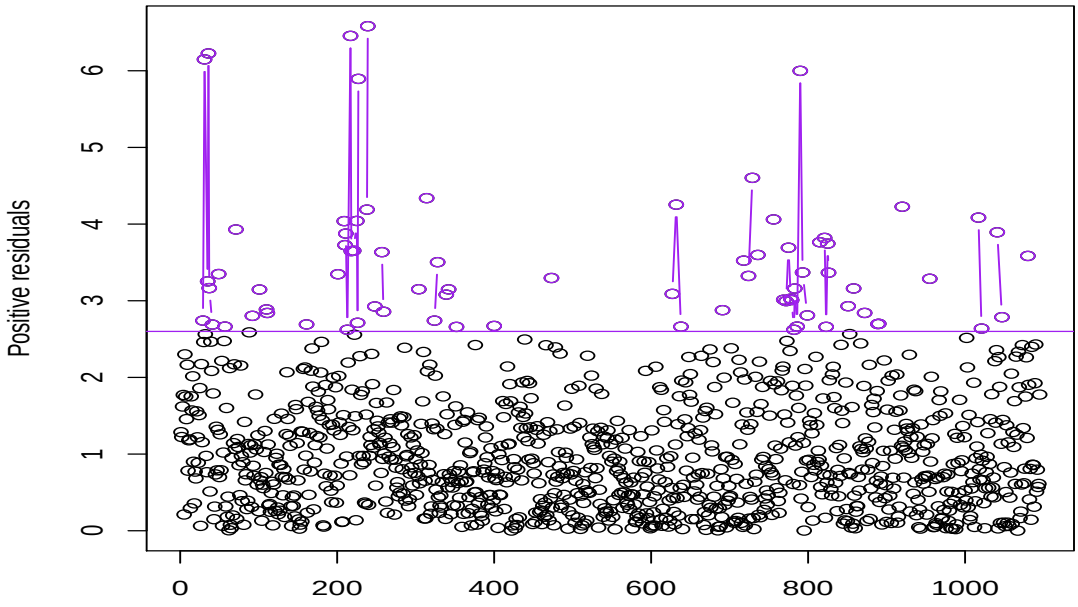


Figure 5.43: JALSH: Declustered exceedances at $u = 2.60$ (cluster-maxima).

Table 5.41: JALSH: GPD parameter estimates.

	CI	u	n	w	$\eta_u = \frac{w}{n}$	$\hat{\xi}$	$\hat{\sigma}$	Log. lik
	90%	2.60	1095	36	0.0329	-0.05 (0.20)	1.16 (0.26)	-39.51
						[-0.37; 0.27]	[0.74; 1.58]	
JALSH	95%	2.60	1095	36	0.0329	-0.05 (0.20)	1.16 (0.26)	-39.51
						[-0.43; 0.33]	[0.66; 1.66]	
	99%	2.60	1095	36	0.0329	-0.05 (0.20)	1.16 (0.26)	-39.51
						[-0.55; 0.45]	[0.50; 1.82]	

Table 5.42: JALSH: GPD parameter estimates and profile likelihood intervals.

	CI	u	n	w	$\eta_u = \frac{w}{n}$	$\hat{\xi}$	$\hat{\sigma}$	Log. lik
	90%	2.60	1095	36	0.0329	-0.05 (0.20)	1.16 (0.26)	-39.51
						[-0.29; 0.35]	[0.75; 1.77]	
JALSH	95%	2.60	1095	36	0.0329	-0.05 (0.20)	1.16 (0.26)	-39.51
						[-0.29; 0.44]	[0.68; 1.93]	
	99%	2.60	1095	36	0.0329	-0.05 (0.20)	1.16 (0.26)	-39.51
						[-0.29; 0.67]	[0.57; 2.27]	

Table 5.43: Likelihood ratio test for GPD's ξ estimate.

JALSH index	w	$\hat{\xi}$	LR	CV	LR_{**}	CV
	36	-0.05	-79.02	10%: 2.706 5%: 3.841 1%: 6.635	-72.87	10%: 2.706 5%: 3.841 1%: 6.635

The approximate profile log-likelihood confidence intervals at 90%, 95% and 99% for the GPD's parameter estimates $\hat{\xi}$ and $\hat{\sigma}$ are shown in Table 5.42.

The confidence intervals of the shape parameter ξ in Table 5.42 at the three confidence levels are from negative to positive values including zero, hence a formal hypothesis test is taken to ascertain if $\xi = 0$ or otherwise. From the outcome in Table 5.43, at the 1%, 5% and 10% levels of significance, the values of the likelihood ratio (LR) and the modified likelihood ratio (LR_{**}) tests are less than the critical values. Hence, we fail to reject the null hypothesis and conclude that the shape parameter $\xi = 0$. This suggests that the risk in the South African JALSH market can be modelled by the Gumbel class of distributions.

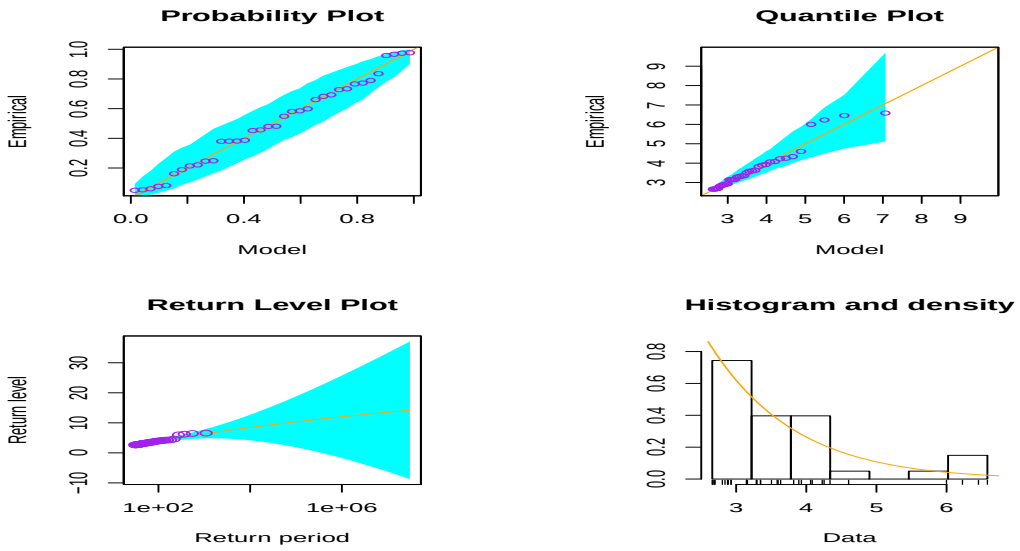


Figure 5.44: JALSH: GPD diagnostic plots.

5.6.8 Diagnostics: Model checking

The four diagnostic plots show no systematic departure of the data points from the CEV model's GPD fit at this choice of threshold $u = 2.60$ (see Figure 5.44). Hence, the fit is reasonable.

The goodness of fit tests of Anderson-Darling and Cramér-von Mises are further used to determine how well the GPD model fits the cluster-maxima exceedances in the JALSH market. The tested hypothesis are:

$$H_0 : \text{GPD fits the cluster-maxima well}$$
$$H_1 : \text{GPD does not fit the cluster-maxima well}$$

Table 5.44 displays the outcomes of the test statistic and p-values of the tested hypothesis of the Anderson-Darling and Cramér-von Mises tests. For both goodness of fit tests, we failed to reject the null hypothesis since the p-value is large (greater than

Table 5.44: JALSH: Goodness of fit test.

Test	GPD values	Statistic values	P-value
Anderson-Darling	σ: 1.16 ξ: -0.05	0.3565	0.6110
Cramér-von Mises	σ: 1.16 ξ: -0.05	0.0458	0.6512

Table 5.45: JALSH: GPD return level ($\ddot{R}_g$) estimates.

$u = 2.60$	2-year	5-year	10-year	20-year	50-year
$\widehat{\ddot{R}_g}$	9.12	9.86	10.41	10.93	11.60
90% CI	(4.43; 13.80)	(3.75; 15.98)	(3.11; 17.70)	(2.38; 19.49)	(1.27; 21.93)
95% CI	(3.53; 14.70)	(2.58; 17.15)	(1.72; 19.10)	(0.74; 21.12)	(-0.71; 23.91)
99% CI	(1.78; 16.45)	(0.29; 19.43)	(-1.01; 21.83)	(-2.46; 24.32)	(-4.57; 27.77)

0.05) in each case. This indicates that the GPD fits the generated cluster-maxima observations well.

Table 5.45 shows the return levels estimates and their corresponding confidence intervals for the South African JALSH market. The table shows that a maximum loss of 9.12% is expected once every 2 years, 9.86% once every 5 years, 10.41% once every 10 years, 10.93% once every 20 years, and 11.60% once every 50 years.

5.6.9 Univariate analysis: Point process

The application of the univariate version of the point process for the risk modelling follows the same procedure used for the conditional extreme value (CEV) model's GPD fit. The same threshold is used for consistency and comparison between the two models. Based on this, the limiting distribution of the point process is fitted to the same 36 cluster-maxima used for the GPD fit.

The 36 cluster-maxima are different risk levels or magnitudes in the South African equity market at threshold $u = 2.60$. The point process is fitted to these cluster-maxima and the result in Table 5.46 is obtained. The results obtained for the shape

Table 5.46: JALSH: Univariate point process parameter estimates.

JALSH	CI	u	μ	$\hat{\xi}$	$\hat{\sigma}$	Log. lik
	90%	2.60	8.53 (2.23) [4.86; 12.19]	-0.05 (0.19) [-0.37; 0.27]	0.86 (0.82) [-0.48; 2.21]	136.88
	95%	2.60	8.53 (2.23) [4.16; 12.89]	-0.05 (0.19) [-0.43; 0.33]	0.86 (0.82) [-0.74; 2.47]	136.88
	99%	2.60	8.53 (2.23) [2.79; 14.26]	-0.05 (0.19) [-0.54; 0.45]	0.86 (0.82) [-1.24; 2.97]	136.88

Table 5.47: Likelihood ratio test for PP's ξ estimate.

JALSH index	w	$\hat{\xi}$	LR	CV	LR_{**}	CV
	36	-0.05	273.76	10%: 2.706 5%: 3.841 1%: 6.635	252.47	10%: 2.706 5%: 3.841 1%: 6.635

parameter ξ are roughly the same with the ones from the GPD fit of the CEV model.

The values of the shape parameter's confidence intervals in Table 5.46 at the three confidence levels are from negative to positive values including zero, hence a formal hypothesis test is required to determine if $\xi = 0$ or otherwise. From the outcome in Table 5.47, at the 1%, 5% and 10% levels of significance, the values of the likelihood ratio (LR) and the modified likelihood ratio (LR_{**}) tests are greater than the critical values. Hence, the null hypothesis is rejected, signifying that the shape parameter $\xi < 0$. This suggests Weibull domain of attraction for the data, hence the risk in the South African JALSH market using the point process approach can be modelled by the Weibull class of distributions.

As shown in Figure 5.45, the data points in the diagnostic probability plot and quantile plot of the point process are close to the 45° lines. Hence, the fit appears to be good.

Table 5.48 shows the point process return level estimates and their corresponding confidence intervals in parentheses for the JALSH index. The table shows that a maximum loss of 8.84% is expected once every 2 years, 9.78% once every 5 years,

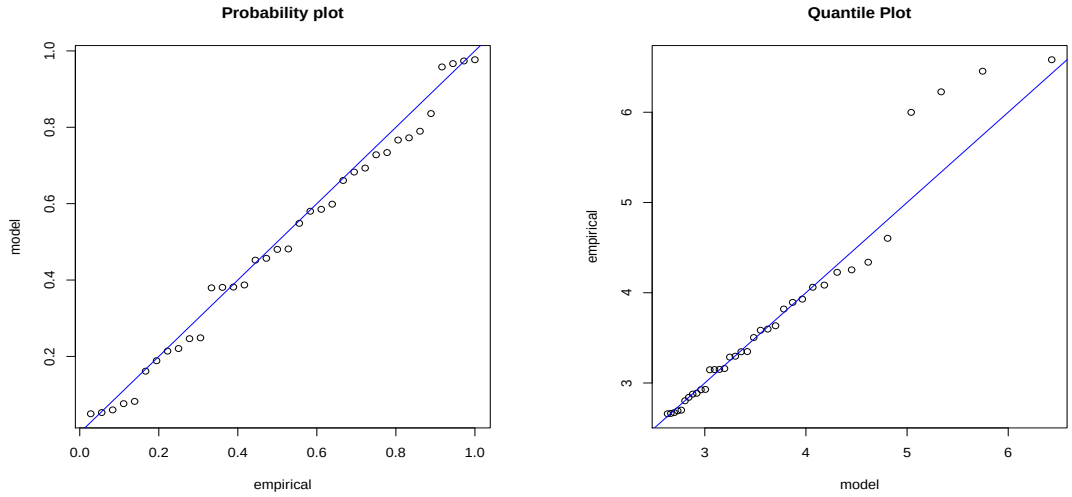


Figure 5.45: JALSH: Point process diagnostic plots.

10.37% once every 10 years, 10.92% once every 20 years, and 11.60% once every 50 years. The return level plot is displayed in panel (e) of Figure 7.18 in the Appendix. The return level estimates from the point process is similar to that of the GPD in Table 5.45.

5.6.10 Combining the GPD and PP return levels' interval forecasts

The PINAW tables for the simple average (Ave) and the envelop (Env) are shown in Tables 5.49 and 5.50. For the simple average in Table 5.49, unlike the previous four markets, the GPD model outperforms the combined PIs approach and the point process with the lowest PINAW values at the three confidence levels. For the envelop heuristic in Table 5.50 however, both the GPD and the combined PIs approach produced the same lower PINAW values when compared with the point process, hence they both performed better than the latter for the South African JALSH return levels

Table 5.48: JALSH: Point process return level (R_g) estimates.

$u = 2.25$	2-year	5-year	10-year	20-year	50-year
$\widehat{R_g}$	8.84	9.78	10.37	10.92	11.60
90% CI	(4.75; 8.52)	(4.60; 9.49)	(4.45; 10.09)	(4.29; 10.64)	(4.06; 11.31)
95% CI	(4.39; 8.88)	(4.13; 9.96)	(3.91; 10.63)	(3.68; 11.25)	(3.36; 12.00)
99% CI	(3.69; 9.59)	(3.21; 10.87)	(2.85; 11.69)	(2.49; 12.44)	(2.01; 13.36)

Table 5.49: PINAW values for Avg heuristic.

JALSH index	GPD	PP	Avg combined PIs
Confidence levels	PINAW	PINAW	PINAW
90%	1.3102	1.6035	1.3793
95%	1.3103	1.6024	1.3792
99%	1.3104	1.6037	1.3795

Table 5.50: PINAW values for Env heuristic.

JALSH index	GPD	PP	Env combined PIs
Confidence levels	PINAW	PINAW	PINAW
90%	1.3102	1.6035	1.3102
95%	1.3103	1.6024	1.3103
99%	1.3104	1.6037	1.3104

Table 5.51: GPD bootstrap tests for the BRICS markets.

	IBO		IMO		NIF		SHC		JAL	
	$\hat{\phi}$	$\hat{\xi}$	$\hat{\phi}$	$\hat{\xi}$	$\hat{\phi}$	$\hat{\xi}$	$\hat{\phi}$	$\hat{\xi}$	$\hat{\phi}$	$\hat{\xi}$
Original	-0.068	-0.051	0.226	0.144	0.127	-0.169	-0.024	-0.248	0.147	-0.050
Bootstrap mean	-0.042	-0.093	0.257	0.080	0.181	-0.260	0.089	-0.420	0.197	-0.139
Bias	0.026	-0.042	0.031	-0.064	0.054	-0.091	0.114	-0.172	0.049	-0.089
Standard deviation	0.166	0.125	0.237	0.201	0.241	0.191	0.331	0.298	0.262	0.212
Bootstrap median	-0.047	-0.086	0.263	0.092	0.186	-0.248	0.084	-0.380	0.204	-0.127

Note: IBO is Brazil-IBOV, IMO is Russia-IMOEX, NIF is India-NIFTY, SHC is China-SHCOMP, and JAL is South Africa-JALSH.

PIs construction.

5.7 Estimation of uncertainty using parametric bootstrap approach

Uncertainty and possible bias in the estimated GPD parameters can be estimated using a parametric bootstrap scheme. This approach can capture the asymmetry of the log-likelihood surface around the (GPD's) maximum likelihood estimates (Harry and Heffernan, 2013). The bootstrap method is a great tool particularly when small samples are used to predict the behaviour of a system or process. Information about the degree of uncertainty in a model's parameters estimation is vital for decision making in financial data modelling. Table 5.51 displays a summary of the bootstrap uncertainty estimation in the five BRICS markets.

The uncertainty in the GPD's parameter estimates as displayed in the table shows that the estimated bias and standard deviations are relatively small in the Brazilian, Russian, Indian, Chinese and South African markets. The bootstrap plots in Figures 5.46 to 5.50 show a visual display of the estimated uncertainty in each of the markets. The shape parameter ξ is denoted by "xi", while the scale parameter σ (reparameterised as $\phi = \log\sigma$) is denoted by "phi". The plots display nearly symmetric shapes for the five BRICS markets.

5.8 Risk hierarchy: comparing the markets' risks

The magnitudes or levels of risks in the BRICS equity markets are compared using the boxplot in Figure 5.51. The plot displays the risk hierarchy from the highest to the lowest, with their associated densities in Figure 5.52. Based on the relative spread (or variability) of each box in Figure 5.51, it can be observed that the Russian IMOEX

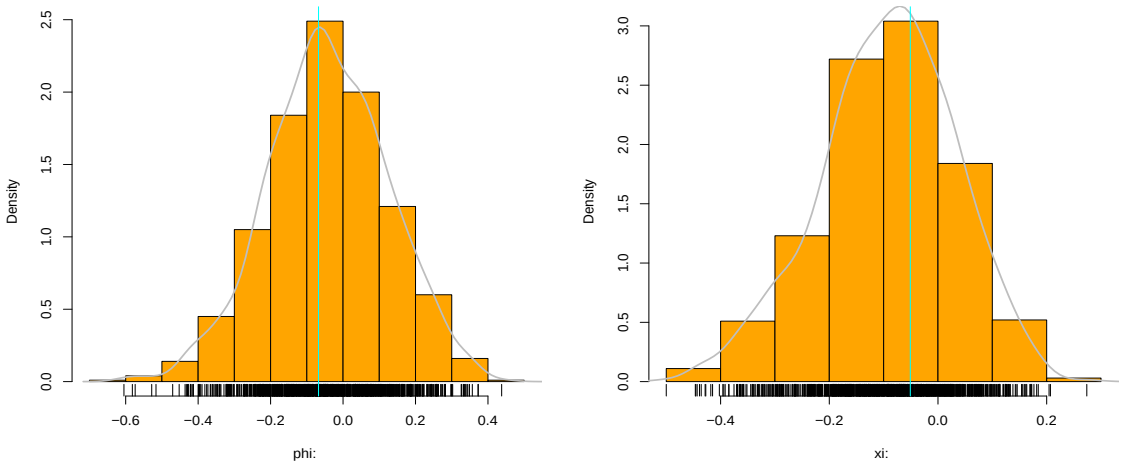


Figure 5.46: Brazilian-IBOV: Bootstrap plots.

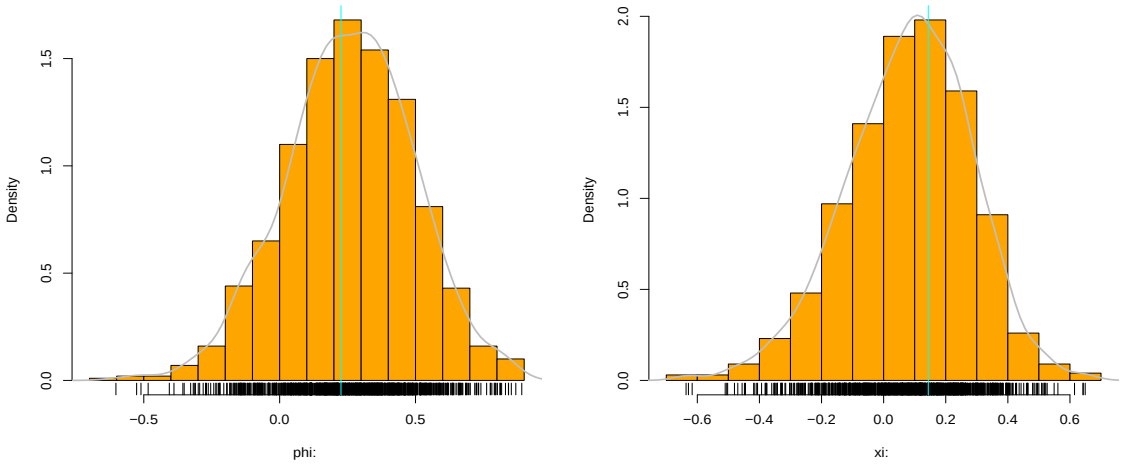


Figure 5.47: Russian-IMOEX: Bootstrap plots.

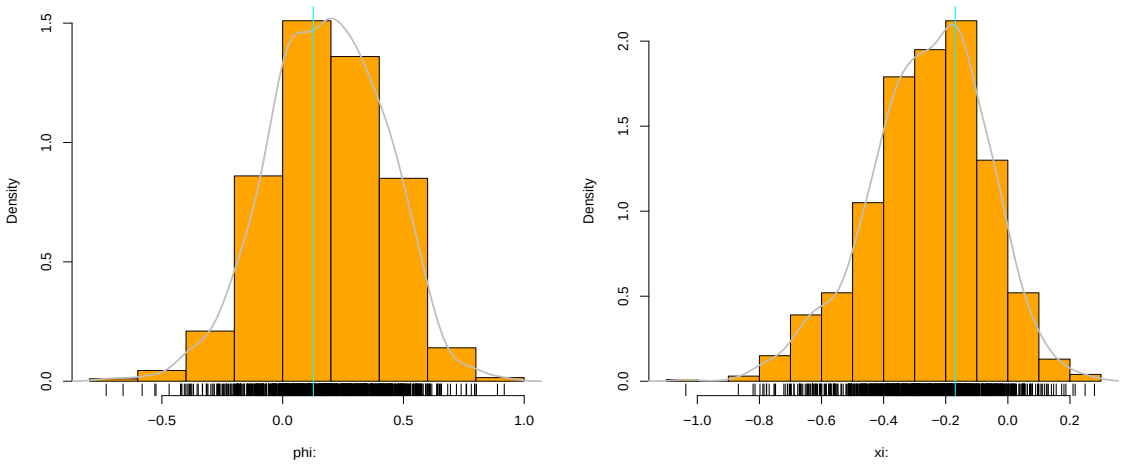


Figure 5.48: Indian-NIFTY: Bootstrap plots.

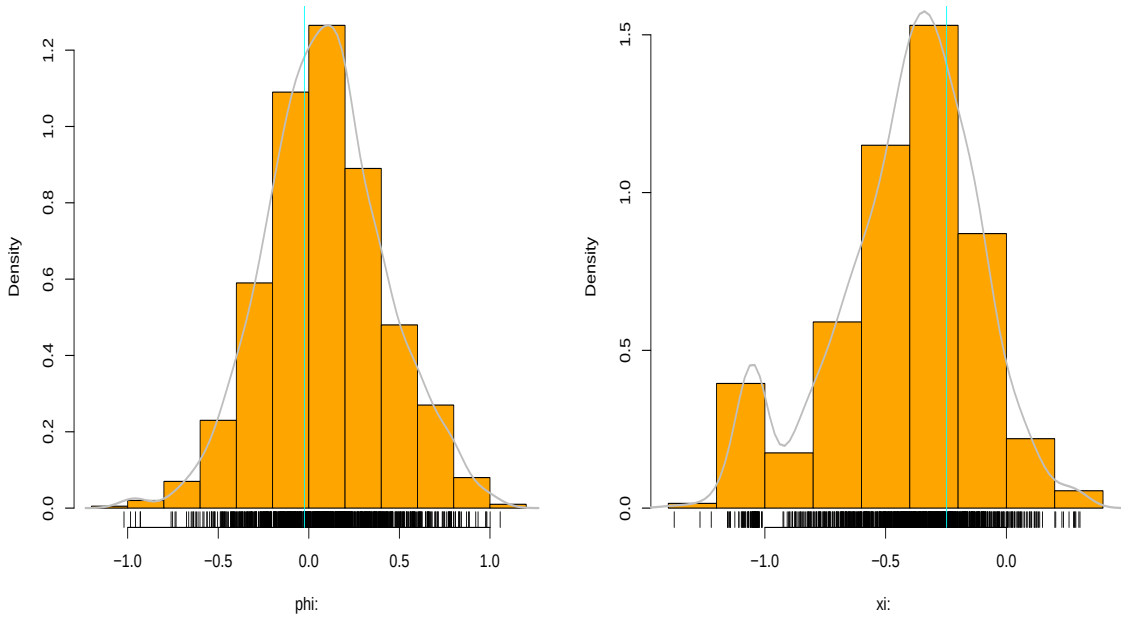


Figure 5.49: Chinese-SHCOMP: Bootstrap plots.

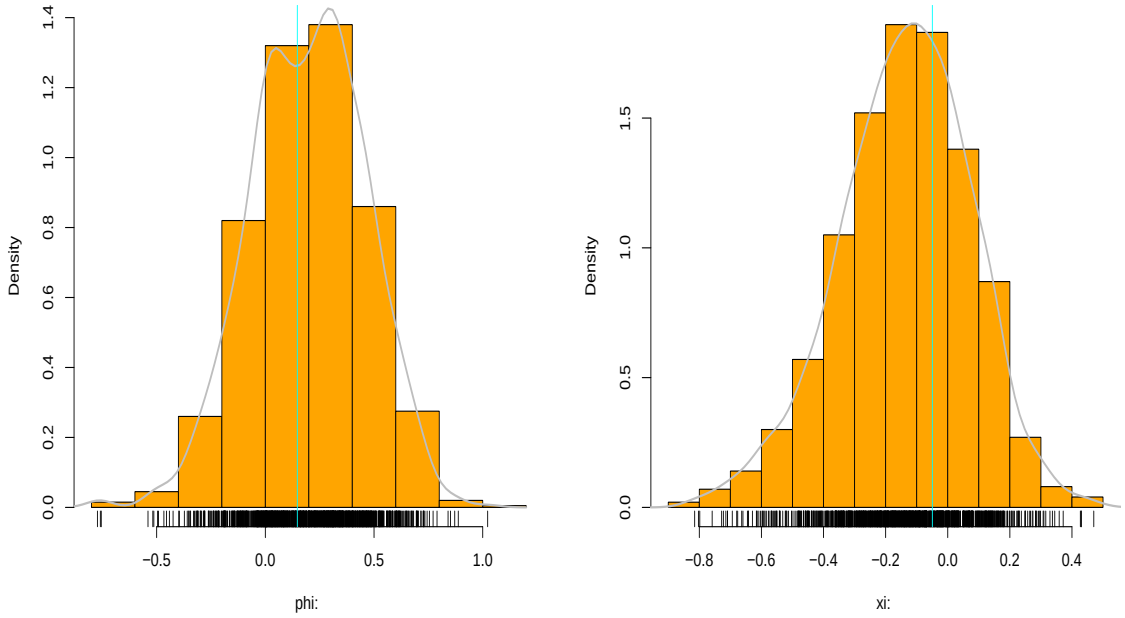


Figure 5.50: South African-JALSH: Bootstrap plots.

market has the highest level of risk, followed by the South African (SA) JALSH market, then the Chinese SHCOMP, Brazilian IBOV and Indian NIFTY markets respectively. This finding shows that the Russian IMOEX market is the most risk-prone, while the least risky is the Indian NIFTY market, with the remaining three markets in between them. The Brazilian IBOV market is however very slightly higher than the Indian NIFTY market in risk level as shown in the plot. High investment risk may either yield potential high returns as a reward or a huge loss to an investor.

The boxplots further show the shapes of the risk in each of the markets based on the concentration of the cluster-maxima observations on the scale. From Figure 5.51, the distribution of the risk observations in the IBOV, NIFTY, SHCOMP and JALSH markets are all skewed to the right, while the IMOEX market is near-symmetric.

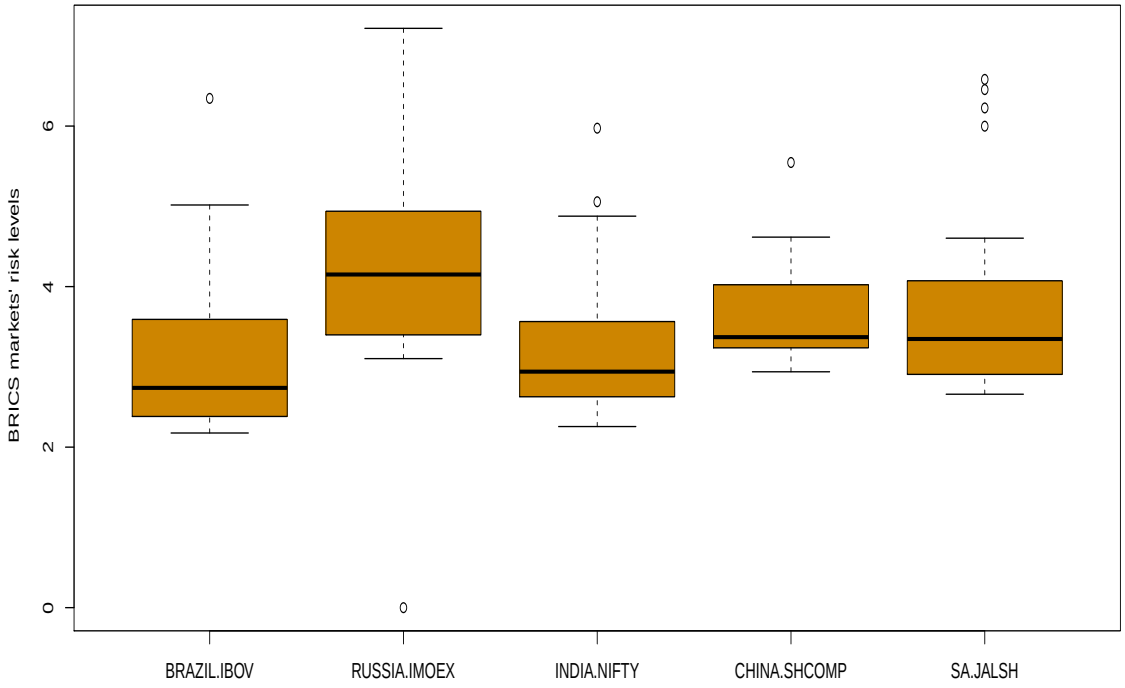


Figure 5.51: Risk hierarchy: comparing the BRICS markets' risks.

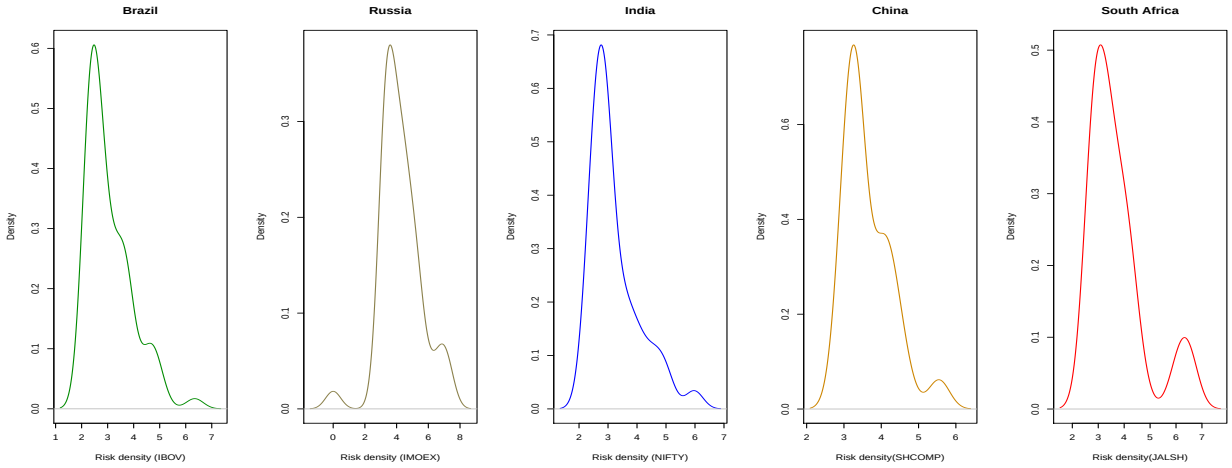


Figure 5.52: Risk densities: comparing the densities of the BRICS markets' risks.

Chapter 6

Extremal dependence of the BRICS stock markets

6.1 Introduction

This chapter of the study focuses on modelling and analysis of extremal (asymptotic) dependence of the pairwise combinations of the five BRICS stock markets. Based on this, the following 10 pairs of the BRICS bloc will be modelled jointly: (1) the Brazilian IBOV and Russian IMOEX markets, (2) the Brazilian IBOV and Indian NIFTY markets, (3) the Brazilian IBOV and Chinese SHCOMP markets, (4) the Brazilian IBOV and South African JALSH markets, (5) the Russian IMOEX and Indian NIFTY markets, (6) the Russian IMOEX and Chinese SHCOMP market, (7) the Russian IMOEX and South African JALSH markets, (8) the Indian NIFTY and Chinese SHCOMP markets, (9) the Indian NIFTY and South African JALSH markets, (10) the Chinese SHCOMP and South African JALSH markets.

6.2 Multivariate extreme value modelling

Multivariate modelling context can either refer to modelling of multiple random variables at various locations or a single variable at multiple locations, or even a single variable at multiple locations. The MEVT can be used to model the joint distribution

of a multivariate process with dependence. In this study, the five BRICS financial variables are modelled in pairwise combinations using the CMEV model and bivariate point process.

6.3 Conditional multivariate extreme value model

This study applies the multivariate modelling method of Heffernan and Tawn (2004) for the extremal dependence modelling, by conditioning the dependence structure on a variable exceeding a suitably high threshold (Southworth et al., 2017). As an illustration using the BRICS markets' variables, if the threshold exceedance of one of the markets' variables is given, then the conditional distribution of the remaining four markets can be described using a regression type approach as described in Section 3.18.3. The modelling process is carried out by first fitting the GPD to the margins, then the CMEV model is used for the dependence modelling. But prior to modelling the dependence, the paired variables are transformed into standardised Laplace margins. The modelling and analytical pattern begin with various multivariate exploratory data analysing plots as described in Section 6.3.1.

6.3.1 Multivariate exploratory plots

An insightful examination using exploratory plots can be made into the pairwise dependence of the BRICS variables under consideration via a pairwise scatterplot. But it should be acknowledged that pairwise dependence between variables in the data body does not automatically indicate extremal dependence (Southworth et al., 2017). Section 6.3.2 and Figure 6.1 describe the exploratory scatterplots of the five BRICS stock markets, while Figures 6.2 to 6.6 display the scatterplots of each stock market variable against the rest.

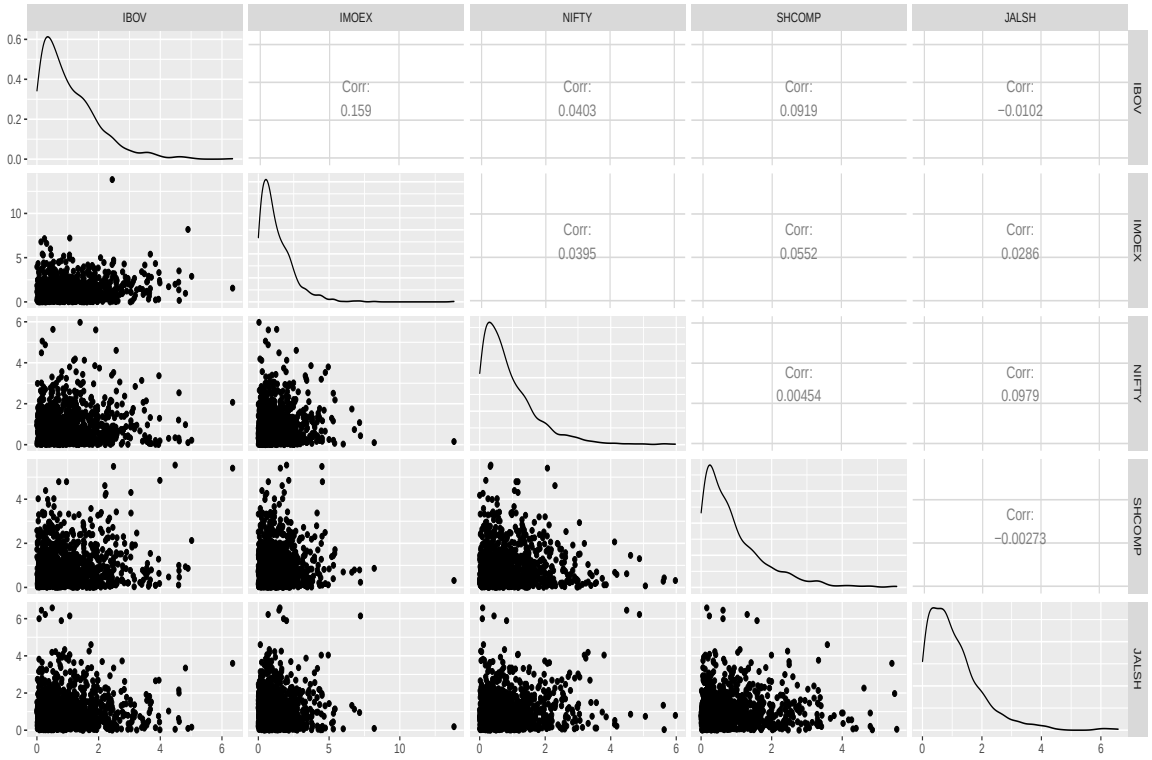


Figure 6.1: Pairwise scatterplot of the markets data.

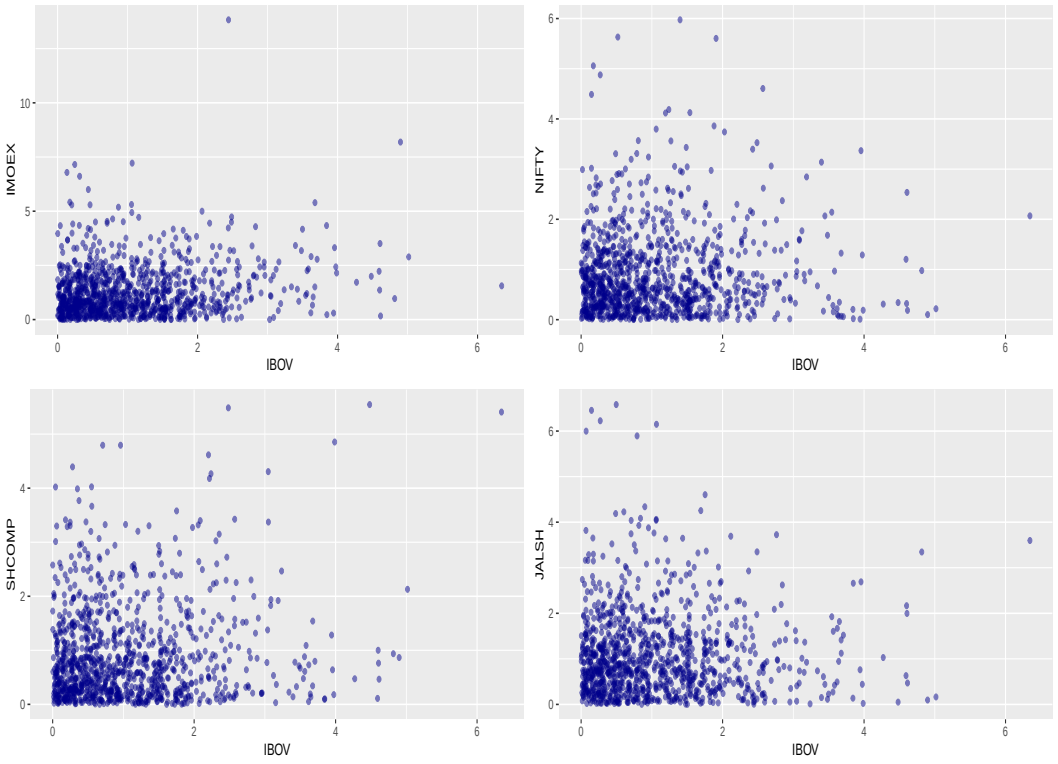


Figure 6.2: Scatterplots of IBOV against the other indices.

6.3.2 Exploratory plots of the five BRICS markets

The pairwise scatterplots of the BRICS residual data as shown in Figure 6.1 mirror absence of extremal dependence between many of the paired variables, while a few market-pairs like the Brazilian IBOV and Russian IMOEX, Brazilian IBOV and Chinese SHCOMP, Indian NIFTY and South African JALSH, and Chinese SHCOMP and Russian IMOEX markets show some low levels of extremal dependence. The highest correlation value is between the Brazilian IBOV and Russian IMOEX markets, and it is suggesting the markets pair with the highest extremal dependence among the entire BRICS markets pairwise combinations.

A closer view on the spread of the scatterplots can be shown when one of the market variables is plotted against each of the others as displayed in Figures 6.2 to 6.6. The quartet pairwise plots in the figures suggest near independence or very weak dependence across the board except between each pair of Brazilian IBOV and Russian IMOEX, Brazilian IBOV and Chinese SHCOMP, and Indian NIFTY and South African JALSH markets that displays weak dependence.

Pairwise extremal dependence in each of the BRICS markets' paired associations can be inspected with the plots of summary statistics χ (Chi) and $\bar{\chi}$ (Chi-bar) described by Coles et al. (1999) as:

- (x) As the quantile u tends to 1, the limiting value of the function $\bar{\chi}(u)$ plot gives a diagnostic to know if asymptotic dependence is exhibited in the residual data. A limiting value where $\bar{\chi}(u) = 1$ denotes asymptotic dependence (Southworth et al., 2017).

- (y) The strength of dependence within the asymptotic dependence region, for limiting value in $(x) = 1$, can be measured using the plot of the function $\chi(u)$. Values of this function closer to 1 denotes stronger dependence.

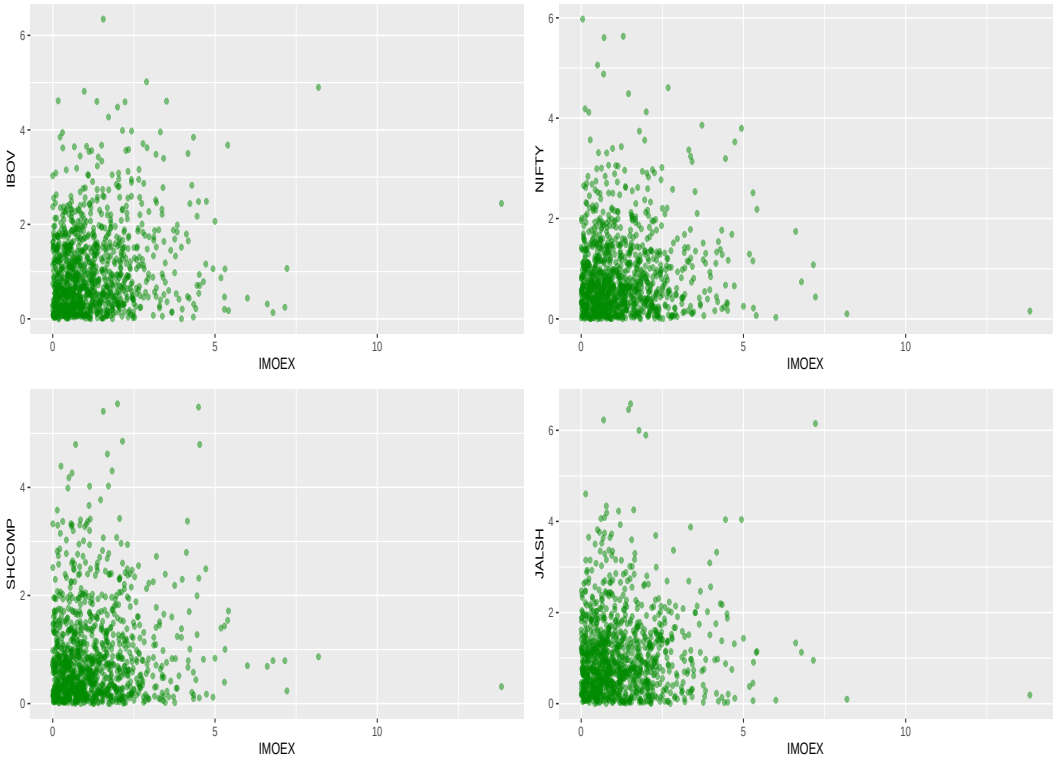


Figure 6.3: Scatterplots of IMOEX against the other indices.

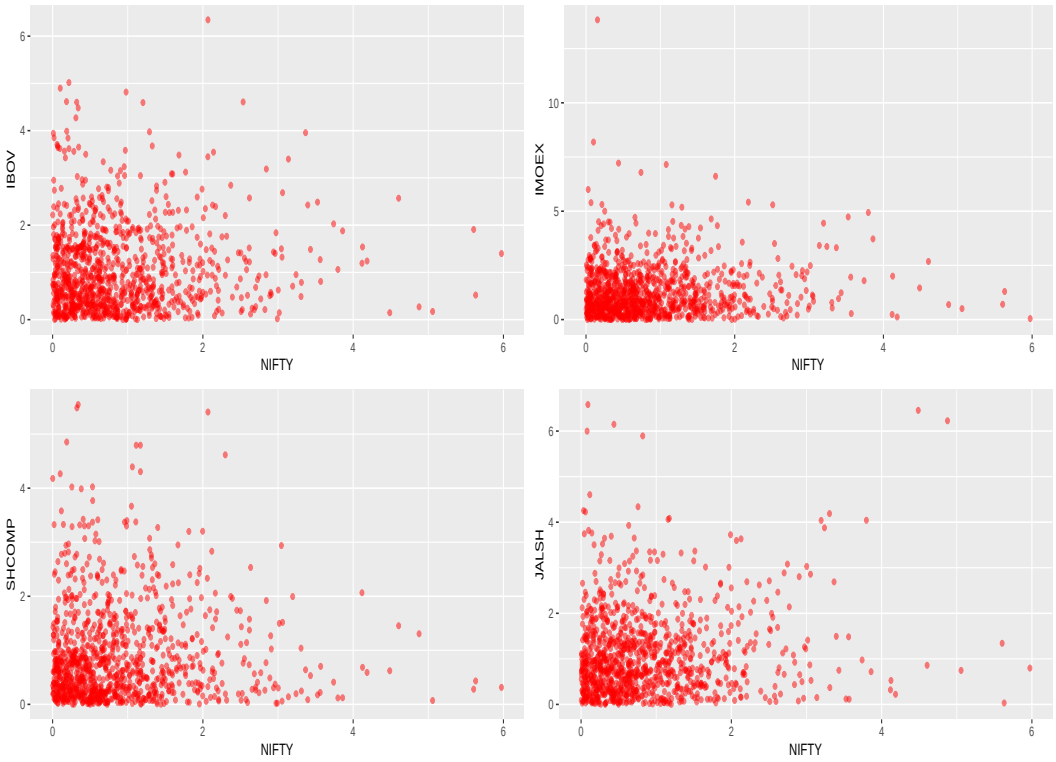


Figure 6.4: Scatterplots of NIFTY against the other indices.

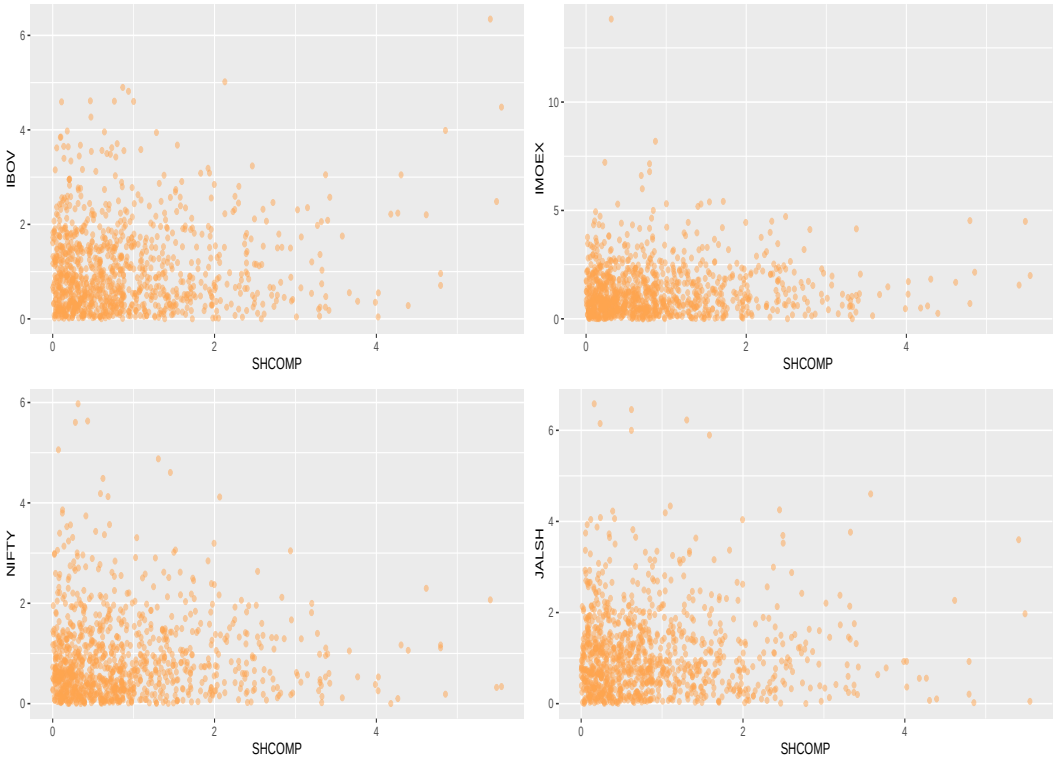


Figure 6.5: Scatterplots of SHCOMP against the other indices.

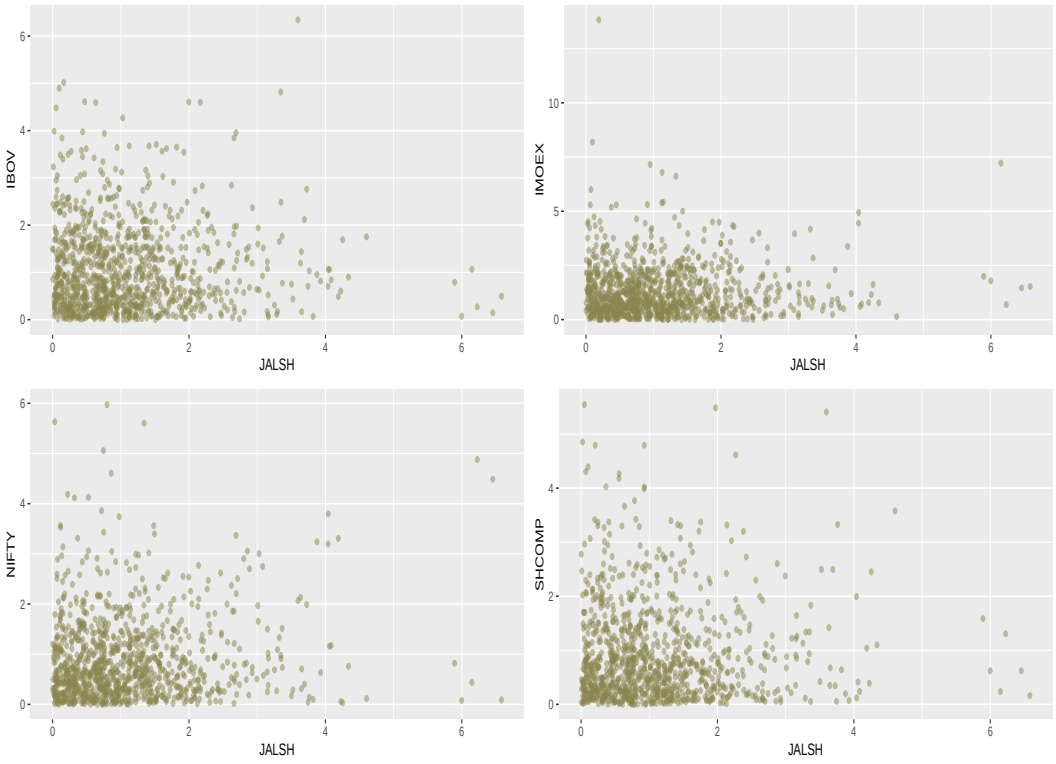


Figure 6.6: Scatterplots of JALSH against the other indices.

- (z) If limit in $(x) < 1$, the strength of dependence can be measured within the asymptotic independence region only by examining $\bar{\chi}(u)$, since $\chi(u) = 0$ here (Coles, 2001). This in other words means that the plot of summary statistic $\chi(u)$ is not relevant for strength of dependence measurement within the asymptotic independence region, but can show levels or a measure of the strength of dependence only within the asymptotic dependence region. The limiting value of this function $\bar{\chi}(u)$ as $u \to 1$ summarises the strength of dependence, where values closer to zero suggest asymptotic near independence. Furthermore, positive (negative) values closer to 1 (-1) indicate stronger positive (negative) dependence (Southworth et al., 2017).

An alternative diagnostic method for inspecting the pairwise extremal dependence is through the multivariate conditional Spearman's (MCS) correlation coefficient ρ introduced by Schmidt and Schmitt (2007). This approach is used to augment the diagnostics of the summary statistics $\bar{\chi}$ and χ for better comprehension of the BRICS markets' dependence structure.

For market pair Brazilian IBOV and Russian IMOEX, and pair Brazilian IBOV and Chinese SHCOMP in Figures 6.7 and 6.10 respectively, the exploratory plots of empirical estimates of $\bar{\chi}$ increase towards the right with the inclusion of 1 as a possible limit, which potentially indicate asymptotic dependence. The χ plots in the respective figures indicate moderately strong positive dependence within the class. The plots of the multivariate conditional Spearman's (MCS) ρ with the 95% confidence interval spread in Figures 6.8 (top panel) and 6.11 (top panel) for the two market pairs further support the evidence of moderately strong positive dependence. However, the estimated strength of dependence is greater between the former pair of markets than the latter. These results are suggesting that we should envisage some moderately strong levels of extremal dependence between each of these paired markets when the CMEV modelling is carried out.

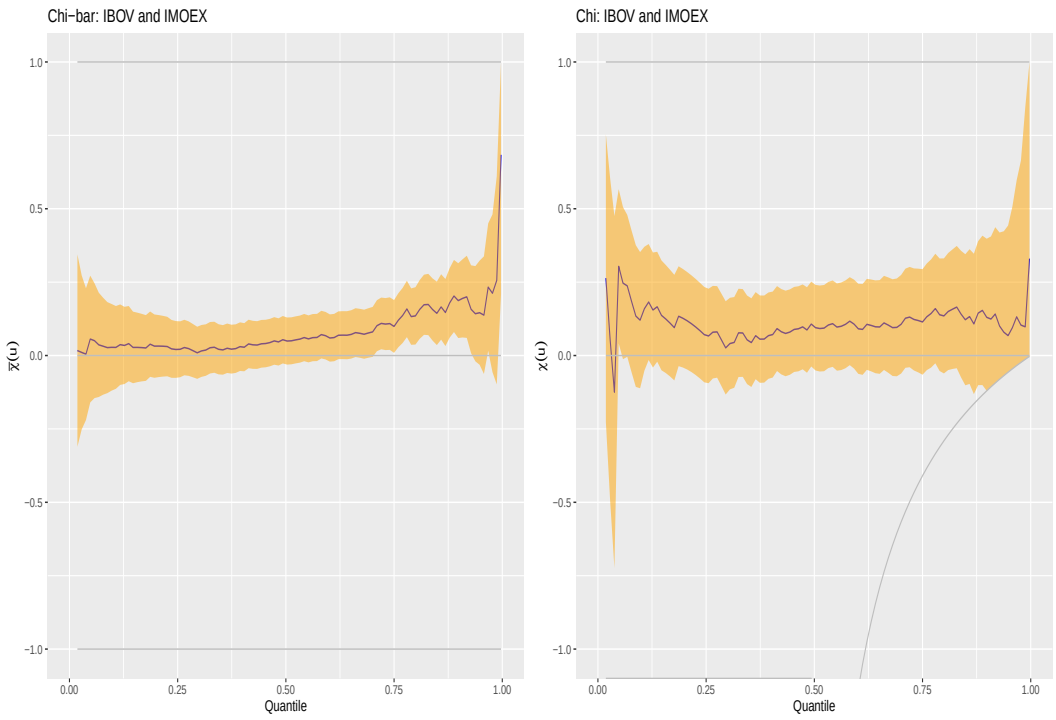


Figure 6.7: Pairwise extremal dependence exploratory plots: IBOV and IMOEX indices.

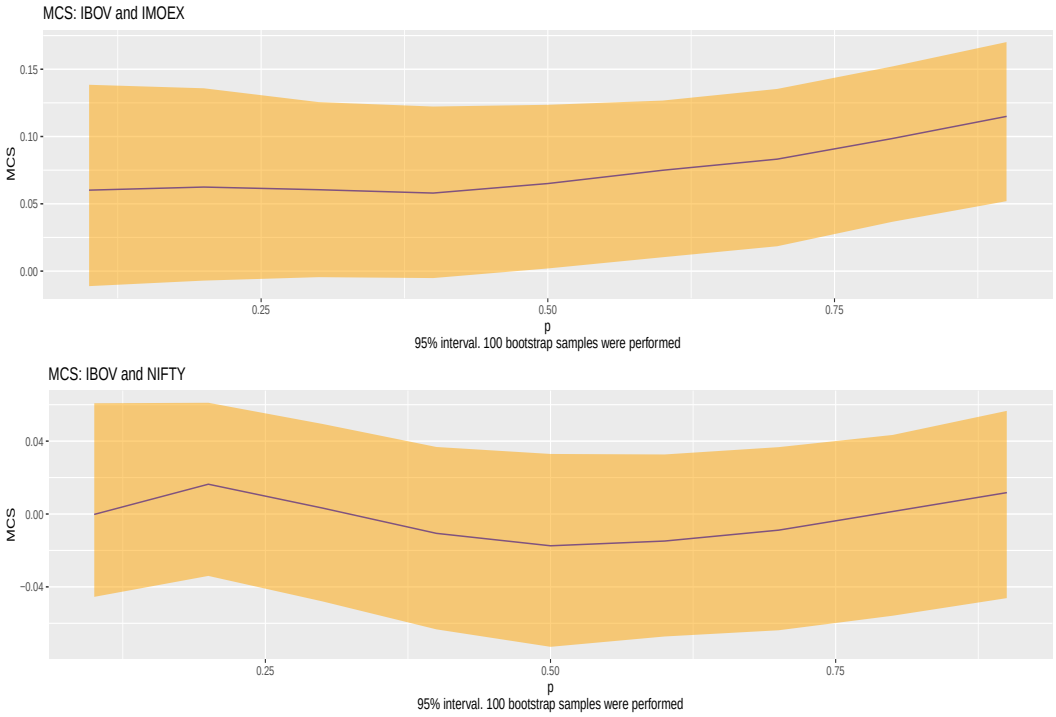


Figure 6.8: MCS ρ plots: IBOV and IMOEX indices (top panel) with IBOV and NIFTY indices (bottom panel).

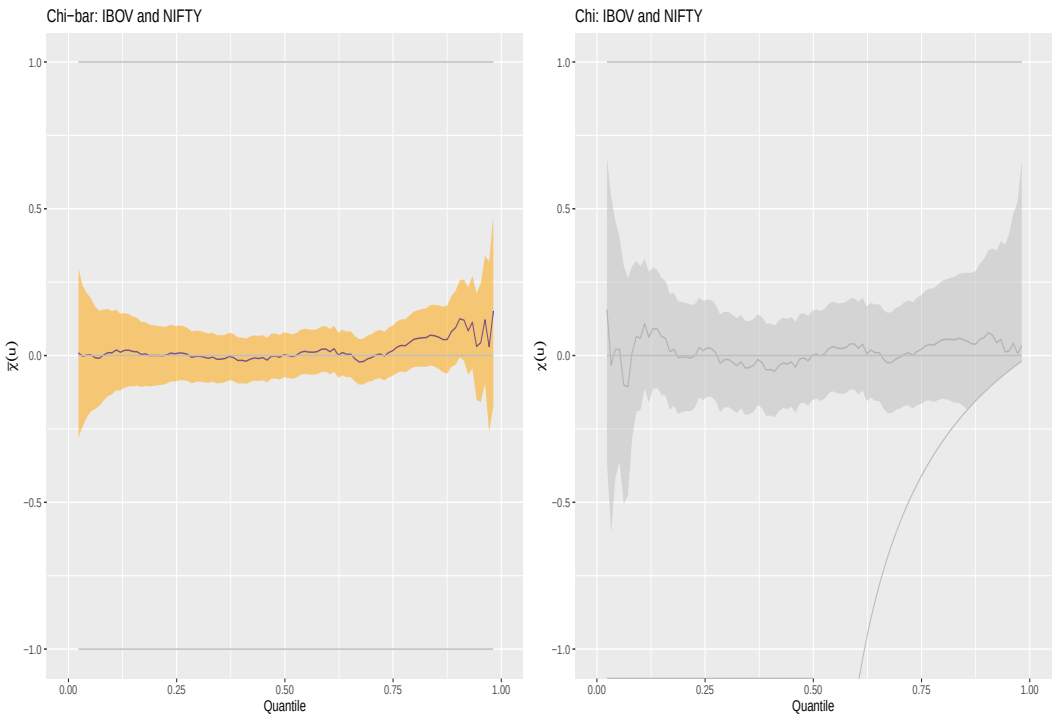


Figure 6.9: Pairwise extremal dependence exploratory plots: IBOV and NIFTY indices.

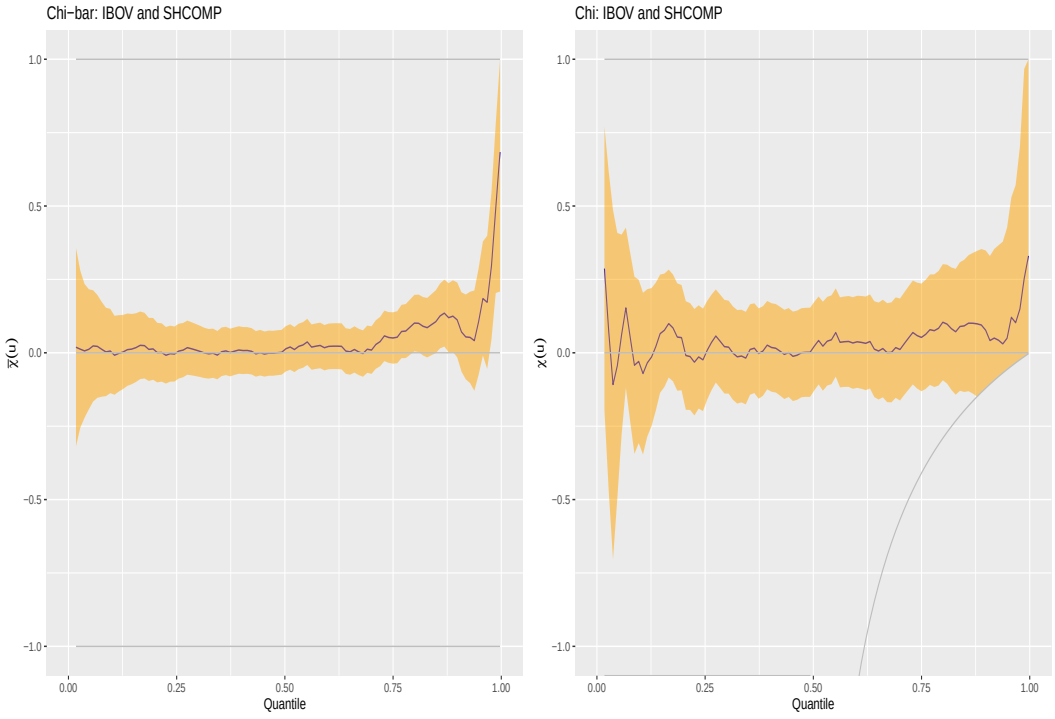


Figure 6.10: Pairwise extremal dependence exploratory plots: IBOV and SHCOMP indices.

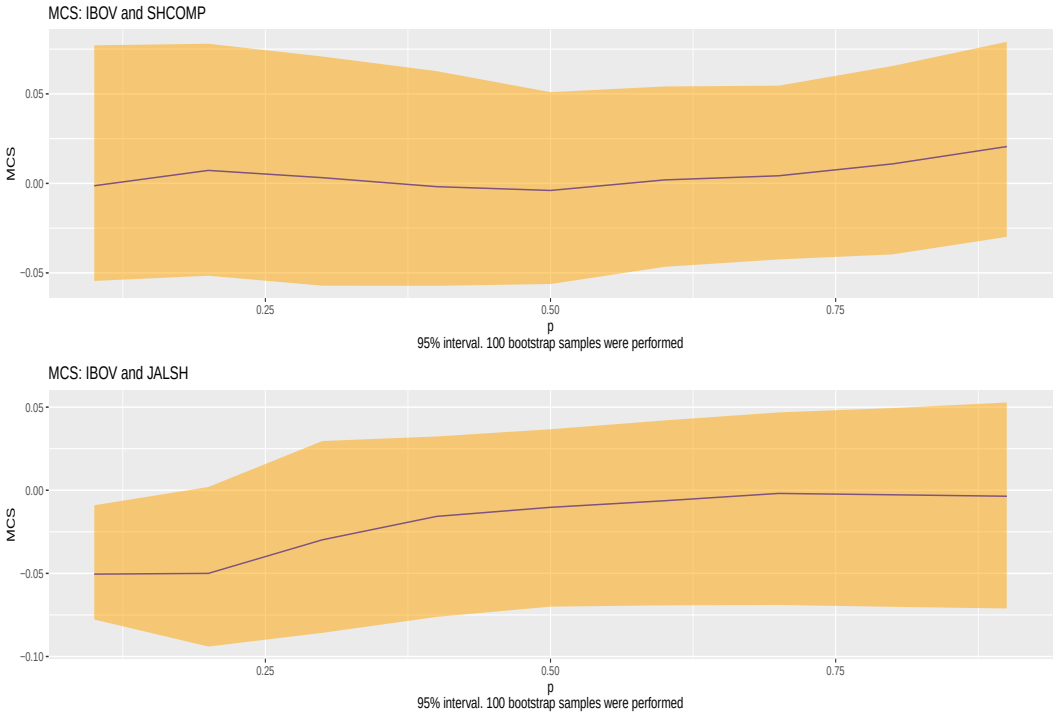


Figure 6.11: MCS ρ plots: IBOV and SHCOMP indices (top panel) with IBOV and JALSH indices (bottom panel).

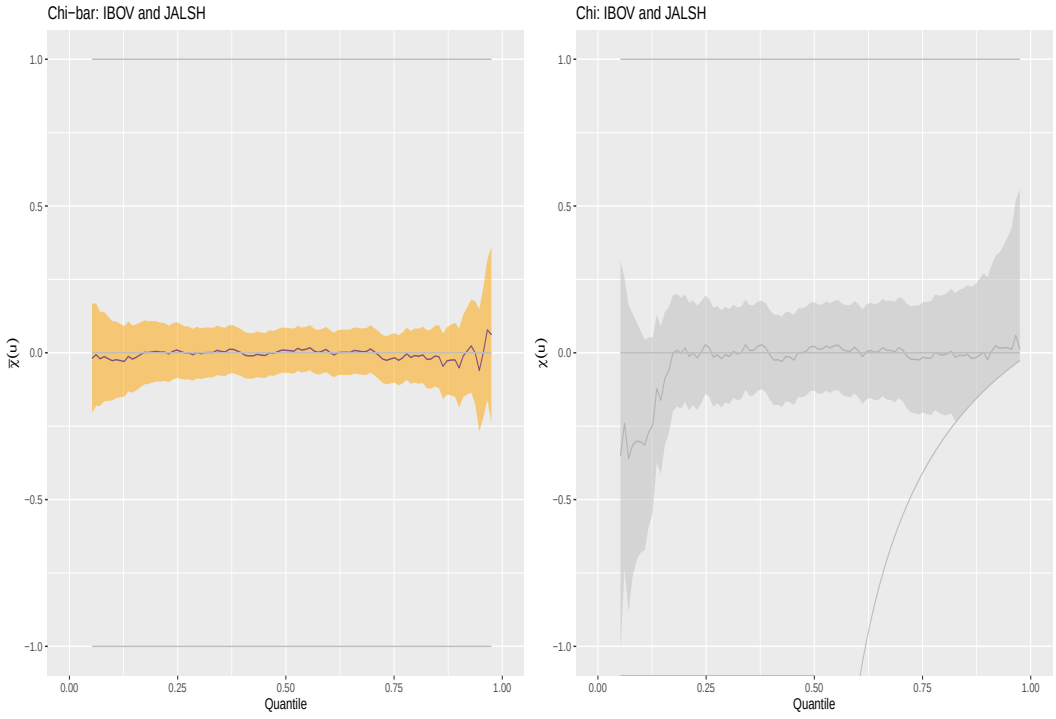


Figure 6.12: Pairwise extremal dependence exploratory plots: IBOV and JALSH indices.

For the pair of Brazilian IBOV and Indian NIFTY markets in Figure 6.9, the exploratory plot of the $\bar{\chi}$ shows asymptotic dependence with a very weak positive dependence as displayed by the χ plot. The plot of the MCS ρ in Figure 6.8 (bottom panel) also indicates a very weak positive dependence, hence this dependence outcome should be expected between the association of these markets when the CMEV model is used. The $\bar{\chi}$ plot for Brazilian IBOV and South African JALSH in Figure 6.12 shows that these markets' association fluctuates between asymptotic independence and asymptotic dependence, with the χ plot displaying near independence to a very weak positive dependence in the class. The plot of the 95% confidence interval MCS ρ in Figure 6.11 (bottom panel) for this pair spreads from a weak negative dependence to a very weak positive dependence. Hence, this is suggesting that when we come to the CMEV modelling, we should expect to see a weak negative dependence or a very weak positive dependence between the extremes of these markets.

For the trio of pair Russian IMOEX and India NIFTY, pair Russian IMOEX and Chinese SHCOMP, and pair Russian IMOEX and South African JALSH, $\bar{\chi}$ in Figures 6.13, 6.15 and 6.16 respectively shows that these markets are likely to be asymptotically dependent, with weak positive dependence within each of the classes as shown by the χ plots. The plots of the respective MCS ρ in Figures 6.14 (top panel), 6.14 (bottom panel), and 6.17 (top panel) also show weak positive dependencies between the extremes of each of the three paired markets. These plots suggest that a weak positive dependence should be expected in each of the associations of the paired markets when the CMEV modelling is applied.

The $\bar{\chi}$ for the pair of Indian NIFTY and Chinese SHCOMP in Figure 6.18 shows that these markets are likely to be asymptotically independent, with weak negative dependence. The plot of χ is irrelevant in an asymptotically independent situation, hence it is not examined here. The plot of the multivariate conditional Spearman's (MCS) ρ in Figure 6.17 (bottom panel) displays an association that moves from a

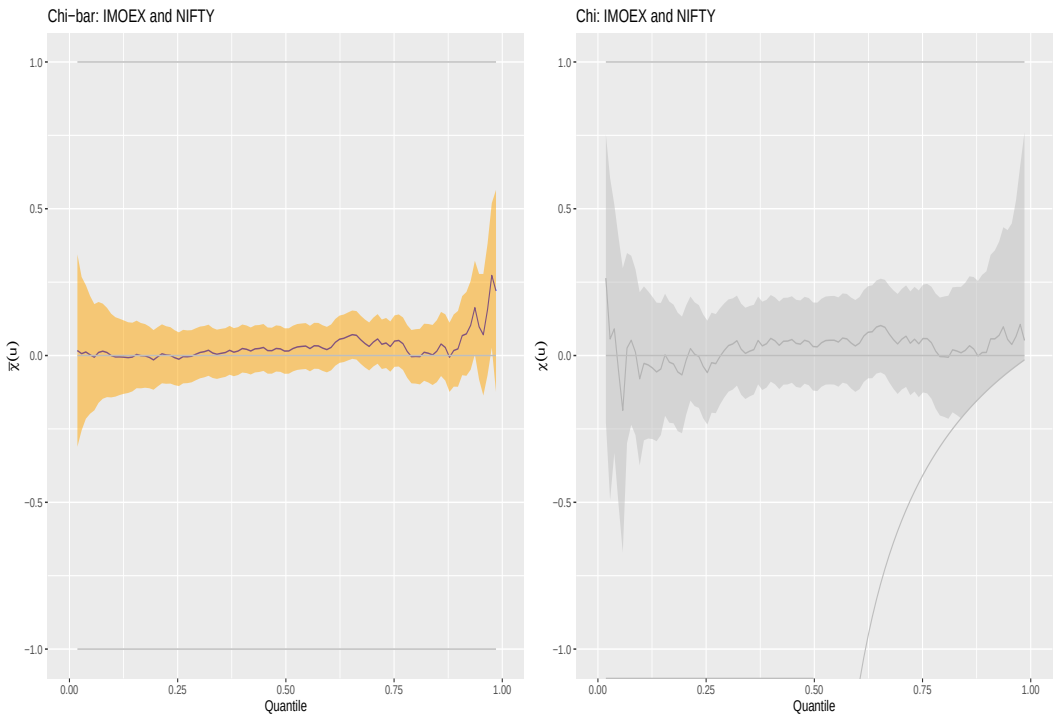


Figure 6.13: Pairwise extremal dependence exploratory plots: IMOEX and NIFTY indices.

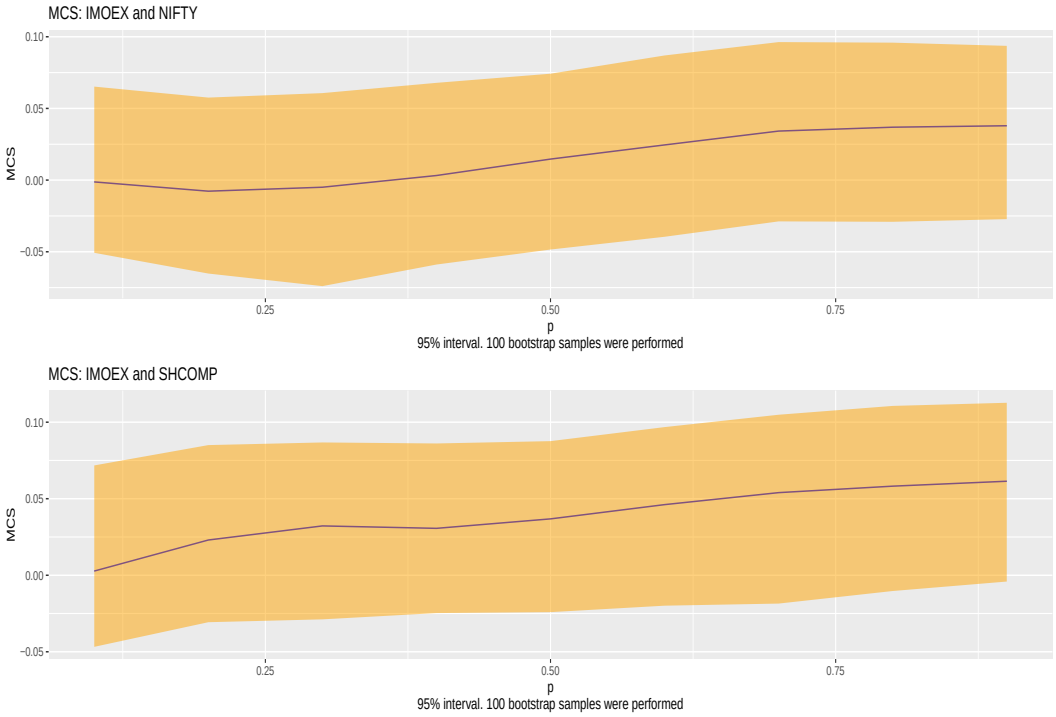


Figure 6.14: MCS ρ plots: IMOEX and NIFTY indices (top panel) with IMOEX and SHCOMP indices (bottom panel).

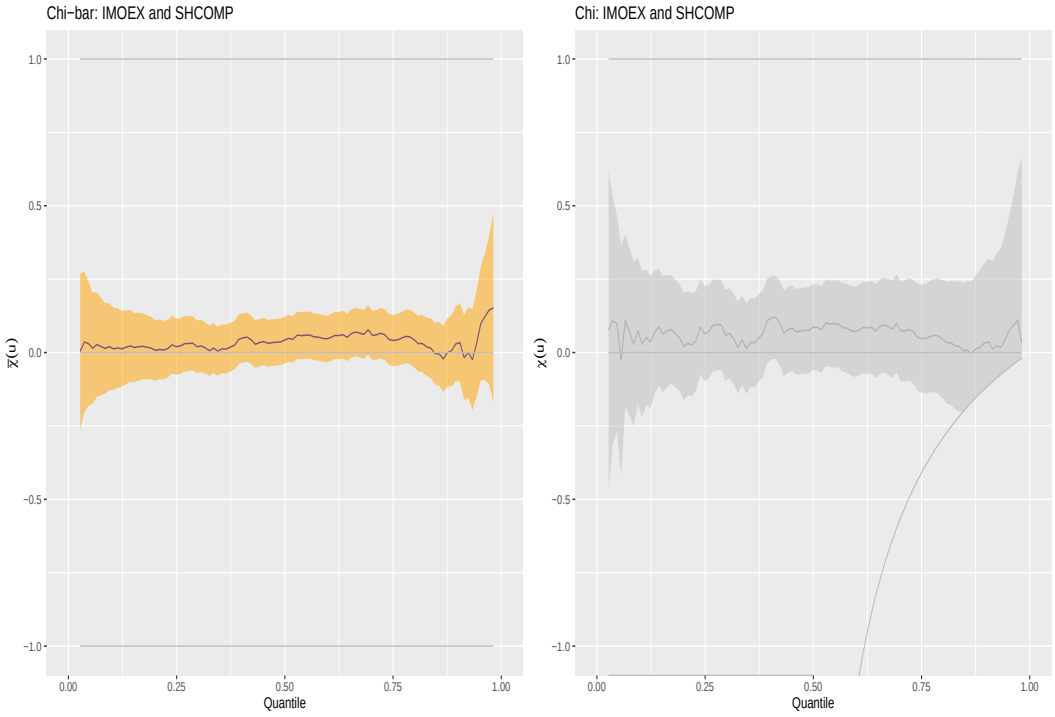


Figure 6.15: Pairwise extremal dependence exploratory plots: IMOEX and SHCOMP indices.

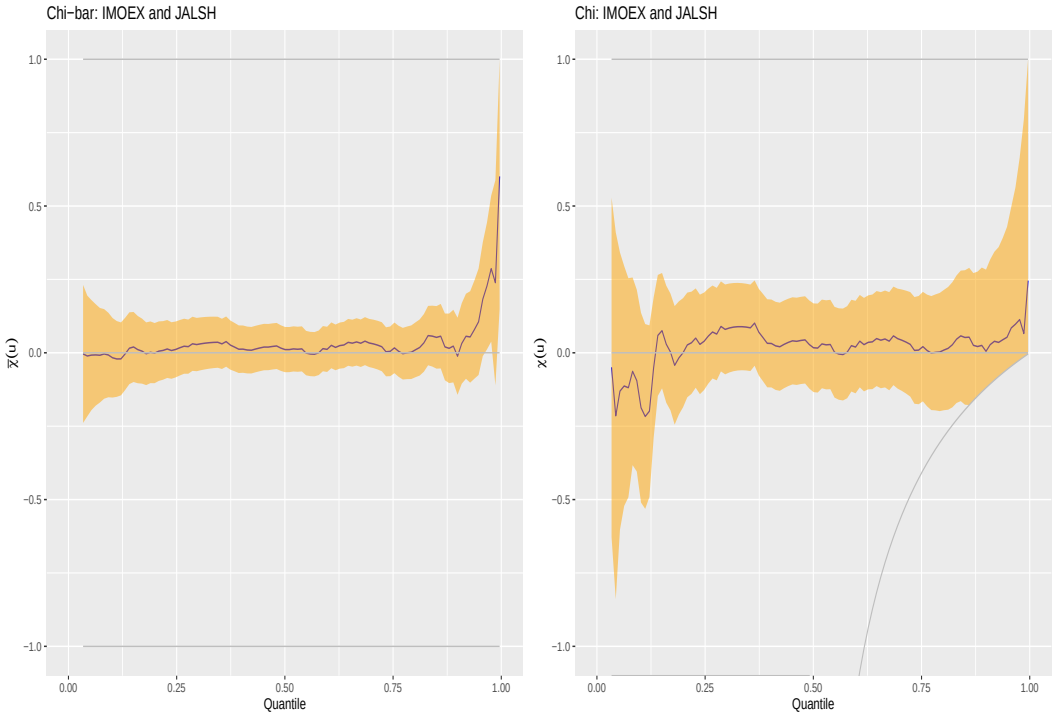


Figure 6.16: Pairwise extremal dependence exploratory plots: IMOEX and JALSH indices.

weak negative dependence to a weak positive dependence in the class. Hence, the three plot-types of $\bar{\chi}$, χ, and MCS ρ altogether suggest a most likely weak negative dependence or a less likely weak positive association between these markets when we come to the CMEV modelling. For the Indian NIFTY and South African JALSH pair in Figure 6.19, it is observed that $\bar{\chi} \to 1$ as $u \to 1$, which indicates possible asymptotic dependence, with the χ plot showing moderately strong positive dependence within the class. The exploratory summary of the MCS ρ in Figure 6.20 (top panel) in addition to those of the summary statistics $\bar{\chi}$ and χ suggest that a moderately strong dependence should be expected between this paired markets with CMEV modelling.

Lastly, for the pair of Chinese SHCOMP and South African JALSH markets in Figure 6.21, the $\bar{\chi}$ plot shows they are likely to be asymptotically dependent with the χ plot displaying near independence to a weak positive dependence within the class. However, as opposed to the plots of $\bar{\chi}$ and χ, the MCS ρ plot with the 95% confidence interval in Figure 6.20 (bottom panel) shows a weak negative association between these markets. Hence, for the proposed extremal dependence modelling using the CMEV model, a weak positive or negative extremal dependence outcome should be expected.

6.3.3 CMEV model fitting and diagnostics

Following the preliminary analysis using the exploratory plots, we then proceed to the extremal dependence modelling with the application of the CMEV model of Heffernan and Tawn (2004). As stated earlier, this conditional multivariate modelling begins with the GPD models fitted to the five BRICS marginal variables, after which the dependence structure is estimated. In other words, the dependence component of the CMEV model also conditions on a variable exceeding a threshold in the same way the GPD models exceedances above a threshold.

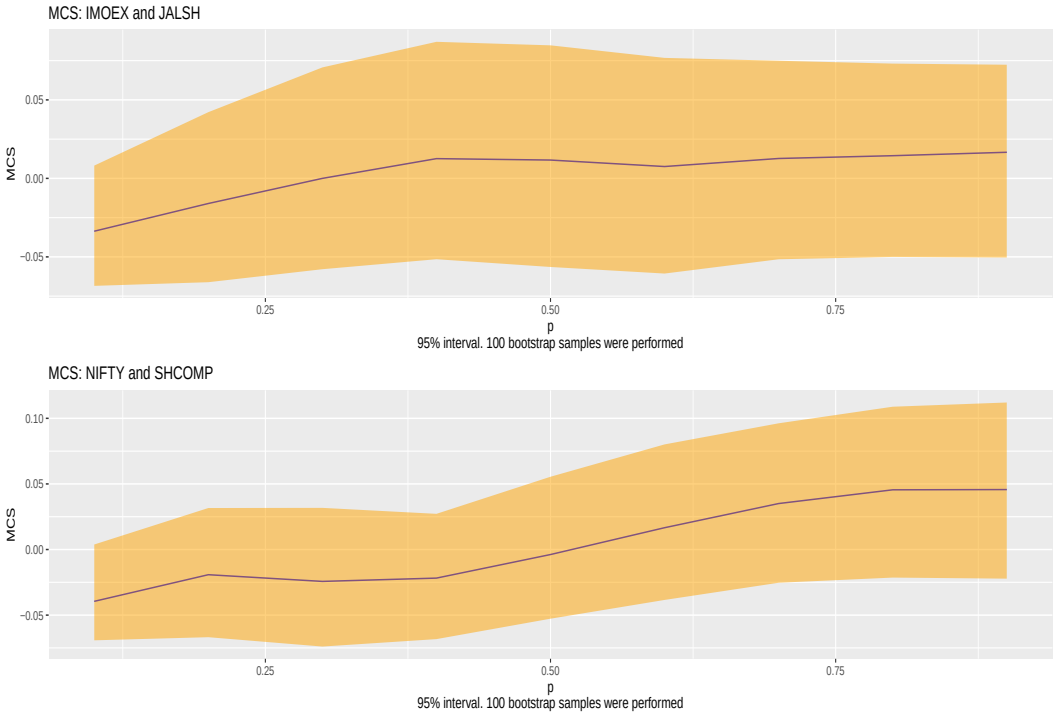


Figure 6.17: MCS ρ plots: IMOEX and JALSH indices (top panel) with NIFTY and SHCOMP indices (bottom panel).

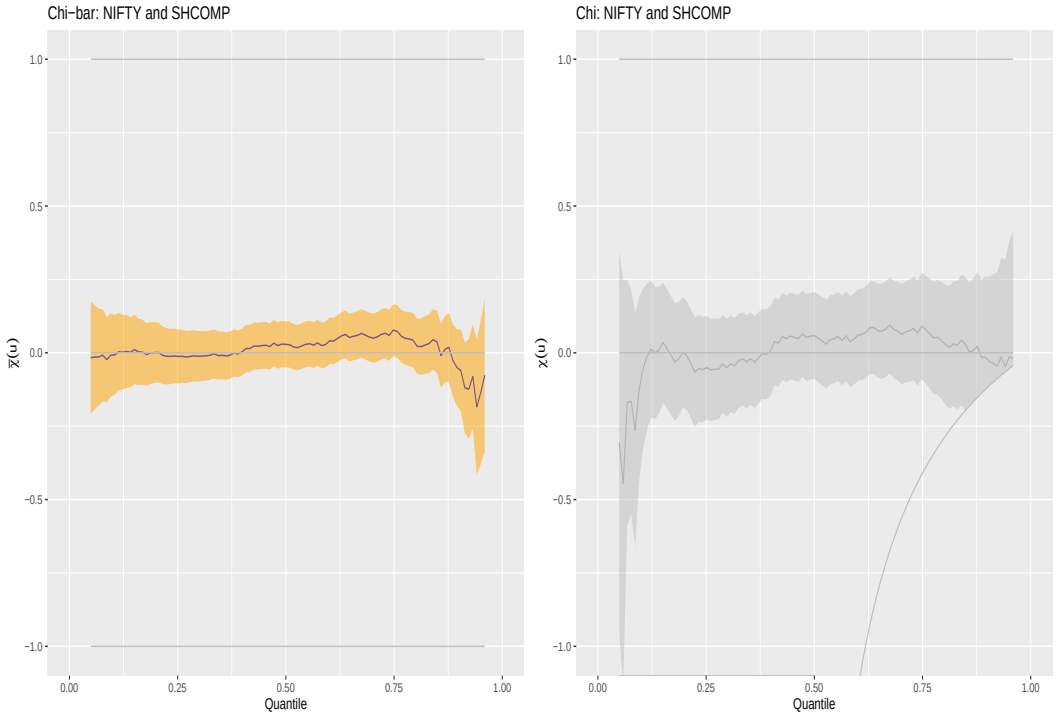


Figure 6.18: Pairwise extremal dependence exploratory plots: NIFTY and SHCOMP indices.

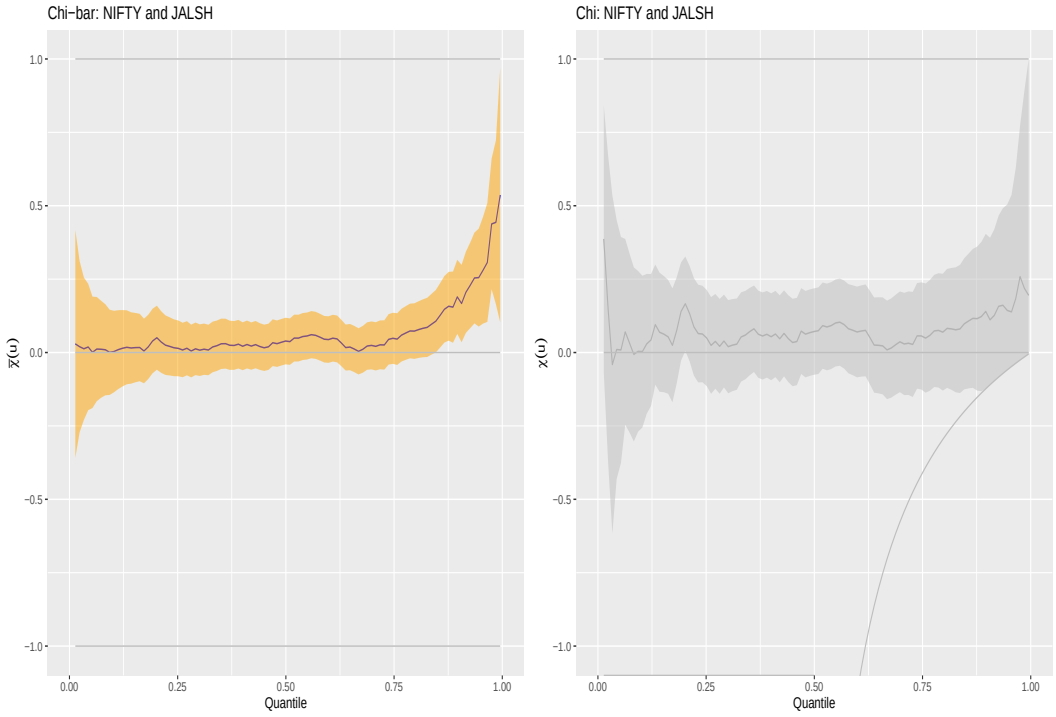


Figure 6.19: Pairwise extremal dependence exploratory plots: NIFTY and JALSH indices.

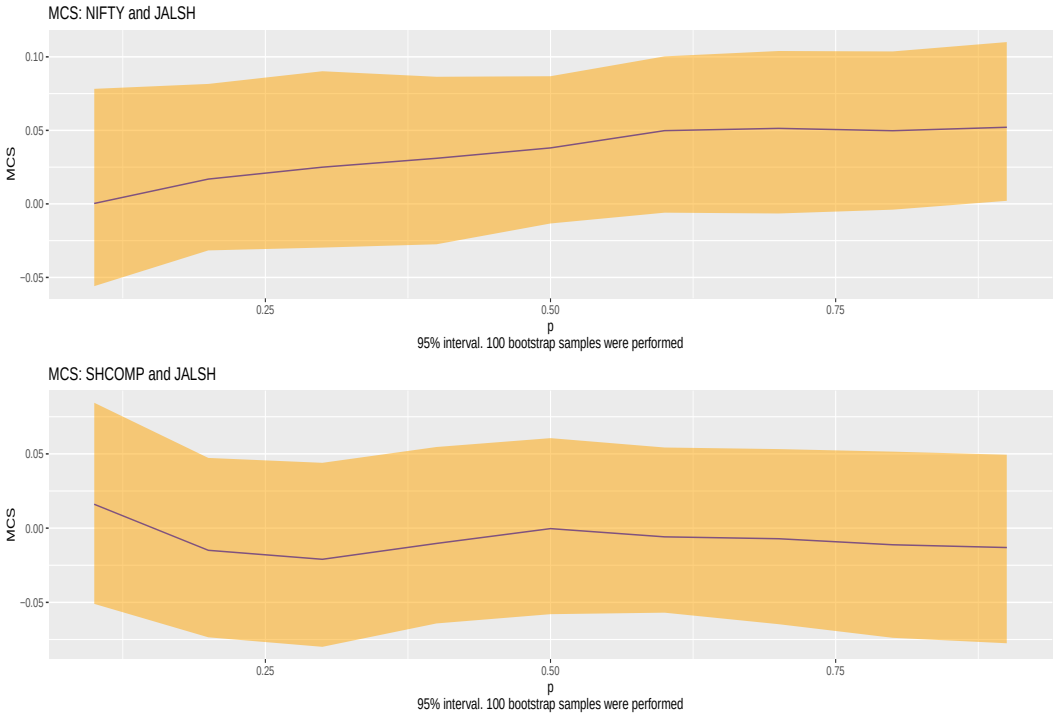


Figure 6.20: MCS ρ plots: NIFTY and JALSH indices (top panel) with SHCOMP and JALSH indices (bottom panel).

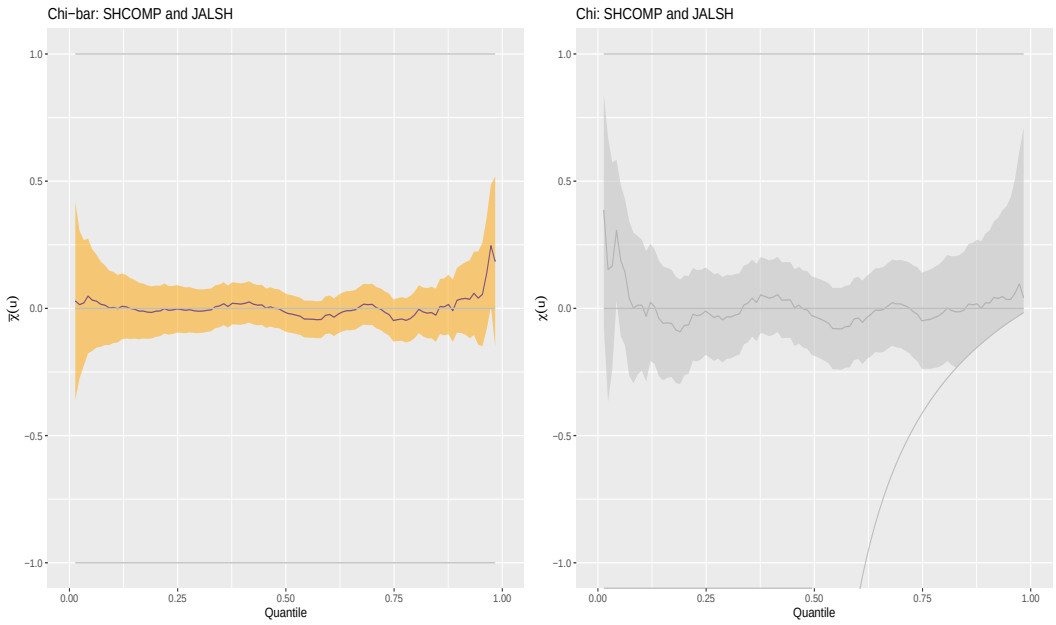


Figure 6.21: Pairwise extremal dependence exploratory plots: SHCOMP and JALSH indices.

6.3.4 Marginal GPD modelling and model diagnostics

Here, the CMEV model is fitted to the BRICS stock dataset, conditioning on each of the five margins one after another. However, we need to specify an appropriate marginal quantile that describes the threshold above which to fit the marginal GPD models for each conditioning variable. To determine this threshold quantile, a series of candidate marginal quantiles were examined, and the validity of each was assessed using quartet diagnostics to ascertain which is the most suitable under the fitted marginal GPD model. The quartet diagnostics are "probability plot", "quantile plot", "return level plot" and "histogram and density". For each market, the examined marginal quantiles were 70th, 75th, 80th, 85th, 90th and 95th percentiles, and the best of these quantiles, based on diagnostics appropriateness, was selected. After thorough examinations of the quartet diagnostic plots under the stated quantiles, it is observed

that the most suitable threshold quantiles for the Brazilian IBOV, Russian IMOEX, Indian NIFTY, Chinese SHCOMP and South African JALSH market-variables using these diagnostics are 70th, 80th, 70th, 70th and 70th percentiles respectively, with the diagnostics displayed in Figures 6.22, 6.23, 6.24, 6.25 and 6.26 in that order. These diagnostic plots are the most accurate when compared with the diagnostics of the remaining unchosen quantiles, hence that informed their selections.

6.3.5 Dependence modelling and model diagnostics

Having obtained the marginal threshold quantiles for the GPD modelling, the CMEV modelling for the extremal dependence parameter estimations is carried out by fitting the model to the dataset in turn, conditioning on each of the five marginal variables. Like the marginal process, the thresholds for the dependence modelling are obtained by examining various candidate quantiles and testing their validity using diagnostic plots to know which is the most appropriate (Southworth et al., 2017). The examined dependence threshold quantiles were 70th, 75th, 80th, 85th, 90th and 95th percentiles. The most suitable threshold quantile for each market variable was selected based on the diagnostic criteria described in Section 6.3.6.

6.3.6 Diagnostics of the dependence model

The diagnostic plots, as shown in Figures 6.27, 6.28, 6.29, 6.30 and 6.31, for the fitted CMEV dependence model introduced by Heffernan and Tawn (2004) can be described using three diagnostic layers produced for each dependent variable. 1) The first layer contains residuals $\mathbf{Z}$ (from the fitted CMEV model) plotted against the conditioning variable's quantile, with a lowess (i.e., locally weighted scatterplot smoothing) curve that shows the local mean of the points. A lowess is a tool that helps to see how variables relate and for predicting or foreseeing trends in a regression analysis by creating a smooth line through scatter plot. 2) The second layer contains the plots

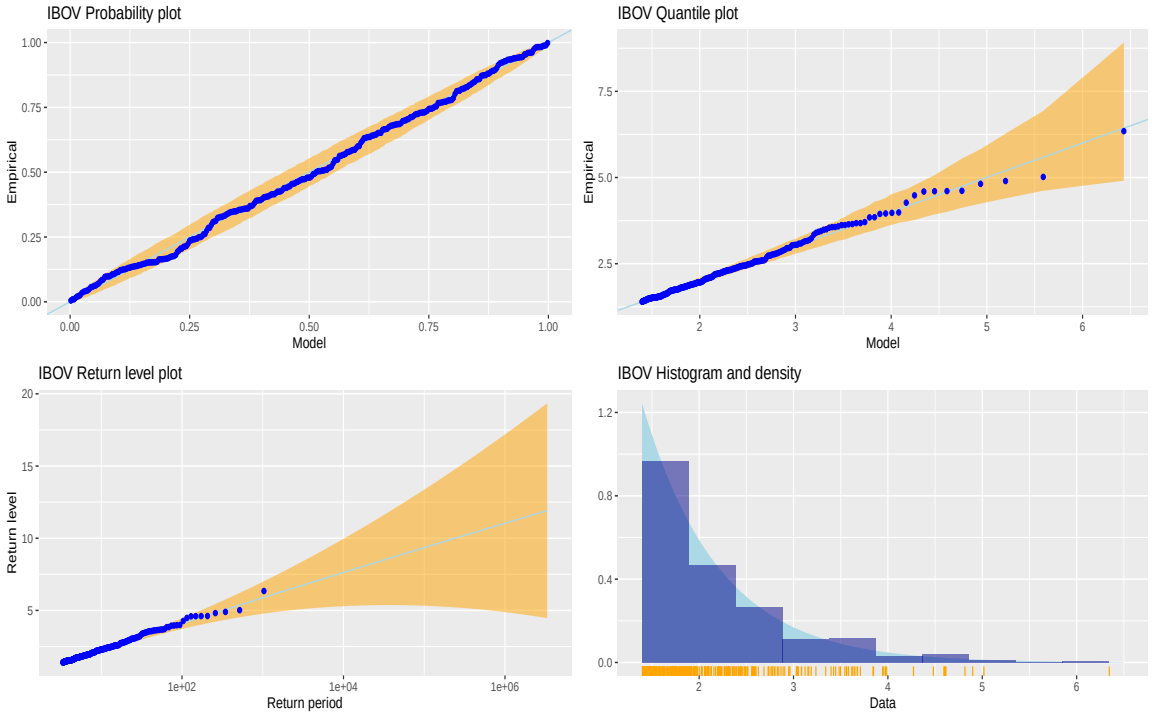


Figure 6.22: Marginal model diagnostics for IBOV variable.

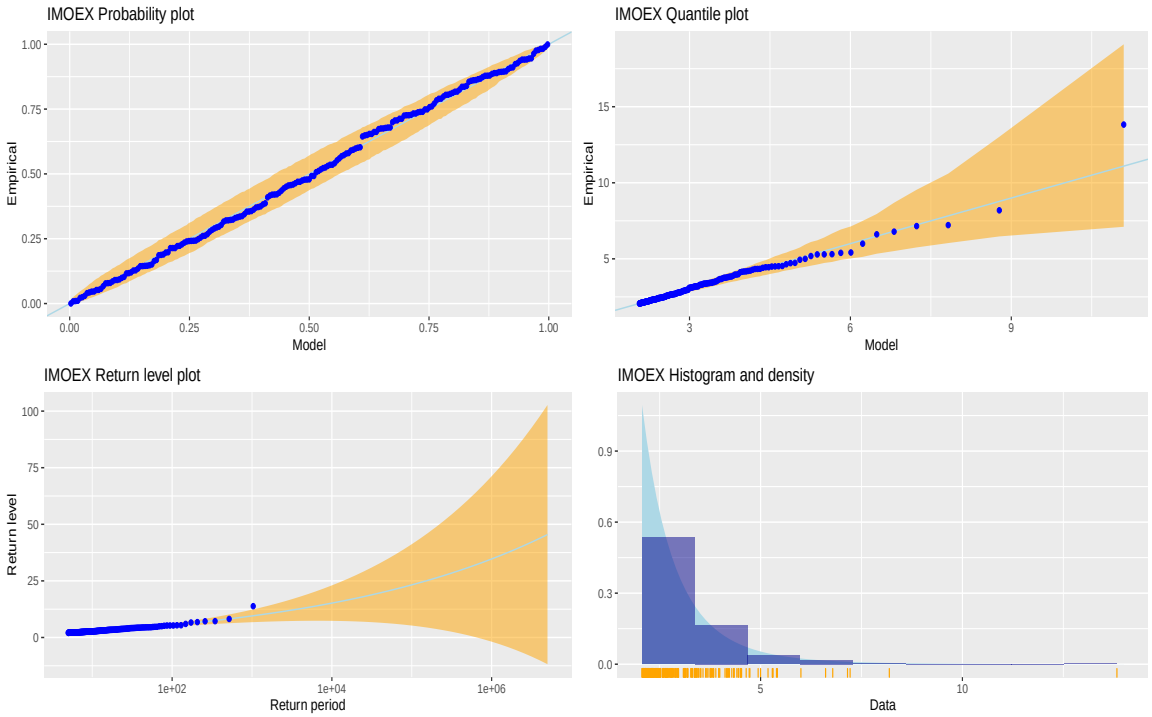


Figure 6.23: Marginal model diagnostics for IMOEX variable.

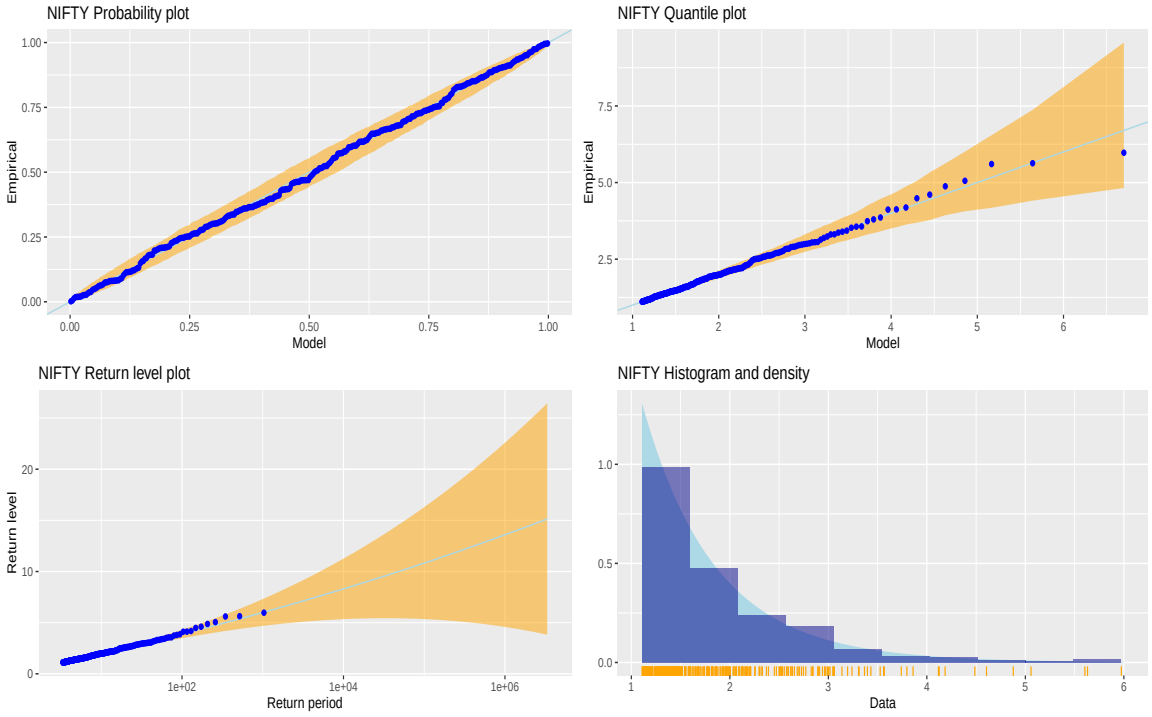


Figure 6.24: Marginal model diagnostics for NIFTY variable.

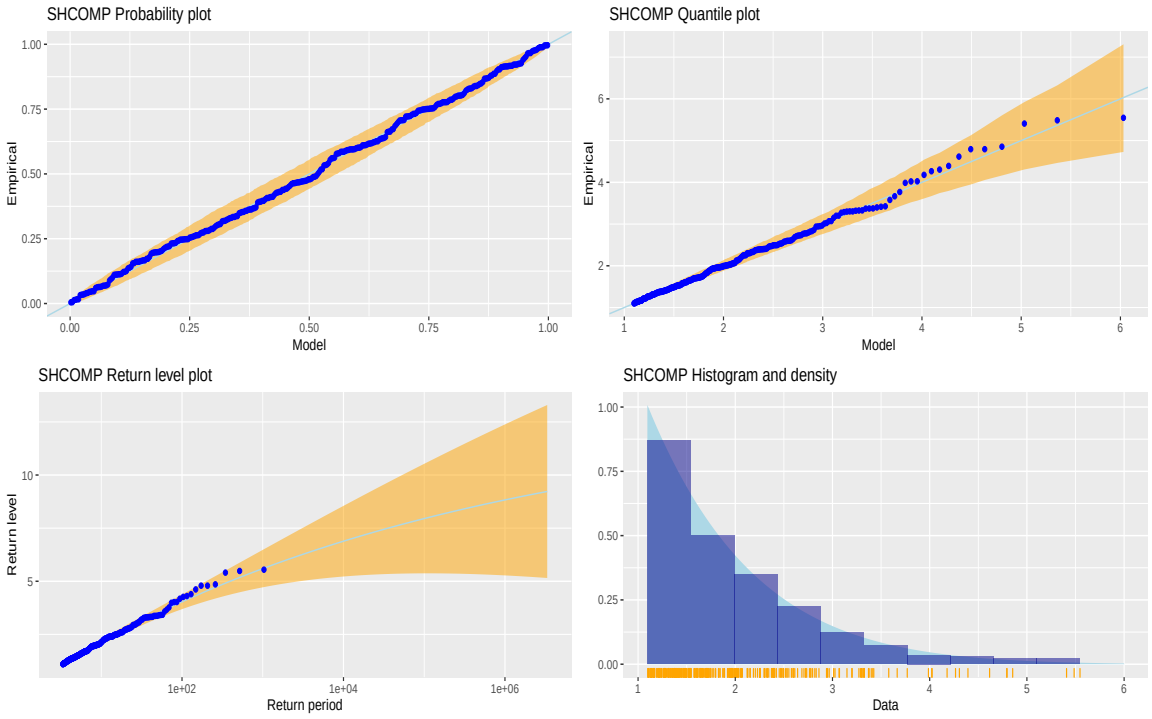


Figure 6.25: Marginal model diagnostics for SHCOMP variable.

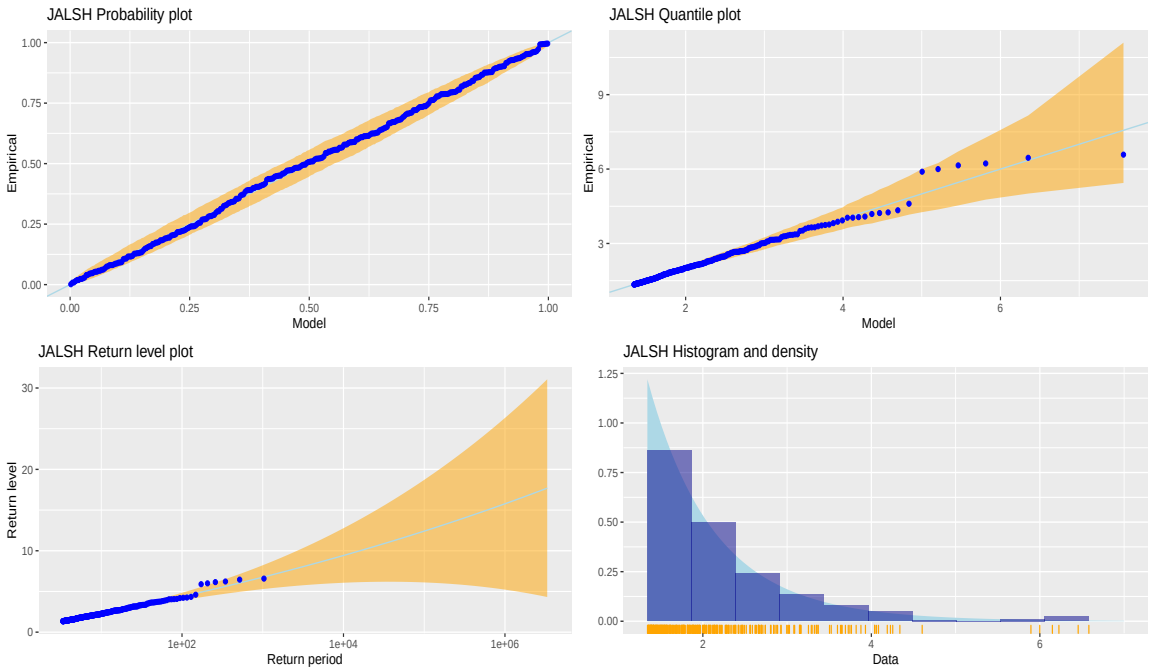


Figure 6.26: Marginal model diagnostics for JALSH variable.

of the absolute value of **Z**-mean(**Z**), where the lowess curve shows the local mean of the points. 3) The third layer displays the plots showing the original data that is not transformed and the quantiles of the CMEV model fit.

As a condition for good fitness, the plots in layers 1 and 2 are meant to display no lowess curve (or scatterplot smoother) that is more or less horizontal (Southworth et al., 2017; Southworth et al., 2020). Or, as stated by Southworth et al. (2020), any trend in the scatter or location of the variables with the conditioning variable violates the assumption of the model that the residuals **Z** are independent of the conditioning variable. Hence, the straighter (i.e., no trend) the lowess curve or scatterplot smoother is, the better the fit. Furthermore, in layer 3, a model that is well fitted is expected to have a satisfactory agreement between the fitted quantile and the scatter plot (i.e., the raw data distribution) (Southworth et al., 2017; Southworth et al., 2020).

From the dependence diagnostic plots in Figures 6.27 to 6.31, it is observed that

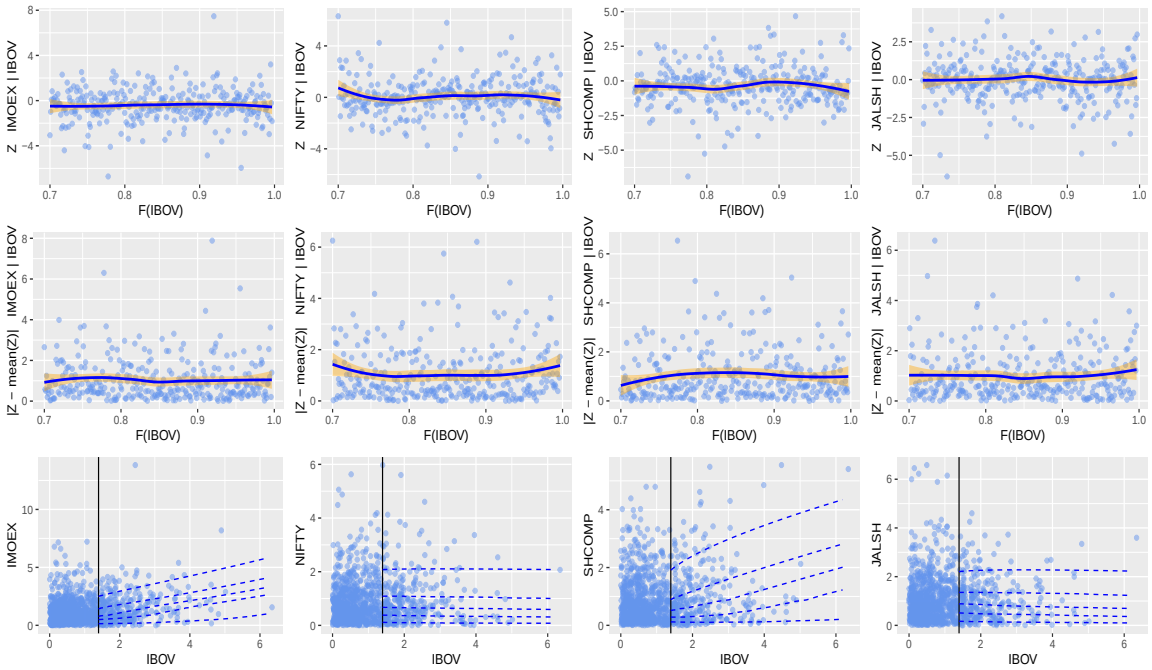


Figure 6.27: Dependence model diagnostics: conditioning on the IBOV variable.

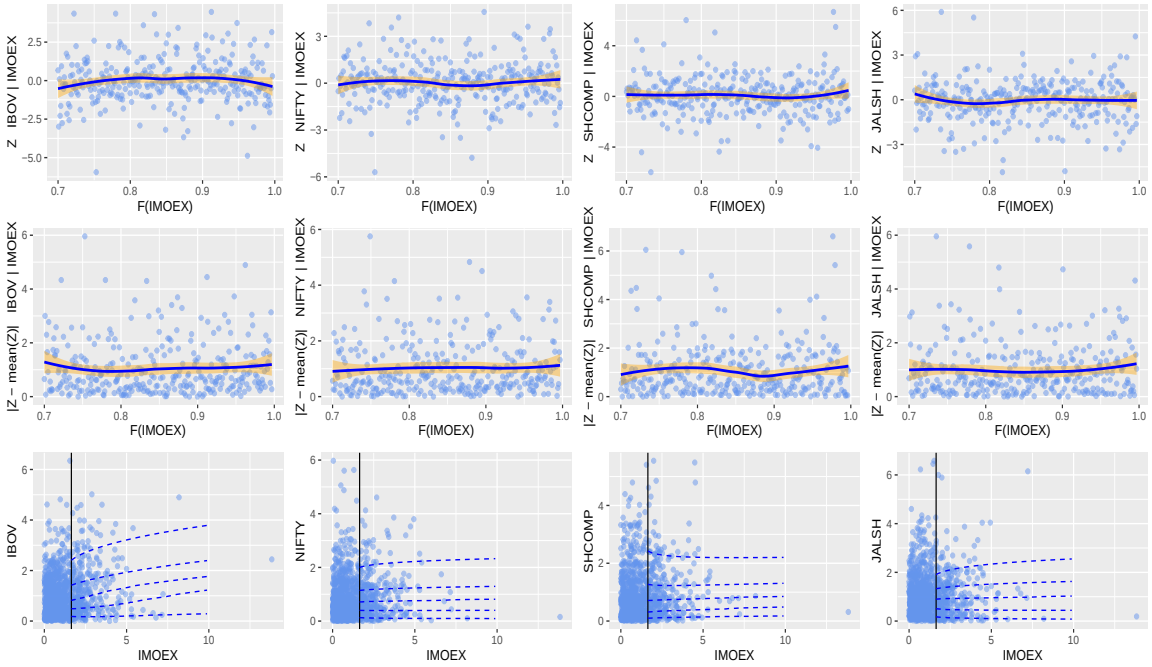


Figure 6.28: Dependence model diagnostics: conditioning on the IMOEX variable.

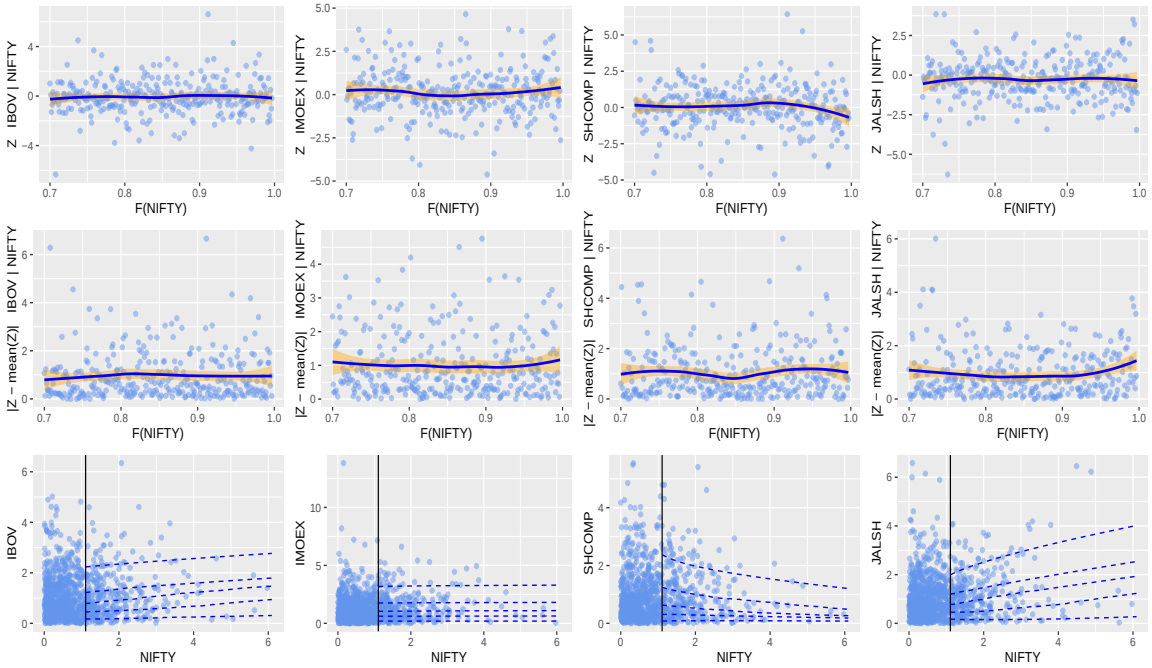


Figure 6.29: Dependence model diagnostics: conditioning on the NIFTY variable.

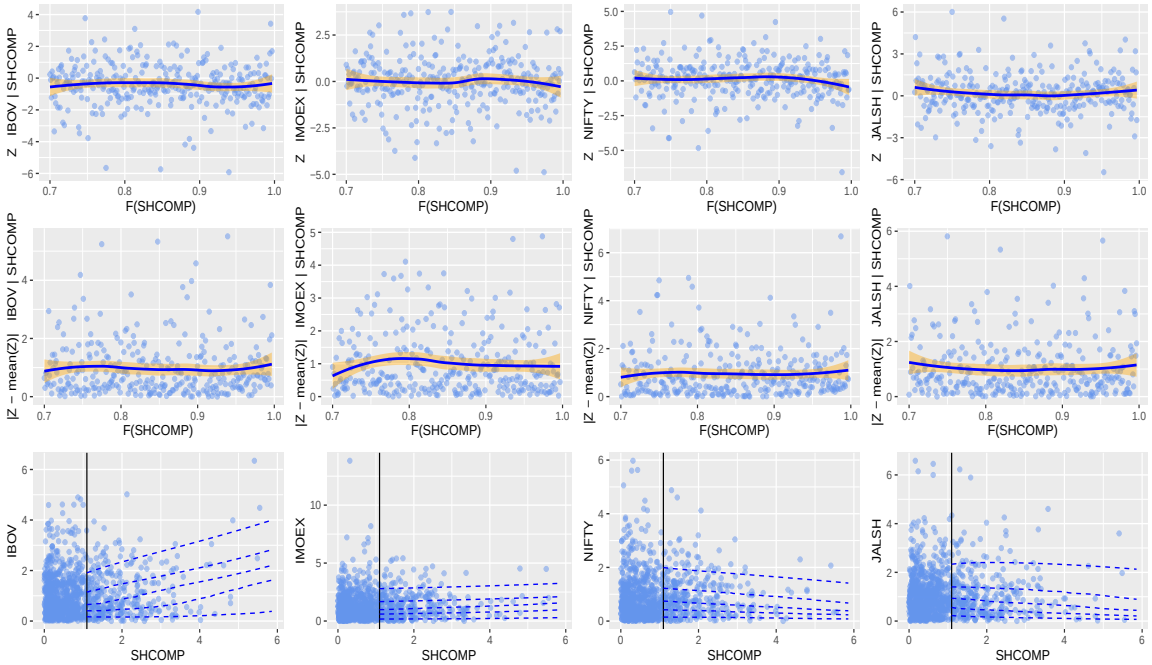


Figure 6.30: Dependence model diagnostics: conditioning on the SHCOMP variable.

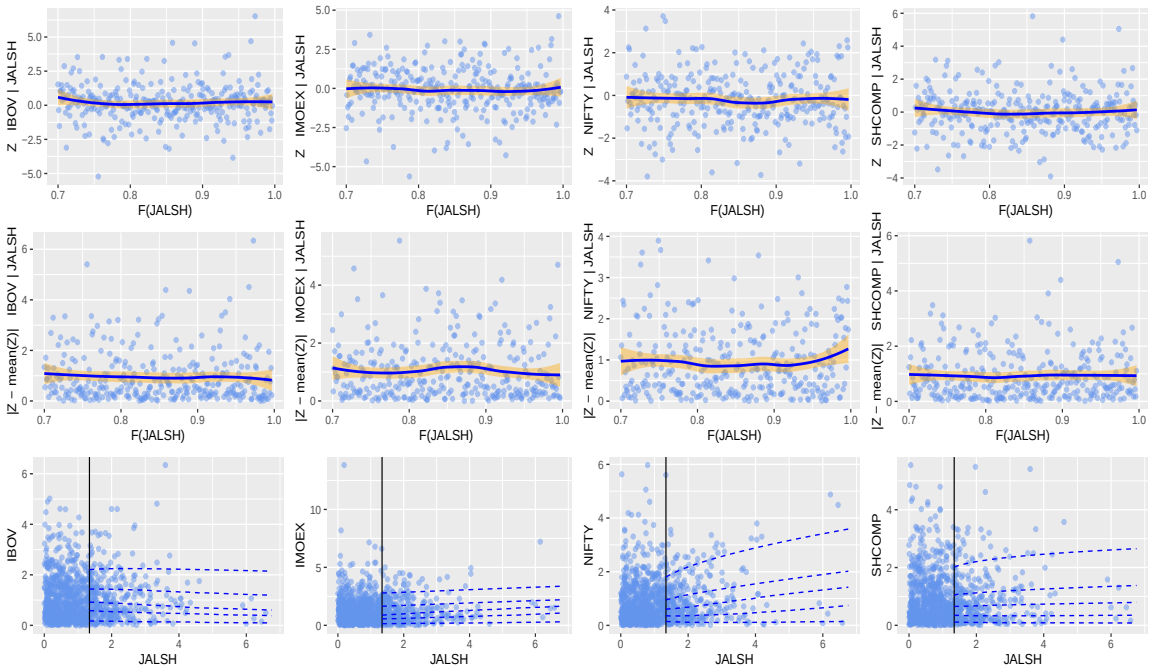


Figure 6.31: Dependence model diagnostics: conditioning on the JALSH variable.

the parameter estimates of the conditional multivariate extreme value model are most accurate where the lowess curves are smoothest at the 70th percentile for all the BRICS stock markets. That is, these diagnostic plots support the selection of the 70th percentile dependence quantile. Moreover, at this chosen quantile, the raw data distribution is approximately in good agreement with the fitted quantiles (i.e., the solid vertical line) and the scatter plots (the dotted lines) as displayed in layer 3 of each market's diagnostic plots. It is necessary to know that different threshold quantiles can eventually be used for the marginal and dependence modelling (Southworth et al., 2017), since the choice of a suitable threshold for each of the modelling is strictly based on the appropriateness of the diagnostic assessment plots. This is observed in the Russian IMOEX market where 80th percentile was obtained as the marginal quantile while 70th percentile was used as the dependence quantile.

6.3.7 CMEV model parameter estimates

Table 6.1 shows the estimated parameters of the BRICS stock markets' dependence structure conditioning on each of the five margins one after the other. That is, when the threshold excesses of one of the markets are given, the conditional distribution of the remaining four market variables are described. Here, we described the dependence between market pairs by a pair of parameters "A" and "B", with more attention focused on the former, such that the values of "A" near 1 or -1 indicate strong positive or negative extremal dependence (Southworth et al., 2017).

6.3.8 Extremal dependence results of the CMEV model

The dependence parameter estimates of the CMEV modelling generates the following results as shown in Table 6.1.

1. Conditioning on Brazilian IBOV market: From the table, it is clearly shown that the Russian IMOEX and Chinese SHCOMP markets have fairly strong positive extremal dependence on large values of the Brazilian IBOV market, with the Russian IMOEX market having stronger dependence than the Chinese SHCOMP market on the Brazilian IBOV market. The Indian NIFTY and South African JALSH markets on the other hand, have a very weak negative extremal dependence on the conditioning Brazilian IBOV market.

2. Conditioning on Russian IMOEX market: The Brazilian IBOV, Indian NIFTY, Chinese SHCOMP and South African JALSH markets have a relatively weak positive extremal dependence on the Russian IMOEX market, with the strongest of this weak dependence being between the Russian IMOEX and Brazilian IBOV markets.

3. Conditioning on Indian NIFTY market: Here, it is observed that the Brazilian IBOV and Russian IMOEX markets have varying levels of weak positive

Table 6.1: Dependence structure parameter estimates of the CMEV model.

	Dependence parameters	IMOEX	NIFTY	SHCOMP	JALSH
Conditioning on: IBOV	A	0.3888	-0.0194	0.3157	-0.0408
	B	0.1019	0.0257	0.2452	0.0633
	Dependence parameters	IBOV	NIFTY	SHCOMP	JALSH
Conditioning on: IMOEX	A	0.1655	0.0151	0.0314	0.0184
	B	0.1386	0.0669	-0.1024	0.1650
	Dependence parameters	IBOV	IMOEX	SHCOMP	JALSH
Conditioning on: NIFTY	A	0.1116	0.0059	-0.1110	0.2531
	B	0.0066	0.0140	-0.1518	0.2034
	Dependence parameters	IBOV	IMOEX	NIFTY	JALSH
Conditioning on: SHCOMP	A	0.3159	0.0717	-0.0957	-0.1235
	B	0.2081	-0.0055	-0.0352	0.1015
	Dependence parameters	IBOV	IMOEX	NIFTY	SHCOMP
Conditioning on: JALSH	A	-0.0606	0.1018	0.2011	0.0311
	B	0.0642	-0.0138	0.2315	0.1255

extremal dependencies on the Indian NIFTY market, while the asymptotic dependence between the Indian NIFTY and South African JALSH markets is moderately strong. The Chinese SHCOMP market however has a weak negative dependence on the Indian NIFTY market.

4. Conditioning on Chinese SHCOMP market: The values of this dependence parameter shows that the Brazilian IBOV is the most (fairly) strongly positively dependent on large values of the Chinese SHCOMP market, while the Russian IMOEX, Indian NIFTY and South African JALSH markets have only weak extremal dependence on the Chinese SHCOMP market. More specifically, the Indian NIFTY and South African JALSH markets have weak negative levels of dependence while the Russian IMOEX market has a weak positive dependence on the Chinese SHCOMP market.

5. Conditioning on South African JALSH market: The values of the dependence parameter estimates show that the Russian IMOEX, Indian NIFTY and Chinese SHCOMP markets all have different weak positive extremal dependencies on the South African JALSH market, the strongest of these is between the South African JALSH and Indian NIFTY markets. The Brazilian IBOV market has a weak negative extremal dependence on the South African JALSH market.

6.3.9 Prediction under the CMEV model

The fitted CMEV model can possibly be interpreted well through variables prediction given extreme values of a conditioning variable (Southworth et al., 2017). With the use of "importance sampling", prediction can be made by estimating quantiles or probabilities of threshold exceedances (as shown in Table 6.2) for the fitted CMEV model given the conditioning variable above the threshold for extrapolation (Southworth et al., 2020). Importance sampling are samples obtainable from important or

Table 6.2: Predicted conditional probability of threshold exceedance.

Conditioning on IBOV	IMOEX	NIFTY	SHCOMP	JALSH
	0.510	0.305	0.454	0.302
Conditioning on IMOEX	IBOV	NIFTY	SHCOMP	JALSH
	0.433	0.346	0.358	0.351
Conditioning on NIFTY	IBOV	IMOEX	SHCOMP	JALSH
	0.323	0.363	0.315	0.402
Conditioning on SHCOMP	IBOV	IMOEX	NIFTY	JALSH
	0.425	0.347	0.342	0.278
Conditioning on JALSH	IBOV	IMOEX	NIFTY	SHCOMP
	0.280	0.337	0.408	0.318

interesting region by sampling from a distribution overweighting the region (Owen, 2018). It is an estimator's variance reduction method, that is, it reduces statistical error such that an estimator can converge to the true value faster (Lu and Zhang, 2003).

The fundamental concept underpinning importance sampling methods is that certain of the input random variables' values have more meaningful influence than others on the estimated parameters, and if such meaningful values are sampled more regularly than others, the estimator's variance can be reduced (Lu and Zhang, 2003). Hence, the objective of the method is to get precise estimate of quantities, like the mean and variance, with fewer samples than needed (Lu and Zhang, 2003).

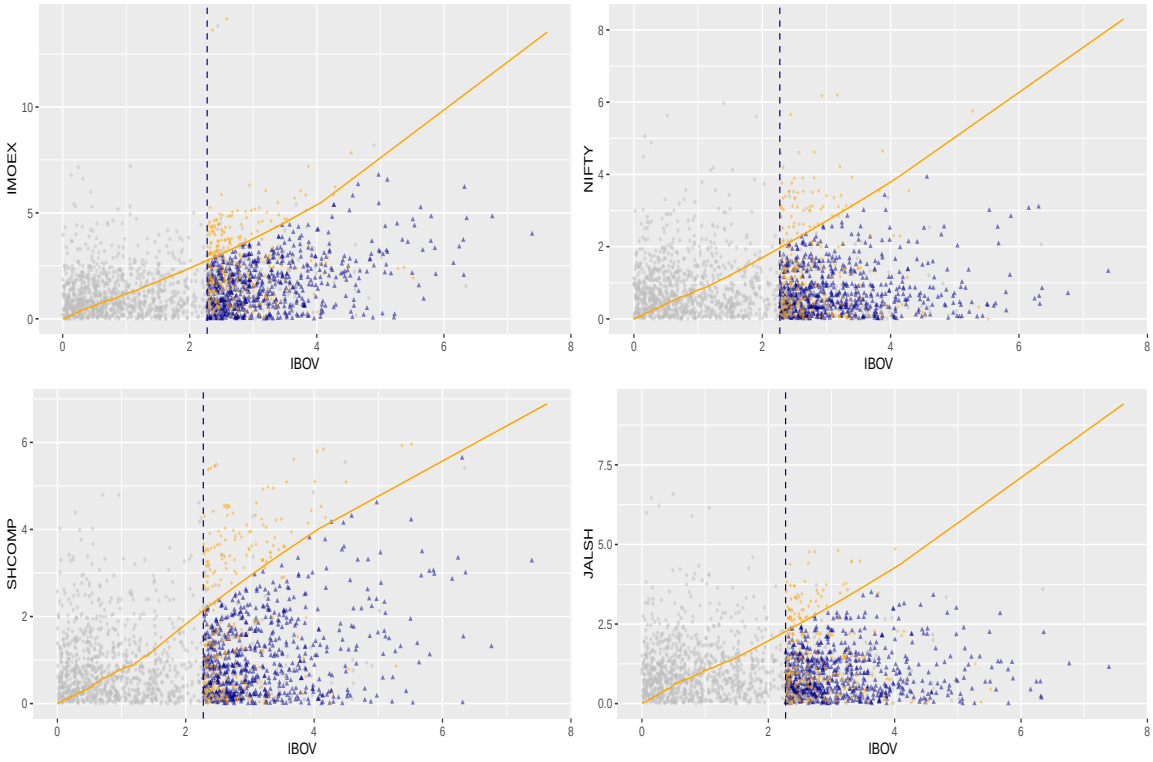


Figure 6.32: Prediction plot conditioning on IBOV being above its 90th percentile.

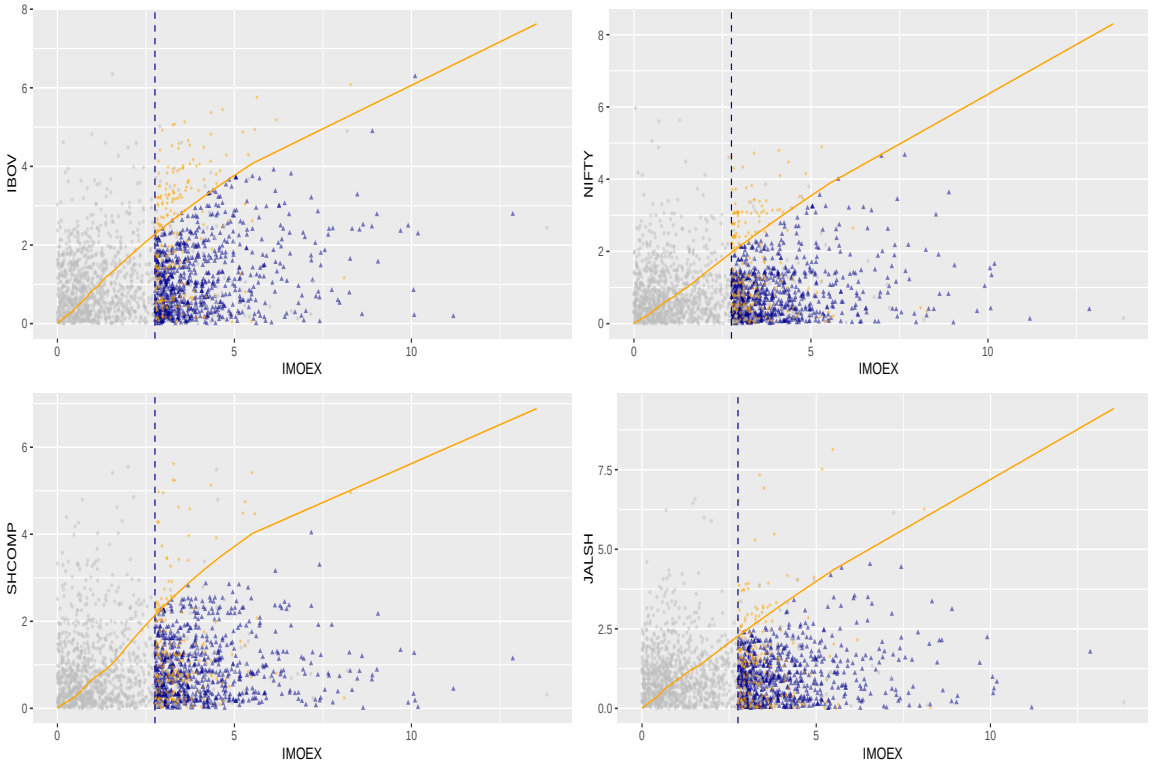


Figure 6.33: Prediction plot conditioning on IMOEX being above its 90th percentile.

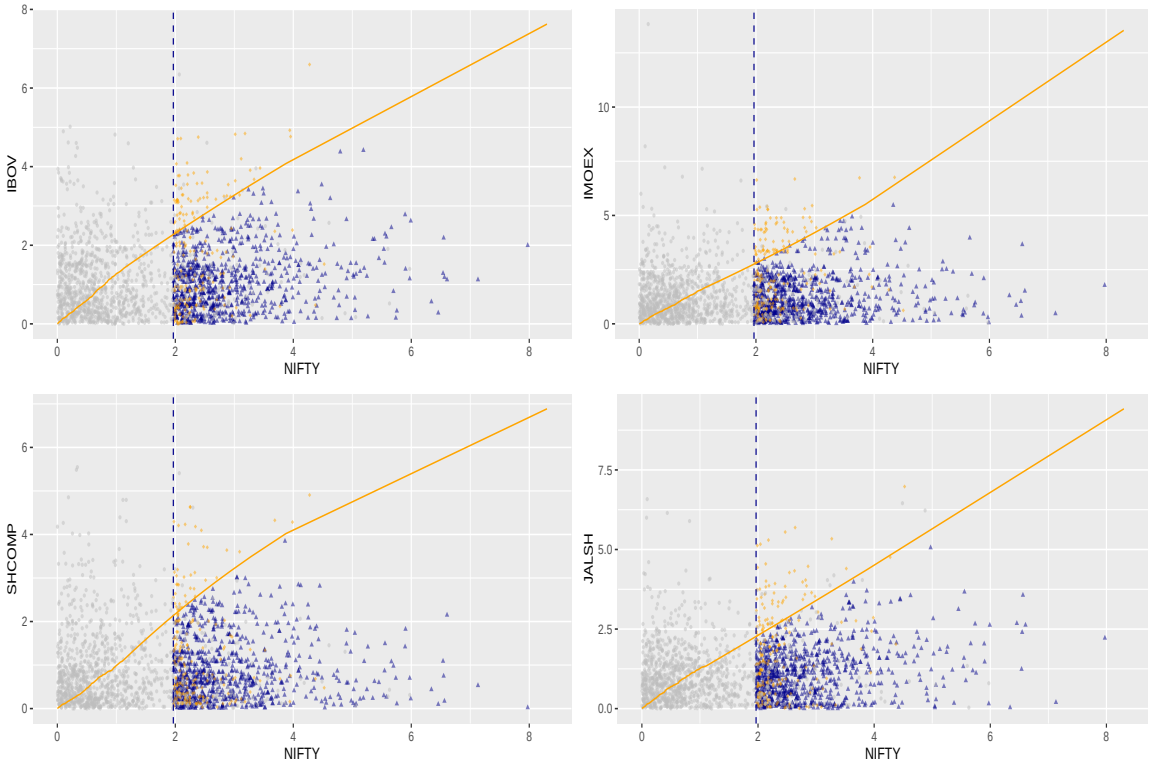


Figure 6.34: Prediction plot conditioning on NIFTY being above its 90th percentile.

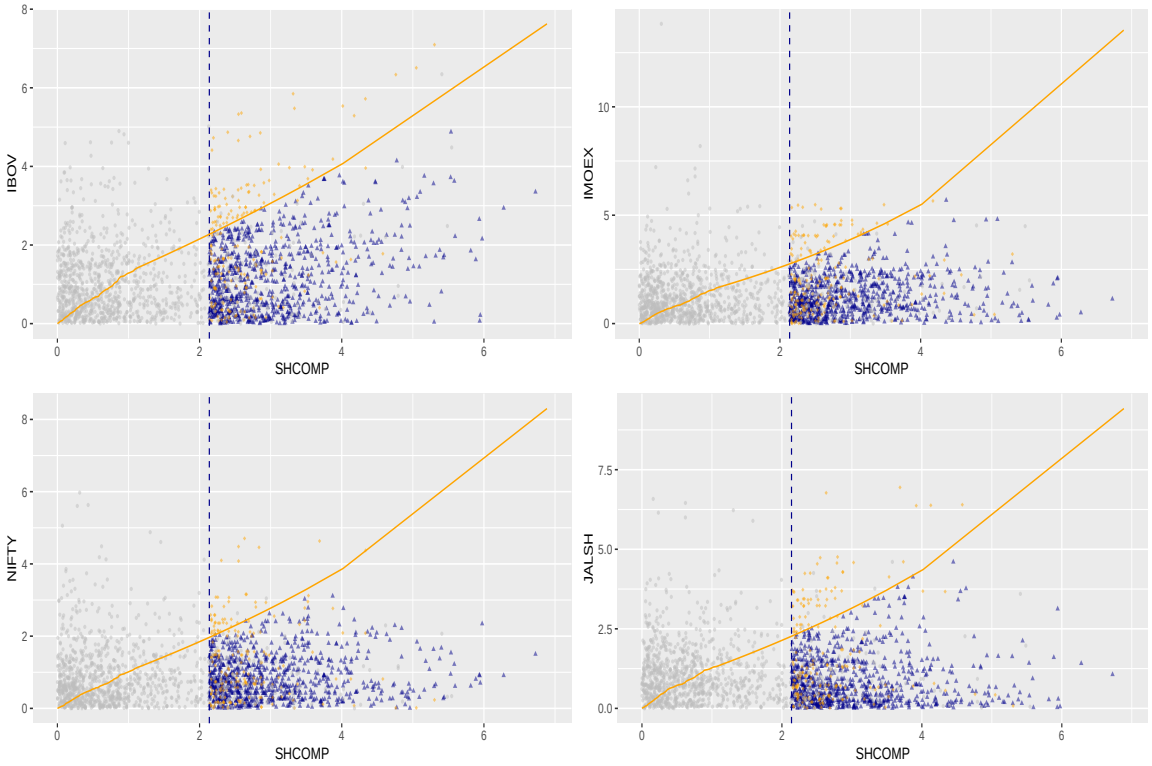


Figure 6.35: Prediction plot conditioning on SHCOMP being above its 90th percentile.

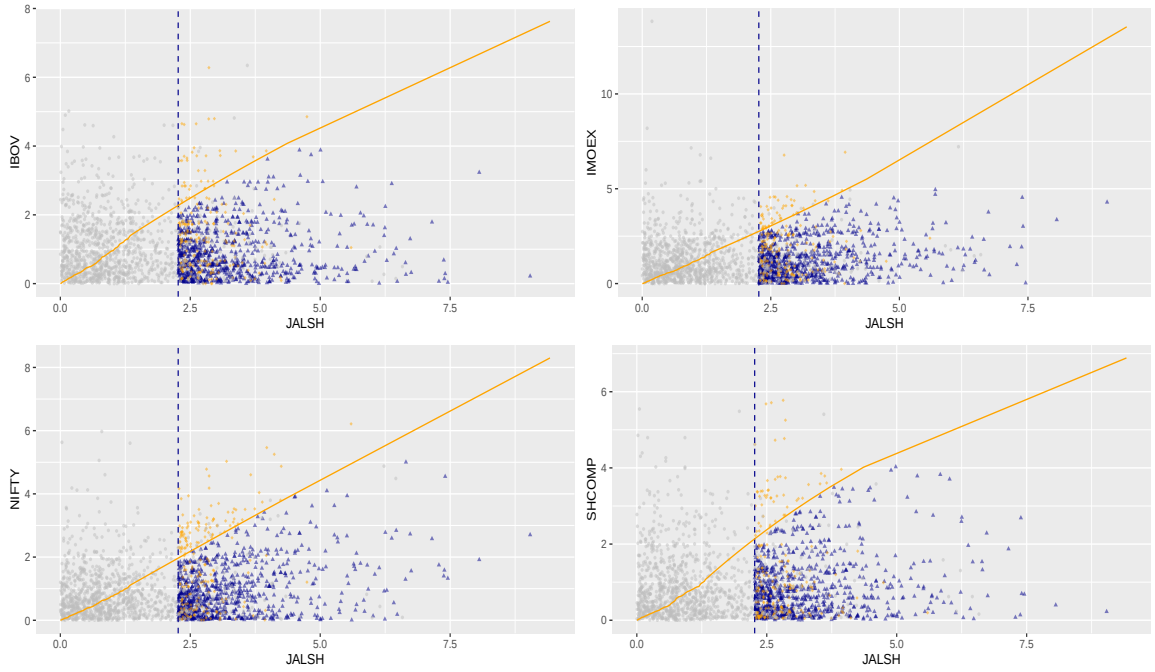


Figure 6.36: Prediction plot conditioning on JALSH being above its 90th percentile.

To begin with, the predictions under the CMEV model fit are made by importance sampling through simulating the values of the other four (remaining) market-variables given the conditioning variable being above a large prediction quantile (see Southworth et al., 2017). In order to choose a suitable prediction quantile, different candidate quantiles of 80th, 90th, 95th and 99th percentiles were examined, and 90th percentile was found to be the most appropriate considering bias and variance trade-off. That is, the highest 10% of the conditioning variable's values are being used. It should be noted that importance samples are known to have few values in the conditional tails (Southworth et al., 2020), hence the higher the prediction quantiles, the fewer the values in the conditional tails. This reason also makes the 90th percentile a more suitable choice in preference to the higher variance 95th and 99th percentiles. As for the other four market-variables (excluding the conditioning variable), any value of threshold quantile can be used for the prediction (Southworth et al., 2017), and we

used 0.7 quantiles or 70th percentiles since the diagnostic plots of the lowess curves are smoothest for the dependence modelling at these quantiles.

Having obtained the required prediction quantile, values of the market-variables are simulated conditional on each of the five variables (in turn) being above its 90th percentile. For instance, with the Brazilian IBOV market, values of the Russian IMOEX, Indian NIFTY, Chinese SHCOMP and South African JALSH market-variables are simulated conditional on the Brazilian IBOV variable being above its 90th percentile. Table 6.2 shows the outcomes of the conditional distributions, i.e. the predicted (estimated) quantiles or probabilities of threshold exceedances.

The prediction plots in Figures 6.32 to 6.36 are used for a visual display of the CMEV model fit and its prediction using importance samples. The plots show grey circles that denote the original data, and data importance samples, represented by the blue triangles and orange diamonds, under the fitted CMEV model above the threshold for forecasting. The orange solid curves or lines in the plots are for references and each curve joins equal quantiles of the BRICS marginal distributions. The markets' paired variables with a perfect dependence will lie precisely on this line, where the line is comparable to a QQ plot's diagonal line. However, the curve is not a straight line because the two paired margins are not equal. From Figure 6.32 for instance (by conditioning on the Brazilian IBOV market), it can be observed, when evaluated on a common quantile scale, that all the blue triangles and orange diamonds are large in the (conditioning) IBOV variable, but only those displayed by the blue triangles are largest in IBOV variable, while the orange diamonds are largest in each of the remaining four market-variables. This visual scenario of the conditioning IBOV market-variable in Figure 6.32 is also experienced by the remaining four conditioning markets, i.e., the IMOEX, NIFTY, SHCOMP and JALSH market-variables in Figures 6.33, 6.34, 6.35 and 6.36 respectively.

Now for the prediction, we begin with conditioning on the Brazilian IBOV market

in Figure 6.32 and observed that contrary to the outcomes of the dependence structure in Table 6.1 where the extremal dependence between market pair Brazilian IBOV and Russian IMOEX is the strongest, followed by the dependence between markets Brazilian IBOV and Chinese SHCOMP, the future relationship as shown in the figure is predicted to be approximately strongest between market pair Brazilian IBOV and Indian NIFTY, followed by the pair of Brazilian IBOV and South African JALSH markets. This predicted extremal dependence within each pair of these latter markets (i.e. the Brazilian IBOV and Indian NIFTY, and Brazilian IBOV and South African JALSH) is better than it is in the two market-pairs Brazilian IBOV and Russian IMOEX, and Brazilian IBOV and Chinese SHCOMP in terms of near-linearity into the future as displayed in the figure.

Next, by conditioning on the Russian IMOEX market in Figure 6.33, it is also observed that as opposed to the results of the dependence structure in Table 6.1, where the extremal dependence between market pair Russian IMOEX and Brazilian IBOV is the strongest, followed by the dependence between markets Russian IMOEX and Chinese SHCOMP, the future predicted relationship is strongest between Russian IMOEX and South African JALSH markets, followed by the dependence between pair Russian IMOEX and Indian NIFTY markets as displayed in the figure due to the sampled data that follow closely the curves of equal marginal quantiles.

By conditioning on the Indian NIFTY market in Figure 6.34, the evidence of strongest extremal dependence between market pair Indian NIFTY and South African JALSH, followed by the pair of Indian NIFTY and Brazilian IBOV markets as shown in Table 6.1 is clearly extended into future dependence predictions as shown in the figure. That is, from the figure, the market pair Indian NIFTY and South African JALSH has the strongest predicted strength of direct proportionality (followed by the Indian NIFTY and Brazilian IBOV pairwise combination) than the rest of the paired markets.

Next, by conditioning on the Chinese SHCOMP market in Figure 6.35, it is also observed that the strongest extremal dependence between Chinese SHCOMP and Brazilian IBOV markets as shown by the result of the dependence structure in Table 6.1 is carried into future extremal dependence as shown by the near-linearity of their diagonal line in the prediction plot. Lastly, when conditioning on the South African JALSH market in Figure 6.36, the pattern of the strongest extremal dependence between the pair of South African JALSH and Indian NIFTY markets, followed by the market pair South African JALSH and Russian IMOEX as shown in Table 6.1 is further projected into the future as displayed in the figure. That is, predictions based on the levels of future relationship (extremal dependence) are greatest in these pairs of markets, with the latter pair following the former.

6.4 Bivariate point process modelling

The bivariate point process is now used for the extremal dependence modelling via the six parametric models that include the logistic, bilogistic, Husler-Reiss, negative logistic, negative bilogistic and the Dirichlet (or Coles-Tawn, abbreviated "ct") dependence models as described in Section 3.24.1. The model that gives the best fit for the dependence structure in each of the ten paired markets is selected. Decision relating to model selection (or comparison) for dependence structure of nested models can be addressed using standard likelihood ratio tests (Coles and Tawn, 1991; Lipika, 2018) and analysis of variance (anova) (Stephenson, 2018), whereas for non-nested models, analytic goodness of fit statistic like the Akaike information criterion (AIC) can be used (Coles and Tawn, 1991; Lipika 2018; Stephenson, 2018). Hence, the AIC as described in Section 3.10.2 and in the applied package "evd" (Stephenson, 2018) is used for the model selection in each pair of the markets.

Tables 6.3, 6.4 and 6.5 show the dependence parameters α and β along with their standard error likelihood estimates in parentheses. The logistic, negative logistic

and Husler-Reiss models have a single dependence parameter α, therefore the β is tabulated as "Nil" in the tables. The model with the lowest AIC value denotes the best fitting model for the dependence estimate of each pair of the BRICS markets.

6.4.1 The findings: Bivariate point process

There are ten pairwise combinations of the five BRICS markets as presented in Tables 6.3, 6.4 and 6.5. The results from the tables show that the model that best describes all the ten paired markets is the Husler-Reiss, with the lowest AIC value in each pair. As described in Section 3.24.1, the Husler-Reiss model produces complete dependence between variables as α tends to ∞, while independence is obtained as α approaches 0. The tables show that maximization of likelihood under the Husler-Reiss model for the ten paired markets yields estimates of $1.2883 \leq \hat{\alpha} \leq 1.3413$, and corresponding standard errors of between 0.0320 and 0.0345 inclusive.

These dependence estimates $\hat{\alpha}$ tend to move further away from 0, and are therefore significantly different from independence, but reasonably correspond to weak levels of dependence. However, since the larger (or smaller) the likelihood estimate of α, the stronger (or weaker) the strength of dependence (see Coles 2001), the dependence values of 1.3413 and 1.3200 for the pairs of Brazilian IBOV and Russian IMOEX, and Indian NIFTY and South African JALSH markets respectively as shown in Tables 6.3 and 6.5 can be classified as approximately fairly strong. These entire findings are consistent with the results obtained from the CMEV modelling. The only likely exception to the consistency is between the pair of Brazilian IBOV and Chinese SHCOMP markets which has a fairly strong dependence under the CMEV modelling, but produced a nearly weak dependence under the point process.

Table 6.3: Estimates of the point process dependence modelling.

Brazilian IBOV and Russian IMOEX			
Parametric model	α	β	AIC
Logistic	0.5994 (0.0117)	Nil	3762.38
Negative logistic	0.9055 (0.0314)	Nil	3711.77
Husler-Reiss	1.3413 (0.0345)	Nil	3669.11
Bilogistic	0.5829 (0.0258)	0.6155 (0.0247)	3763.84
Negative bilogistic	1.1938 (0.1134)	1.0210 (0.0978)	3712.99
ct (or Dirichlet)	0.8195 (0.1010)	1.0172 (0.1369)	3718.47
Brazilian IBOV and Indian NIFTY			
Parametric model	α	β	AIC
Logistic	0.6066 (0.0114)	Nil	3756.67
Negative logistic	0.8854 (0.0300)	Nil	3701.53
Husler-Reiss	1.3174 (0.0330)	Nil	3656.31
Bilogistic	0.6209 (0.0244)	0.5909 (0.0267)	3758.23
Negative bilogistic	1.0470 (0.1029)	1.2140 (0.1141)	3702.85
ct (or Dirichlet)	0.9829 (0.1364)	0.7940 (0.0959)	3710.40
Brazilian IBOV and Chinese SHCOMP			
Parametric model	α	β	AIC
Logistic	0.6086 (0.0114)	Nil	3788.63
Negative logistic	0.8779 (0.0296)	Nil	3733.10
Husler-Reiss	1.3100 (0.0326)	Nil	3683.98
Bilogistic	0.6264 (0.0235)	0.5901 (0.0250)	3789.91
Negative bilogistic	1.0629 (0.0988)	1.2186 (0.1111)	3734.46
ct (or Dirichlet)	0.9515 (0.1229)	0.7931 (0.0949)	3742.99
Brazilian IBOV and S/African JALSH			
Parametric model	α	β	AIC
Logistic	0.6132 (0.0114)	Nil	3817.37
Negative logistic	0.8646 (0.0292)	Nil	3761.19
Husler-Reiss	1.2906 (0.0321)	Nil	3715.79
Bilogistic	0.6137 (0.0246)	0.6125 (0.0254)	3819.37
Negative bilogistic	1.1414 (0.1098)	1.1707 (0.1096)	3763.17
ct (or Dirichlet)	0.8591 (0.1121)	0.8303 (0.1027)	3772.86

Table 6.4: Estimates of the point process dependence modelling.

Russian IMOEX and Indian NIFTY			
Parametric model	α	β	AIC
Logistic	0.6073 (0.0116)	Nil	3873.50
Negative logistic	0.8818 (0.0302)	Nil	3821.21
Husler-Reiss	1.3110 (0.0332)	Nil	3780.47
Bilogistic	0.6185 (0.0240)	0.5955 (0.0254)	3875.22
Negative bilogistic	1.0850 (0.1033)	1.1839 (0.1100)	3822.96
ct (or Dirichlet)	0.9256 (0.1208)	0.8203 (0.0997)	3830.56
Russian IMOEX and Chinese SHCOMP			
Parametric model	α	β	AIC
Logistic	0.6072 (0.0116)	Nil	3941.75
Negative logistic	0.8806 (0.0303)	Nil	3887.51
Husler-Reiss	1.3145 (0.0334)	Nil	3838.94
Bilogistic	0.6140 (0.0244)	0.6003 (0.0252)	3943.65
Negative bilogistic	1.1083 (0.1068)	1.1628 (0.1106)	3889.43
ct (or Dirichlet)	0.9066 (0.1206)	0.8415 (0.1072)	3897.57
Russian IMOEX and S/African JALSH			
Parametric model	α	β	AIC
Logistic	0.6126 (0.0114)	Nil	3935.83
Negative logistic	0.8653 (0.0293)	Nil	3883.64
Husler-Reiss	1.2908 (0.0321)	Nil	3839.97
Bilogistic	0.6068 (0.0232)	0.6184 (0.0231)	3937.75
Negative bilogistic	1.1356 (0.1036)	1.1759 (0.1065)	3885.60
ct (or Dirichlet)	0.8727 (0.1101)	0.8151 (0.0967)	3894.70
Indian NIFTY and Chinese SHCOMP			
Parametric model	α	β	AIC
Logistic	0.6063 (0.0117)	Nil	3805.05
Negative logistic	0.8826 (0.0305)	Nil	3758.71
Husler-Reiss	1.3070 (0.0332)	Nil	3719.32
Bilogistic	0.6142 (0.0231)	0.5980 (0.0245)	3806.90
Negative bilogistic	1.1027 (0.1029)	1.1632 (0.1055)	3760.61
ct (or Dirichlet)	0.8996 (0.1146)	0.8365 (0.1005)	3767.56

Table 6.5: Estimates of the point process dependence modelling.

Indian NIFTY and S/African JALSH			
Parametric model	α	β	AIC
Logistic	0.6040 (0.0115)	Nil	3745.97
Negative logistic	0.8924 (0.0304)	Nil	3699.66
Husler-Reiss	1.3200 (0.0332)	Nil	3664.18
Bilogistic	0.6054 (0.0229)	0.6026 (0.0229)	3747.97
Negative bilogistic	1.1070 (0.0979)	1.1343 (0.1005)	3701.64
ct (or Dirichlet)	0.8946 (0.1064)	0.8649 (0.1018)	3707.80
Chinese SHCOMP and S/African JALSH			
Parametric model	α	β	AIC
Logistic	0.6149 (0.0114)	Nil	3869.14
Negative logistic	0.8605 (0.0290)	Nil	3810.94
Husler-Reiss	1.2883 (0.0320)	Nil	3764.18
Bilogistic	0.6283 (0.0242)	0.6006 (0.0260)	3870.76
Negative bilogistic	1.1129 (0.1076)	1.2121 (0.1143)	3812.71
ct (or Dirichlet)	0.8926 (0.1183)	0.7915 (0.0977)	3823.01

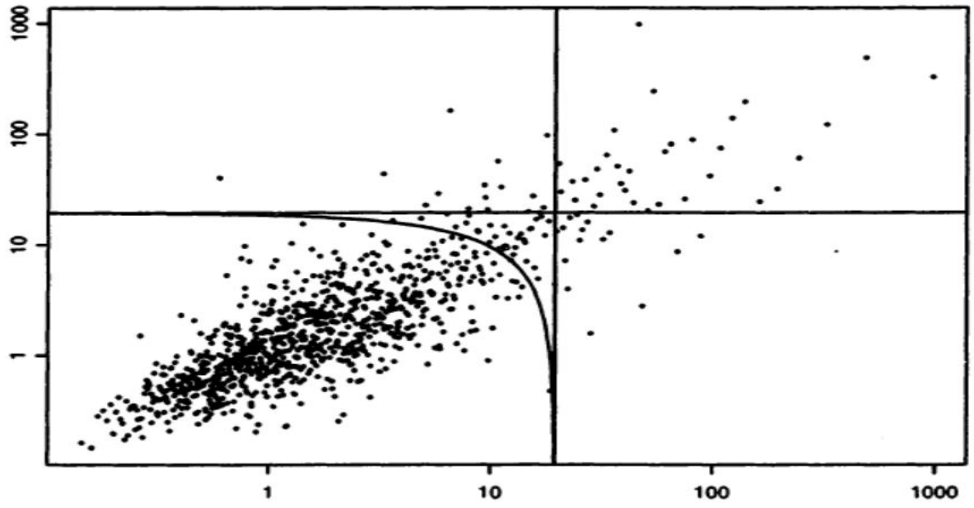


Figure 6.37: The CMEV and bivariate point process models' same points thresholds.

6.4.2 Point process and CMEV models compared

This section compares the outcomes and modelling approaches of the bivariate point process and the CMEV models. An adequately objective comparison of the dependence approach of the point process with the threshold excess method of the CMEV model is made by using the dependence threshold u that corresponds to the 70% quantile that was used for the CMEV modelling. This is made possible when a threshold u is selected where its intersection with the axes takes place at the same points for both models, as exemplified in Figure 6.37 (see Coles, 2001). At these points (in this study), the thresholds for the point process and threshold excess approach of the CMEV models are selected to intersect the axes at the marginal 70% quantiles.

The difference between the two models is due to the different regions where their approximations assume validity. From Figure 6.37, the limit of the point process model is assumed to be valid in the entire region above the curve, to the right of the boundary, i.e., the boundary of the threshold and the axes. More precisely, the point

Table 6.6: Extremal dependence structures of the CMEV and point process models.

Panel A		
Paired markets	**CMEV model**	**Bivariate point process model**
Brazilian IBOV and Indian NIFTY	Weak	Weak
Brazilian IBOV and S/African JALSH	Weak	Weak
Russian IMOEX and Indian NIFTY	Weak	Weak
Russian IMOEX and Chinese SHCOMP	Weak	Weak
Russian IMOEX and S/African JALSH	Weak	Weak
Indian NIFTY and Chinese SHCOMP	Weak	Weak
Chinese SHCOMP and S/African JALSH	Weak	Weak
Panel B		
Brazilian IBOV and Russian IMOEX	Fairly strong	Fairly strong
Indian NIFTY and S/African JALSH	Fairly strong	Fairly strong
Panel C		
Brazilian IBOV and Chinese SHCOMP	Fairly strong	Nearly weak

process can use exceedance observations above the curve in the two margins together and in the joint upper (right) quadrant. This upward region of the boundary or curve is considered appropriately extreme for the limit outcome of the point process to offer a valid approximation. As opposed to this, given the same marginal threshold as the point process, the limit of the threshold excess method of the CMEV model takes accuracy only in the joint upper (right) quadrant (Coles, 2001).

Hence, since every extreme observation above the curved region is used, the point process is able to model many more points or exceedances that contribute to the likelihood estimation and it gives more information than the threshold excess method of the CMEV model. Table 6.6 summarily presents the extremal dependence structures of the ten paired markets under the two models: the CMEV and bivariate point process models. It can be observed as shown in panel A of the table that asymptotic dependence is weak but significant for seven out of the ten pairs of the markets under both models. Next, panel B displays a fairly strong dependence outcome for two paired markets under the two comparative models. In panel C however, the two models, as discussed earlier, yielded a slightly different outcome, with the CMEV model resulting in a fairly strong dependence, while the point process produced a nearly weak dependence for the same paired markets. Hence, there is consistency in the modelling capabilities of the two models for nine out of the ten paired markets, and near consistency in the tenth market pair.

6.4.3 Extremal dependence impacts on investment and portfolio diversification

One barrier on the route to the creation of wealth is the risk of extreme losses, and any portfolio that seeks wealth maximisation should mitigate this risk. Risk is basically defined as the instability or unpredictability of returns from an investment. Intensely advanced risk management systems and rigorous research have been carried out over

time to comprehend and tackle this risk. The most robust outcome from the works so far is rather straightforward: diversify.

Creating a well-diversified portfolio within an asset class and across classes of assets, and also geographically by investing domestically and in foreign markets can help investors to deal with returns variability, thereby forestalling extreme losses. Diversification is a risk management approach where a broad variety of investments are mixed within a portfolio. Although portfolio diversification may limit investment gains in the short-term, but in the long-term, it decreases portfolio risk, hedges against volatility in markets, and gives higher returns (Segal, 2020).

Diversification is the fundamental for creating portfolios that maximise return for a certain risk level or on the contrary, minimise risk for a certain return level. Reisen (2000) in his findings revealed that risk is easier to reduce via international diversification than through domestic diversification. Chen (2020) further showed that having an international portfolio can be used for risk reduction in investment. This is because if equities underperform in one nation's domestic market, gain in the international holdings for the investor in any other nation(s) can smooth out the return. Hence, through international investment, an investor is able to raise return by way of risk reduction.

Amongst others, Solnik (1974) indicated that international (portfolio) diversification enables investors to achieve greater efficient edge when compared to diversifying domestically (Aloui et al., 2011). Moreover, because of different structures of industry in different nations and since different economies do not precisely adhere to similar business cycle, diversifying across countries whose economic (or trade) cycles do not have perfect correlation can typically reduce returns variability for investors and portfolio managers (Solnik, 1974; Zonouzi et al., 2014). This is so because different factors drive each nation's market at any given time. More directly, the benefits of diversification are obtained from risk reduction in nearly uncorrelated and negatively

correlated markets (Odit et al., 2011). This, in other words, would mean the existence of low correlations of stock returns between different economies (Solnik, 1974).

Yavas (2007) addressed the subject of risk reduction via international diversification by revealing that diversification across nations within an industry can produce much more effective risk reduction than industry diversification within a nation. In essence, if an entire industry fails in one nation but succeeds in another, investing in the same industry in both nations may well hedge the risk. Numerous possible benefits like the mentioned risk reduction and increase in returns have led to investors internationalizing their portfolios. These apparent benefits are seen by Bartram and Dufey (2001) as essential motivations needed for international portfolio investments.

Results in Tables 6.1 and 6.3 – 6.6 describe the weak (or low) positive and negative extremal dependence associations and the fairly strong (but still low) relationship exhibited by the pairs of the BRICS markets. All these outcomes from the ten pairwise combinations of the BRICS stock markets signify varied levels of low extremal dependence. With low correlation or simply weak asymptotic dependence and negative extremal dependence being the fundamental requirements for efficient portfolio diversification, investors and all market participants can seize the rich investment opportunities presented by these markets as presented in the tables. The weak extremal dependence indicates that extreme losses in one market do not easily spill-over to the other market. On the other hand, the negative asymptotic dependence (for instance, in the four paired markets under the CMEV modelling in Table 6.1) means that underperformance or negative outcome caused by extreme losses in one market can be offset or hedged by good (positive) returns performance in the other market for a paired market.

To begin with, the findings as shown in panel A of the summary Table 6.6 show weak extremal (asymptotic) dependence between each of the seven (out of ten) paired

markets. That is, from the table, beneficial risk reduction and high investment returns through international portfolio diversifications can be derived from any of the following as indicated in panel A: 1) if a Brazilian investor invests in the domestic market and any of the Indian and South African markets; 2) if a Russian investor jointly invests in the domestic market and any of the Indian, Chinese, and South African market; 3) if an Indian investor invests in the domestic market and in the Chinese market; 4) if a Chinese investor invests in the domestic market and in the South African market. Moreover, an authorised international investor outside any of these countries can likewise invest comfortably and obtain good returns from any of the markets' combinations.

Next, in panel B (that contains two paired markets), a fairly good investment opportunity derivable from international portfolio diversifications can also be expected. This is so because the extremal dependence between the markets in these market pairs is "fairly strong" as compared to the "weak asymptotic" dependence for panel A markets, hence the diversification benefits are meant to be lower for these two paired markets when compared to the seven paired markets in panel A. Following these outcomes in panel B, 1) a Brazilian investor can earn a fairly good returns by diversifying in the domestic market and in the Russian stock market, and 2) an Indian investor can also obtain beneficially fair portfolio diversification and investment returns jointly from the domestic market and the South African stock market. This opportunity is also open to international investors who are interested in portfolio investments in these market pairs.

The findings in panel C for the tenth paired markets are not outrightly direct since the two models (CMEV and bivariate point process) produced slightly different outcomes. However, from the table, these two outcomes (fairly strong and nearly weak extremal dependencies) are still categorised approximately under low asymptotic dependence for any investor with investment interest in these markets. Hence, from

panel C, a Brazilian investor can derive a beneficial or fairly beneficial investment return outcome from diversifying in the domestic market and in the Chinese stock market.

It can be observed from the findings in Tables 6.1, 6.3 – 6.6 that investors will possibly achieve the most desirable portfolio diversification benefits in the Chinese−South African combination with the lowest extremal dependence and the least desirable diversified portfolio investment in the Brazilian−Russian combination with the highest asymptotic dependence. The summary extremal dependence results in panels A, B and C of Table 6.6 are highly beneficial to investors, traders, portfolio managers and other market participants who are interested in maximising their investment returns and financial gains, and in the process mitigate possible investment downturns through international portfolio diversification in the BRICS stock markets.

It is also important to know that in these days, unlike in the past, investors can effectively build internationally diversified portfolios through mutual funds and international exchange-traded funds (ETFs) which focus on handling foreign equities, with quite reasonable and quick way to diversify (Yavas, 2007; Driessen and Laeven, 2007; Kuepper, 2019; Chen, 2020).

Chapter 7

Discussion of key findings and concluding remarks

7.1 Introduction

In this thesis, we modelled and analytically presented the dynamics of financial returns volatility, risks and extremal dependence of the BRICS stock markets, using various robust statistical models and tests via the algorithms of different packages in *R* software. This chapter conclusively explains the study's modelling outcomes and their relevance to investors, academics, policy makers, portfolio managers and other market participants for profitable investments and relevant decision making.

7.2 Summary of modelling by chapters

- In Chapter 4, this study extends the literature by modelling and statistically analysing the BRICS stock markets' volatility using GARCH (1, 1) model under the assumptions of seven error distributions of the normal, skewed-normal, Student's t, skewed-Student's t, GED, skewed-GED and the generalised hyperbolic (GHYP). The study further provides analytical details on the best error distribution that describes each of the markets.

- In Chapter 5, we modelled the equity risks of the BRICS stock markets using the univariate versions of EVT models via the conditional univariate extreme value's GPD and point process.

- In Chapter 6, the extremal dependence structures of each of the ten pairs of the BRICS stock markets were modelled using multivariate extreme value theory via the CMEV and bivariate point process models. The fitting efficacies of these two models were compared to ascertain the one that gave a better fit. The chapter also includes the predictions of future extremal dependence for each pair of the BRICS markets.

7.3 Modelling discussions and summary of key findings

7.3.1 Volatility modelling

From various proposed candidate models, ARMA(1, 1) and GARCH (1, 1) models were jointly selected as the most adequate to remove linear dependency and heteroscedasticity in the return series of the Brazilian, Russian, Indian, Chinese and South African markets. The fitting of this parsimonious ARMA-GARCH model was carried out under the seven stated error distributions.

It is concluded from the joint ARMA(1, 1)-GARCH(1, 1) model, that the Brazilian Bovespa and Russian IMOEX markets can both be well characterised (or described) by a heavy-tailed Student's t distribution, while the Indian NIFTY market is best characterised by the generalised hyperbolic (GHYP) distribution. Also, the Chinese SHCOMP and South African JALSH markets are best described by the skew-GED and skew-Student's t distributions, respectively.

These outcomes are consistent with the findings of Romero and Kasibhatla (2013) who studied the volatility dynamics of the BRIC (excluding South Africa) from 2000 to 2010 using GARCH models under the assumptions of three distributions of a normal, Student's t, and GED. That is, similar to our findings, the authors also concluded that the Student's t error distribution was the best to describe the volatility of the Brazilian and Russian stock markets. However, since we used a wider scope of error distributions, i.e., we used seven error distributions (including the skewed versions and GHYP) as compared to their only three distributions, our study shows skew-GED and GHYP as the best error distributions to describe the Chinese and Indian markets contrary to their choice of GED and Student's t respectively.

Volatility modelling in financial markets is very vital for risk management, portfolio administration and options trading, hence this study is motivated by the need to model the volatility of the BRICS return series in order to ascertain the level of potential risk associated with each of the markets for investors' decision making. Findings from this study show that under each of the best describing error distributions of the Student's t, skew-Student's t, GED, skew-GED and GHYP, the values of volatility persistence is close to one, indicating a near integration process. However, it is observed that the persistence of volatility in the BRICS markets does not follow the same hierarchical pattern under these error distributions, except under the skew-Student's t and GHYP distributions where the pattern is the same. Under these two assumptions, i.e., the skew-Student's t and GHYP, in a descending hierarchical order of magnitudes, volatility with persistence is highest in the Chinese market, followed by the South African market, then the Russian, Indian, and Brazilian markets, respectively.

For the Student's t distribution, the Chinese market has the highest volatility persistence, followed by the Russian, South African, Indian, and Brazilian markets in that order. For the GED, the pattern of volatility persistence in a descending hierarchy

is the Chinese, Indian, South African, Russian, and Brazilian markets, respectively. Lastly, for the skew-GED, the descending hierarchical pattern is the Chinese, South African, Indian, Russian, and Brazilian markets, respectively. However, under each of these error distributions, the Chinese market is the most volatile, while the Brazilian market is the least volatile.

This result is consistent with the findings of Ijumba (2013) who also modelled the BRICS markets' volatility, and observed the presence of volatility persistence in all the BRICS markets' stock returns using VAR and GARCH(1, 1) models. The study by the author also showed that Chinese market is the most volatile among the bloc but his conclusion on South African market being the least volatile market contradicts our findings of Brazil as the least volatile. The contradiction may be attributed to the scope and frequency of the data used for the two studies. The author (Ijumba, 2013) used sampled weekly data from 2000 to 2012, while this present study used daily closing data from 2010 to 2018, and it should be noted that South Africa only joined the BRICS bloc in 2010. Hence, the author only had access to two years weekly data for the South African JSE whereas this study utilised eight years of daily closing South African JALSH data with wider coverage and much more market activities.

Based on the above findings, the main potential contribution of this work to the growing studies on the volatility of the BRICS stock returns is twofold. First, this study gives new insights into the literature by providing a more robust explanation in more depth to the modelling and analysis of volatility of the BRICS stock returns by using the specified collective seven error distributions which as far as we know has not been done before this study. Second, as opposed to a major study carried out by Ijumba (2013) for the entire five BRICS markets using weekly data that covered only 2 years of markets' activities from 2010 to 2012, this study uses a wider coverage of activities from 2010 to 2018 using daily closing data, which can potentially give more accurate results.

7.3.2 Risk modelling via univariate models

Effective modelling of extreme financial losses is a key investment strategy required by investors and portfolio managers for successful assessment of risk in any financial market. This study compares the modelling capabilities of two EVT models via the conditional extreme value's (CEV's) GPD and point process for risk management and risk forecasting in the BRICS equity markets. The findings reveal that under the GPD model, the risks in the five BRICS equity markets can all be modelled by the Gumbel class of distributions. Under the point process approach however, the risk in the Russian equity market can be modelled by the Fréchet-Pareto class of distributions, while the risks in the Brazilian, Indian, Chinese, and South African equity markets can be modelled by the Weibull class of distributions. Furthermore, in terms of risk levels, the findings show that the Russian IMOEX market is the most risk-prone, while the least risky is the Indian NIFTY market, with the remaining three markets in between them. That is, the Russian IMOEX market has the highest level of risk, followed by the South African JALSH market, then the Chinese SHCOMP, Brazilian IBOV and Indian NIFTY markets, respectively.

To find out the validity of the fitted models, diagnostics checks were carried out using various diagnostic plots for the CEV's GPD and point process models, and also goodness of fit via the Cramer-von Mises and Anderson-Darling tests for the GPD, and the results show that the models' fits are adequately satisfactory. The model diagnostics used are the probability plots, quantile plot, return level plots, and histogram and density plots. Furthermore, the return level estimates are nearly the same for the two models. In conclusion, the likelihood estimations of the two models are approximately the same, but the GPD has fewer parameters to compute than the point process.

7.3.3 Extremal dependence modelling via multivariate models

The two models, i.e., bivariate point process and threshold excess method of CMEV model, used for the extremal dependence modelling were compared in this study using the same dependence threshold. However, it is observed that the bivariate point process was able to model many more extreme observations or exceedances that contribute to the likelihood estimation and it gives more information than the threshold excess method of the CMEV model.

The outcomes of both models signify varied levels of low extremal dependence that can be stated under three distinct categories. Under the first category of weak extremal (asymptotic) dependence, our findings show the possibility of a beneficial risk reduction and high investment returns through international portfolio diversifications as follows: First, if a Brazilian investor invests in the domestic market and any of the Indian and South African markets; second, if a Russian investor jointly invests in the domestic market and any of the Indian, Chinese, and South African markets; third, if an Indian investor invests in the domestic market and in the Chinese market; fourth, if a Chinese investor invests in the domestic market and in the South African market. Moreover, any authorised international investor outside any of these countries may likewise invest comfortably and obtain good returns from any of the markets' combinations.

Under the second category of fairly strong (but still low) extremal dependence, a fairly good investment opportunity derivable from international portfolio diversifications can be expected as follows: First, a Brazilian investor can earn a fairly good returns by diversifying in the domestic market and in the Russian stock market; and second, an Indian investor can also obtain beneficially fair portfolio diversification and investment returns jointly from the domestic market and the South African stock

market. This opportunity is also open to international investors who are interested in portfolio investments in these market pairs.

The findings under the third category show that a Brazilian investor can derive a beneficial or fairly beneficial investment return outcome from diversifying in the domestic market and in the Chinese stock market.

7.4 Concluding remarks

The summary on extremal dependence results as stated will be highly beneficial to investors, portfolio managers, and other market participants who are interested in maximising their investment returns and financial gains, and in the process mitigate possible investment downturns through international portfolio diversification in the BRICS stock markets.

Based on the various modelling results and findings, this study provides answers to the five research questions stated in chapter 1 as follows: First, considering the risk hierarchy of the BRICS stock markets in a descending order of magnitudes, the most susceptible to risk is the Russian IMOEX market. This is followed by the South African JALSH market, then the Chinese SHCOMP and Brazilian IBOV markets respectively, and the least risky is the Indian NIFTY markets. Second, international portfolio diversification is strongly realistic between each of the selected pairs of the BRICS stock markets because of the observed varied levels of low extremal dependence. Third, the trade and economic relations of the BRICS nations do not constitute a high, but rather low, extremal dependencies in the pairs of the markets.

Fourth, it can be seen from the findings and conclusions that fundamental factors like trade or economic relations between each of the market pairs do not result to serious asymptotic dependence in the markets. In that case, what may potentially cause volatility or risk spill-over in the markets may well be "contagion effect". Contagion

happens with investors hastily changing their investment positions due to concerns that crisis in one market will sooner or later spill-over to the other markets, and that may eventually lead to "herd behaviour effects", i.e., panic selling. Fifth, the two EVT models of conditional extreme value (CEV) and point process are adequately satisfactory for modelling the risk, based on the results of various diagnostics and tests under the univariate modelling. For extremal dependence modelling however, the bivariate point process was able to model many more extreme observations or exceedances that contribute to the likelihood estimation and it gives more information than the threshold excess method of the CMEV model.

7.5 Limitations of the Thesis

In this thesis, modelling of each market's volatility was limited to the assumptions of seven error distributions. This is perceived as one of the limitations since we could possibly have tried other fat-tailed error distributions in addition to the ones used, in order to ascertain the most suitable error distribution when the true market's return distribution is unknown. Another supposed limitation is that we only used MLE throughout the study for the parameters' estimations. Another option would have been the use of Bayesian approach of parameter estimation and then compare its results with that of the benchmark MLE.

7.6 Future research studies

The outcomes of this thesis offer possible areas for future research as follows:

- Future research may involve the use of a robust Bayesian method of parameter estimation using different prior distributions and then compare the results with the outcomes from a benchmark MLE approach.

- Although the standard GARCH model is pervasive for filtering and modelling financial return volatility, recent studies have shown that variability involving regime changes in the dynamics of volatility can be swiftly modelled using the Markov-switching GARCH (MSGARCH) model (Ardia et al., 2019a). The features of MSGARCH model also include assessment of risk management metrics like the Value-at-Risk (VaR) and expected shortfall (ES) methods. Moreover, apart from its swift ability to estimate return volatility, like the standard GARCH model, the MSGARCH model can likewise be well used for simulation and forecasting using the *R* package MSGARCH (Ardia et al., 2019b). The model allows for different behaviour of GARCH in each regime by capturing the difference in the variance dynamics of high and low volatility periods. The use of this model on the BRICS equity markets data will be fascinating for future study.

- Another possible future research attempt may be to subject or rather to further open the empirical modelling of these markets to simulation-based tests. One vital approach that is useful for evaluating issues that relate to the applications of statistics is to use Monte Carlo simulation experiments (Sigal and Chalmers, 2016). Simulation modelling and diagnostics of the empirical results may possibly yield more refined results for risk management and portfolio diversification.

- Another area of interest for future research will be the modelling of the BRICS bond markets. This can be interestingly done using both simulation and empirical modelling to model and analyse the volatility, risks, and asymptotic dependencies of the markets.

- It is also worth mentioning that apart from the BRICS equity markets that this study is focused on, future study will be extended to other global sectors with extreme event issues.

References

1. Adua G., Alagidede P. and Karimu A., Stock return distribution in the BRICS, *Review of Development Finance*, 5, (2015), 98-109.

2. Alagidede P., How integrated are Africa's stock markets with the rest of the world?, *Department of Economics, University of Stirling*, (2008).

3. Afuecheta E., Utazi C., Ranganai E. and Nnanatu C., An Application of extreme value theory for measuring Financial Risk in BRICS Economies, *Annals of Data Science*, from https://doi.org/10.1007/s40745-020-00294-w, (2020).

4. Aggarwal S. K, Saini L. M. and Kumar A., Load/price forecasting and managing demand response for smart grids, *Electr Power Energy Syst*, 31, (2009), 13-22.

5. Akaike H., A new look at the statistical model identification, *IEEE Transaction on Automatic Control*, 19, (1974), 716-723.

6. Al-Janabi M. A. M., Risk management in trading and investment portfolios: An optimisation algorithm for maximum risk-budgeting threshold, *Journal of Emerging Market Finance, SAGE Publications*, 11, (2012), 189-229.

7. Allen D. E., Singh A. K. and Powell R. J., EVT and tail-risk modelling: Evidence from market indices and volatility series, *The North American Journal of Economics and Finance*, 26, (2013), 355-369.

8. Aloui R., Aissa M. S. B., and Nguyen D. K., Global financial crisis, extreme interdependences, and contagion effects: The role of economic structure? *Journal of Banking & Finance*, 35 (2011) 130-141.

9. Anderson E. W. and Mansi S. A., Does customer satisfaction matter to investors? Findings from the bond market, *Journal of Marketing Research*, XLVI, (2009), 703-714.

10. Andersen T. G. and Bollerslev T., Answering the skeptics: yes, standard volatility models do provide accurate forecasts, *Int Econ Rev* 39, (1998), 885-905.

11. Andersen T. G., Bollerslev T. and Christoffersen P. F., Diebold FX practical volatility and correction modeling for financial markets risk management. In: *Carey M, Schultz R (eds) Risk and financial institutions, University of Chicago Press for NBER, Chicago*, (2006).

12. Ang A., and Bekaert G., Timing and diversification: A state-dependent asset allocation approach, *Financial Analysts Journal*, 60 (2004), 86-99.

13. Ardia D., Bayesian estimation of the GARCH(1,1) model with normal innovations, *Student*, 5, (2006), 283-298 http://ssrn.com/abstract=1543409.

14. Ardia D., Boudt K., and Catania L., Downside risk evaluation with the R package GAS, *The R Journal* 10, (2018), 410-421.

15. Ardia D., Boudt K., and Catania L., Generalized autoregressive score models in R: The GAS package, *Journal of Statistical Software*, doi: 10.18637/jss.v088.i06, 88(6).

16. Ardia D., Boudt K. and Catania L., Value-at-Risk prediction in R with the GAS Package, in *SSRN Electronic Journal*, DOI: 10.2139/ssrn.2871444, (2016), 1-10.

17. Ardia D., Bluteau K., Boudt K., Catania L., Ghalanos A., Peterson B. and Trottier D. A., MSGARCH: Markov-Switching GARCH models in R. *R package version 2.31* (2019b), URL https://CRAN.R-project.org/package=MSGARCH.

18. Ardia D., Bluteau K., Boudt K., Catania L. and Trottier D. A., Markov-Switching GARCH models in R: The MSGARCH Package, *Journal of Statistical Software*, 91, (2019a).

19. Ardia D. and Hoogerheide L. F., Bayesian estimation of the GARCH(1,1) Model with Student-t innovations, *The R Journal*, 2, (2010), 41-47.

20. Ardia D., Lennart H. and Nienke C., Stock index returns' density prediction using GARCH models: Frequentist or Bayesian estimation?, *aeris CAPITAL AG, Tinbergen and Econometric Institutes, Erasmus University Rotterdam, The Netherlands*, (2011), from https://mpra.ub.uni-muenchen.de/28259/.

21. Arltová M. and Fedorová D., Selection of unit root test on the basis of length of the time series and value of AR(1) parameter, *STATISTIKA*, 96, (2016).

22. Ashour S. and Abdel-Hameed M., Approximate skew normal distribution, *Journal of Advanced Research*, 1, (2010), 341-350.

23. Azzalini A. A., Class of distributions which includes the normal ones, *Scand J Stat*, (1985), 171-178.

24. Auguie B. and Antonov A., Miscellaneous functions for "grid" graphics, *Package 'gridExtra Version 2.0.0*, URL https://github.com/baptiste/gridextra, (2015).

25. Azzalini A. and Capitanio A., Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution, *J. R. Stat. Soc. Ser. B*, 65, (2003), 367-389.

26. Azzalini A. and Capitanio A., Statistical applications of the multivariate skew-normal distribution, *Journal of the Royal Statistical Society Series B*, 61, (1999), 579-602.

27. Baba Y., Engle R. F., Krafet D. and Kroner K. F., Multivariate simultaneous generalized ARCH unpublished manuscript, *Department of Economics, University of California, San Diego*, (1990).

28. Babu M., Hariharan C. and Srinivasan S., Testing the co-movement of BRICS nations' capital markets, *IIMS Journal of Management Science*, 6, (2015), 213-222.

29. Bader B. and Yan J., eva: Extreme value analysis with goodness-of-fit testing, *R package version 0.2.3*, (2016).

30. Bae K., Karolyi G. A. and Stulz R. M., A new approach to measuring financial contagion, *Review of Financial Studies*, 16, (2003), 717-63.

31. Bali T. G., An extreme value approach to estimating volatility and value at risk, *Journal of Business*, 76, (2003), 83-108.

32. Bali T. G., Testing the empirical performance of stochastic volatility models of the short term interest rate, *Journal of Financial and Quantitative Analysis*, 35, (2000), 191-215.

33. Bali T. G. and Neftci S. N., Disturbing extremal behavior of spot price dynamics, *Journal of Empirical Finance*, 10, (2003), 455-477.

34. Bartram S. M. and Dufey G., International portfolio investment: theory, evidence, and institutional framework, *Financial Markets, Institutions & Instruments*, 10, (2001), 85-155.

35. Bates J. M. and Granger C. W. J., The combination of forecasts, *Operational Research*, 20, (1969), 451-468.

36. Batista L., Gong Y., Pereira S. and Jia F., Circular supply chains in emerging economies – a comparative study of packaging recovery ecosystems in China and Brazil, *International Journal of Production Research*, 57, (2018), 1-21, DOI:10.1080/00207543.2018.1558295.

37. Baur D. and Jung R. C., Return and volatility linkages between the US and the German stock market, *Journal of International Money and Finance*, 25, (2006), 598-613.

38. Bekaert G, Harvey C. and Ng A., Market integration and contagion, *The Journal of Business*, 78, (2005), 39-70.

39. Bekiros S. D. and Georgoutsos D. A., The extreme-value dependence of Asia-Pacific equity markets, *Journal of Multinational Financial Management*, 18, (2008), 197–208.

40. Ben-Rejeb A. Volatility spillovers and contagion: An empirical analysis of structural changes in emerging market volatility, *Economics Bulletin*, 33, (2013), 56–71.

41. Bensalah Y., Steps in applying extreme value theory to finance: A review, *Working Paper* 2000-20 (Bank of Canada), (2000).

42. Beirlant J., Goegebeur Y. and Teugels J., *Statistics of extremes: Theory and applications*, Wiley, Chichester, (2004).

43. Bere A., Some non-standard statistical dependence problems, *University of the western cape*, viewed 9 September 2021. from http://etd.uwc.ac.za/xmlui/handle/11394/4868, (2016), 1-132.

44. Berger J. O., Robust Bayesian analysis: sensitivity to the prior, *J Stat Plan Inference*, 25, (1990), 303-328.

45. Berger J. O., Robustness of Bayesian analyses, Studies in Bayesian econometrics, North-Holland, Ch. *The robust Bayesian viewpoint*, (1984), 63-124.

46. Bickel D. R., Blending Bayesian and frequentist methods according to the precision of prior information with applications to hypothesis testing, *Stat Methods Appl.*, 24, (2015), 523-546.

47. Bollerslev T., A conditional heteroskedastic time series model for speculative prices and rates of return, *Review of Economics and Statistics*, (1987), 542-547.

48. Bollerslev T., Generalized autoregressive conditional heteroscedasticity, *Journal of Econometrics*, 31, (1986), 307-327.

49. Bommier E., Peaks-over-threshold modelling of environmental data, *UPPSALA UNIVERSITET, U.U.D.M. Project Report*, (2014), 1-35.

50. Box G. E. P., Science and Statistics (PDF), *Journal of the American Statistical Association*, 71, (1976), 791-799, doi:10.1080/01621459.1976.10480949).

51. Box G. E. P. and Draper N. R., *Empirical model-building and response surfaces*, John Wiley & Sons, (1987).

52. Box G. E. P., Hunter J. S. and Hunter W. G., *Statistics for Experimenters* (2nd ed.), John Wiley & Sons, (2005), 208, 384, 440.

53. Boyer B. H., Gibson M. S. and Loretan M., Pitfalls in tests for changes in correlations, *International Finance Discussion Paper No. 597* (Washington: Board of Governors of the Federal Reserve System), (1997).

54. Bozdogan H., Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions, *Psychometrika*, 52, (1987), 345-370.

55. Branco M. D. and Dey D. K., A general class of multivariate skew-elliptical distributions, *Journal of Multivariate Analysis*, 79, (2001), 99-113.

56. Breheny P., Model selection I, BST 760: *Advanced Regression*, from http://web.as.uky.edu/statistics/users/pbreheny/teaching.html, (2013).

57. Brooks P., Power ARCH modeling of the volatility of emerging equity markets, *Emerging Markets Review*, 8, (2007), 124-133.

58. Buckland S. T., Burnham K. P. and Augustin N. H., Model selection: An integral part of inference, *Biometrics*, 53, (1997), 603-618, doi:10.2307/2533961.

59. Burnham K. P. and Anderson D. R., *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.), New York, NY: Springer, (2003).

60. Buuren S. and Groothuis-Oudshoorn K., MICE: Multivariate imputation by chained equations in R, *Journal of Statistical Software*, 45, (2011), 1-67.

61. Cahill A. T., Significance of AIC differences for precipitation intensity distributions, *Advances in Water Resources*, 26, (2003), 45-464.

62. Castillo E., Ali S., Hadi A. S., Balakrishnan N. and Sarabia J. M., *Extreme value and related models with applications in Engineering and Science*, Wiley, New York, (2005), 241-250.

63. Chandra A. and Thenmozhi M., On asymmetric relationship of India volatility index (India VIX) with stock market return and risk management, *Indian Institute of Management Calcutta*, 42, (2014), 33-55.

64. Chan-Lau, J. A., Mathieson D. and Yao J. Y., Extreme contagion in equity markets, *IMF Staff Papers*, 51, (2004), 386-408.

65. Chatfield C., *Time-series forecasting*, Chapman & Hall/CRC, (2000).

66. Chaurasia A. and Harel O., Model selection rates of information based criteria, *Electronic Journal of Statistics*, 7, (2013), 2762-2793.

67. Chavez-Demoulin V. and McGill J. A., High-frequency financial data modeling using Hawkes processes, *Journal of Banking & Finance*, 36, (2012), 3415-3426.

68. Chen J., Brazil, Russia, India, China and South Africa (BRICS), viewed 17 June 2020, from https://www.investopedia.com/terms/b/brics.asp, (2020).

69. Chen J., International Portfolio, viewed 14 December 2020, from https://www .investopedia.com/terms/i/international-portfolio.asp, (2020).

70. Cheung Y. W. and Ng L. K., The dynamics of S&P 500 index and S&P 500 futures intraday price volatilities, *Review of Futures Markets*, 9, (1990), 458-486.

71. Chiang T. C. and Doong S. C., Empirical analysis of stock returns and volatility: Evidence from seven Asian stock markets based on TAR-GARCH model, *Review of Quantitative Finance and Accounting*, 17, (2001), 301-318.

72. Christensen W., Model selection using information criteria (Made Easy in SAS®), *University of California, Los Angeles*, wchristensen@ucla.edu., https:// github.com/wendychristensen, (2018).

73. Christian P. R., Error and inference: an outsider stand on a frequentist Philosophy, *Theory* Dec., 74, (2013), 447-461.

74. Christiansen C., Decomposing European bond and equity volatility, *International journal of finance & economics*, 15, (2010), 105-122.

75. Claeskens, G. and Hjort N. L., *Model selection and model averaging*, Cambridge University Press, (2008).

76. Clemen R. T., Combining forecasts: a review and annotated bibliography, *Int J Forecast*, 5, (1989), 559-583.

77. Coles S., *An Introduction to statistical modelling of extreme values*, (2001), 81-156.

78. Coles S., Heffernan J. and Tawn J., Dependence measures for extreme value analysis, *Extremes*, 2, (1999), 339-65.

79. Coles S. G. and Tawn J. A., Modelling extreme multivariate events, *Journal of the Royal Statistical Society*, 53, (1991), 377-392.

80. Costinot A., Thierry R. and Jerome T., Revisiting the dependence between financial markets with copulas, *Working Paper (Paris: Credit Lyonnais)*, (2000).

81. Cotter J., Varying the VaR for unconditional and conditional environments, *Journal of International Money and Finance*, 26, (2007), 1338-1354.

82. Cox D. R. and Lewis P. A. W., Multivariate point processes, Proc. 6th Berkeley Symp, *Math. Statist.* Prob. 3, (1972), 401-448.

83. Cox D. R. and Mayo D. G., Objectivity and conditionality in frequentist inference, in *error and inference: Recent exchanges on experimental reasoning, Reliability, and the Objectivity and Rationality of Science, eds. D. Mayo and A. Spanos, Cambridge, Cambridge University Press*, (2010), 276-304.

84. Craven P. and Wahba G., Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross validation, *Numer. Math*, 31, (1979), 377-403.

85. Dacorogna M. M., Müller U. A., Pictet O. V. and DeVries C. G., The Distribution of extremal foreign exchange rate returns in extremely large data sets, Preprint, *O & A Research Group*, (1995).

86. D'Agostino R. B. and Stephens M. A., *Goodness-of-fit techniques*, Marcel Dekker, Inc., New York, (1986).

87. Daniel K., Hirshleifer D., and Teoh S. H., Investor psychology in capital markets: evidence and policy implications, *J Monet Econ*, 49, (2002), 139-209.

88. Danielsson J. and de Vries C., Tail index and quantile estimation with very high frequency data, *Journal of Empirical Finance*, 4, (1997), 241-257.

89. Danielsson J., Hartmann P. and de Vries C., The cost of conservatism, *RISK*, 11, (1998), 101-103.

90. David R. B., Blending Bayesian and frequentist methods according to the precision of prior information with applications to hypothesis testing, *Stat Methods Appl*, 24, (2015), 523-546.

91. Davidson R. and MacKinnon J. G., *Econometric theory and methods*, New York, Oxford University, 613, (2004), ISBN 0-19-512372-7.

92. Davidson A. C. and Smith R. L., Models for exceedances over high thresholds, *Journal of the Royal Statistical Society B*, 52, (1990), 393-442.

93. Davidson R. and MacKinnon J. G., *Econometric theory and methods*, New York, Oxford University, 613, (2004), ISBN 0-19-512372-7.

94. Daziano R. and Bolduc D., Covariance, identification, and finite-sample performance of the MSL and Bayes estimators of a logit model with latent attributes, *Transportation*, 40, (2012), 647-670.

95. de Haan L., Extremes in higher dimensions: The model and some statistics, in *"Proceedings of the 47th Session of the International Statistical Institute"*, Amsterdam, Netherlands, (1985), 1-16.

96. de Haan L. and Ferreira A., *Extreme Value Theory: an Introduction*, Springer Science+Business Media, LLC, New York, (2006).

97. de Haan L. and de Ronde J., Sea and wind: Multivariate sample extremes at work, *Extreme*, 1, (1998), 7-45.

98. de Haan L. and Resnick S. I., Limit theory for multivariate sample extremes, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 40, (1977), 317-337.

99. de Hoog F. R. and Hutchinson M. F., An efficient method for calculating smoothing splines using orthogonal transformations, *Numer. Math*, 50, (1987), 311-319.

100. de Jesus R., Ortiz E. and Cabello A., Long run peso/dollar exchange rates and extreme value behavior: Value at risk modeling, *The North American Journal of Economics and Finance*, 24, (2013), 139-152.

101. de Menezes L. M., Bunn D. W. and Taylor J. W., Review of guidelines for the use of combined forecasts, *Eur J Oper Res*, 120, (2000), 190-204.

102. Devaine M., Gaillard P., Goude Y. and Stoltz G., Forecasting the electricity consumption by aggregating specialized experts: A review of sequential aggregation of specialized experts, with an application to Slovakian and French country-wide one-day-ahead (half-) hourly predictions, *Machine Learning*, 90, (2012), 231-260.

103. Dickey D. A., Estimation and hypothesis testing in nonstationary time series, PhD dissertation, *Iowa State University*, (1976).

104. Ding Z., Granger C. W. J. and Engle R. F., A long memory property of stock

market returns and a new model, *Journal of Empirical Finance*, 1, (1993), 83-106.

105. Driessen J., Melenberg B. and Nijman T., Common factors in international bond returns, *Journal of International Money and Finance*, 22, (2003), 629-656.

106. Driessen J. and Laeven L., International portfolio diversification benefits: Cross-country evidence from a local perspective, *Journal of Banking & Finance*, 31, (2007), 1693-1712.

107. Djakovic V., Andjelic G. and Borocki J., Performance of extreme value theory in emerging markets: An empirical treatment, *African Journal of Business Management*, 5, (2011), 340-369.

108. Dufour J. M., Coefficients of determination, *McGill University*, (2011), 1-12.

109. Dufour J. M., Model selection, *Université de Montréal*, (2007).

110. Edwards A., A history of likelihood, *Internat Stat Rev*, 42, (1974), 9-15.

111. Eldén L., An algorithm for the regularization of ill-conditioned, banded least squares problems, *SIAM J. Sci. Stat. Comput*, 5, (1984), 237-254.

112. Elhorst J., *Spatial econometrics from cross-sectional data to spatial panels*, Berlin, Springer, (2014).

113. Eling M., Fitting asset returns to skewed distributions: Are the skew-normal and skew-student good models?, *Insurance: Mathematics and Economics*, 59, (2014), 45-56.

114. Embrechts P., Klüppelberg C. and Mikosch T., *Modeling extremal events for insurance and finance*, Springer, (1997).

115. Embrechts P., Lindskog F., McNeil A., Modelling dependence with copulas and applications to risk management, *In: Handbook of Heavy Tailed Distributions in Finance*, ed. S. Rachev, Elsevier, (2003), 329-384.

116. Embrechts P., McNeil A. and Straumann D., Correlation: Pitfalls and alternatives, *Risk*, 12, (1999), 69-71.

117. Embrechts P., Resnick S. and Samorodnitsky G., Living on the edge, Risk Magazine, 11, (1998), 96-100.

118. Emerson M., Do the BRICS make a bloc?, *CEPS Commentary*, (2012), 1-7.

119. Engle R., Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation, *Econometrica*, 50, (1982), 987-1006.

120. Engle R., Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business and Economic Statistics*, 20,

121. Engle R., GARCH 101: the use of ARCH/GARCH models in applied econometrics, *Journal of Economic Perspectives*, 15, (2001), 157-168.

122. Engle R. and Bollerslev T., Modelling the persistence of conditional variances, *Econometric Reviews*, 5, (1986), 1-50.

123. Engle R. F., Ito T. and Lin W., Meteor shower or heat wave? Heteroskedastic intra-daily volatility in the foreign exchange market, *Econometrica*, 59, (1990), 524-542.

124. Engle R. F. and Lee G. J., A permanent and transitory component model of stock return volatility, in R. F. Engle and H. White (eds.), cointegration,

causality, and forecasting: A festschrift in honor of Clive W. J. Granger, Oxford: *Oxford University Press*, (1999), 475-497.

125. Engle R. and Ng V. K., Measuring and testing the impact of news on volatility, *Journal of Finance*, 4, (1993), 15-36.

126. Engle R. F. and Susmel R., Common volatility in international equity markets, *Journal of Business and Economic Statistics*, 11, (1993), 167-176.

127. EViews 9.0., *EViews 9 users guide II*, IHS Global Inc, (2016).

128. Fabozzi F. J., *Encyclopedia of financial models*, 3 Volume Set, John Wiley & Sons, Inc., (2013).

129. Fayyad A. and Daly K., The volatility of market returns: a comparative study of emerging versus mature markets, *International Journal of Business Management*, (2010), 24-36.

130. Fazio G., Extreme interdependence and extreme contagion between emerging markets, *Journal of International Money and Finance*, 26, (2007), 1261-1291.

131. Fernandez V., The impact of major global events on volatility shifts: evidence from the Asian crisis and 9/11, *Economic Systems*, 30, (2006), 79-97.

132. Fernández C. and Steel M., On Bayesian modelling of fat tails and skewness, *Journal of the American Statistical Association*, 93, (1998), 359-371.

133. Ferreira A. and de Haan L., On the block maxima method in extreme value theory: PWM estimators, *The Annals of Statistics*, 43, (2015), 276-298.

134. Ferro C. A. T., Statistical methods for clusters of extreme values, *Lancaster University*, (2003) 1-167.

135. Ferro C. A. T. and Segers J., Inference for clusters of extreme values, *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, 65, (2003), 545-556.

136. Fisher T. J. and Gallagher C. M., New weighted portmanteau statistics for time series goodness of fit testing, *Journal of the American Statistical Association*, 107, (2012), 777-78.

137. Fisher R. A. and Tippett L. H. C., Limiting forms of the frequency distribution of the largest and smallest member of a sample, *Proc. Camb. Phil. Soc*, 24, (1928), 180-190.

138. Floros C., Modelling volatility using GARCH models: Evidence from Egypt and Israel, Middle Eastern *Finance and Economics*, 2, (2008), 31-41.

139. Forbes K. and Rigobon R., No contagion, only interdependence: measuring stock market co-movements, *Journal of Finance*, 57, (2002), 2223–61.

140. Forster M. R., Simplicity and unification in model selection, *University of Wisconsin–Madison*, Retrieved from http://philosophy.wisc.edu/forster/520/Chapter

141. French K. R., Schwert G. W. and Stambaugh R. F., Expected stock returns and volatility, *Journal of Financial Economics*, 19, (1987), 3-29.

142. Freund J. E., *Mathematical statistics*, 2nd ed. Englewood Cliffs, (NJ), Prentice Hall, (1971).

143. Fukutome S., Liniger M. A. and Süveges M., Automatic threshold and run parameter selection: a climatology for extreme hourly precipitation in Switzerland, *Theor Appl Climatol*, 120, (2015), 403-416.

144. Gaba A., Tsetlin I. and Winkler R. L., Combining interval forecasts, *Decision Analysis* Published online in Articles in Advance, (2017), 1-20, http://dx.doi.org/10.1287/deca.2016.0340.

145. Gaillard P., Contributions to online robust aggregation: Work on the approximation error and on probabilistic forecasting, *PhD Thesis, University Paris-Sud, France*, (2015).

146. Galambos J., *Asymptotic theory of extreme order statistics*, 2nd edition, Krieger, Malabar (Florida), (1987).

147. Galambos J., Order statistics of samples from multivariate distributions, *J. Amer. Statist. Assoc.*, 70, (1975), 674-680.

148. Galloppo G., Paimanova V. and Aliano M., Volatility and liquidity in Eastern Europe financial markets under efficiency and transparency conditions, *Economics and Sociology*, 8, (2015), 70-92.

149. Genest C. and Favre A. C., Everything you always wanted to know about copula modeling but were afraid to ask, *Journal of Hydrologic Engineering*, 12, (2007), 347-368.

150. Ghalanos A., Introduction to the rugarch package *(version 1.3-1)*, from https://cran.r-project.org/web/packages/rugarch/vignettes /Introduction to the rugarch package.pdf, (2015).

151. Ghini A. E. and Saidi Y., Return and volatility spillovers in the Moroccan stock market during the financial crisis, *Empir Econ*, 52, (2017), 1481-1504.

152. Ghorbel A. and Trabelsi A., Predictive performance of conditional extreme value theory in value-at-risk estimation, *International Journal of Monetary Economics and Finance*, 1, (2008), 121-148.

153. Gilleland E., extRemes: Extreme value analysis, *Version 2.1*, https://cran.r-project.org/web/packages/extRemes/index.html, (2020).

154. Gilleland E extRemes: Extreme value analysis, *R package* version 2.0-8, URL https://CRAN.R-project.org/package=extRemes, (2016).

155. Gilli M. and Kellezi E., An Application of extreme value theory for measuring financial risk, *Comput Econ*, 27, (2006), 1-23.

156. Glosten L., Jagannathan R. and Runkle D., On the relation between expected value and the volatility of the nominal excess return on stocks, *Journal of Finance*, 48, (1993), 1779-1801.

157. Gnedenko B. V., Sur la distribution limite du terme maximum d'une série aléatoire *Ann. Math*, 44, (1943), 423-453.

158. Goldman S. *Information Theory*, New York: Prentice Hall, New York: Dover 1968 ISBN 0-486-62209-6, 2005 ISBN 0-486-44271-3, (1953).

159. Gourieroux C. and Jasiak J., Financial Econometrics, *Princeton University Press*, Princeton, NJ, (2001).

160. Granger C., Robins P. P. and Engle R. F., Wholesale and retail prices: Bivariate time-series modeling with forecastable error variances. In: Belsley, D.A., Kuh, E. (Eds.), *Model Reliability*. MIT Press, Cambridge, (1986), 1-17.

161. Gregory A. W. and Hansen B. E., Residual-based tests for cointegration with model in regime shifts, *Journal of Econometrics*, 70, (1996), 99-126.

162. Grieb T. and Reyes M. G., Random walk tests for Latin American equity indexes and individual firms, *Journal of Financial Research*, 22, (1999), 371-383.

163. Grinstead C. M. and Snell J. L., Introduction to probability, 2nd ed. Washington, *DC: American Mathematical Society*, (1997).

164. Gu C., *Smoothing Spline ANOVA Models*, Springer, New York, (2002).

165. Gumbel E., Bivariate logistic distributions, *Journal of the American Statistical Association*, 56, (1961), 335-349.

166. Gumbel E. J., Bivariate exponential distributions, *Journal of the American Statistical Association*, 55, (1960), 698-707.

167. Hamaker E. L., van Hattum P., Kuiper R. M. and Hoijtink H., Model selection based on information criteria in multilevel modelling, In *Hox, J. J. & Roberts, J. K (Eds.)*, (2011).

168. Hamao Y. R., Masulis R. W. and Ng V. K., Correlations in price changes and volatility across international stock markets, *The Review of Financial Studies*, 3, (1990), 281-307.

169. Hammoudeh S., Kang S. H., Mensi W. and Nguyen D. K., Dynamic global linkages of the BRICS stock markets with the United States and Europe under external crisis shocks: implications for portfolio risk forecasting, *The World Economy*, (2016).

170. Hannan E. J. and Quinn B. G., The Determination of the order of an autoregression, *Journal of the Royal Statistical Society, Series B (Methodological)*, 41, (1979), 190-195, ISSN00359246. URL http://www.jstor.org/stable/2985032.

171. Hartmann P., Straetsman S. and De Vries C. G., Asset market linkages in crisis periods, *Review of Economics and Statistics*, 86, (2001), 313-326.

172. Harvey A. C., Dynamic Models for volatility and heavy tails: With applications to financial and economic time series, *Cambridge University Press*, 52, (2013).

173. Harvey A. and Sucarrat G., Cambridge working papers in Economics 0840, *Faculty of Economics, University of Cambridge*, 2, (2008).

174. Harvey A. and Sucarrat G., EGARCH models with fat tails, skewness and leverage, *Computational Statistics and Data Analysis*, 26, (2014), 320-338.

175. Hassan A. R., The interplay between the Bayesian and frequentist approaches: a general nesting spatial panel data model, *Spatial Economic Analysis*, 12, (2017), 92-112, http://dx.doi.org/10.1080/17421772.2017.1248478.

176. Heffernan J. E. and Southworth H., texmex: Statistical modelling of extreme values, *R package version 2.1.*, (2013).

177. Heffernan J. E., Stephenson A. G. and Gilleland E., Ismev: An introduction to statistical modeling of extreme values, *R package version 1.41*, URL https://cran.r-project.org/package=ismev, (2016).

178. Heffernan J. E. and Tawn J., A conditional approach for multivariate extreme values, *Journal of the Royal Statistical Society Series B*, 56, (2004), 497-546.

179. Hentschel L., All in the family nesting symmetric and asymmetric GARCH models, Journal of Financial Economics, Elsevier, 39, (1995), 71-104.

180. Herrera R. and Schipp B., Statistics of extreme events in risk management: The impact of the subprime and global financial crisis on the German stock market, *North American Journal of Economics and Finance*, 29, (2014), 218-238.

181. Herwartz H. and Hafner C. M., A Lagrange multiplier test for causality in variance, *Economics Letters*, 93, (2006), 137-141.

182. Hill B. M., A simple general approach to inference about the tail of a distribution, *Ann Statist.*, 3, (1975), 1163-1174.

183. Hong Y., A test for volatility spillover with applications to exchange rates, *Journal of Econometrics*, 103, (2001), 183-224.

184. Hong Y., Liu Y. and Wang S., Granger causality in risk and detection of extreme risk spillover between financial markets, *Journal of Econometrics*, 150, (2009), 271-287.

185. Hora S. C., Fransen B. R, Hawkins N. and Susel I., Median aggregation of distribution functions, *Decision Anal*, 10, (2013), 279-291.

186. Hosking J. R. M., Wallis J. R. and Wood E. F., Estimation of the generalised extreme-value distribution by the method of probability-weighted moments, *Technometrics*, 27, (1985), 251-261.

187. Hu Y. and Scarrott C., evmix: An R package for extreme value mixture modeling, threshold estimation and boundary corrected kernel density estimation, *Journal of Statistical Software*, 84, (2018), 1-27.

188. Hutchinson M. F. and de Hoog F. R., Smoothing noisy data with spline functions1q1qaz, *Numer. Math*, 47, (1985), 99-106.

189. Hüsler J. and Reiss R., Maxima of normal random vectors: Between independence and complete dependence, *Statistics & Probability Letters*, 7, (1989), 283-286.

190. Hyndman R. J. and Fan S., Density forecasting for long-term peak electricity demand, *IEEE Transactions on Power Systems*, 25, (2010), 1142-1153.

191. Ijumba C., Multivariate analysis of the BRICS financial markets *(Unpublished master's thesis)*, http://hdl.handle.net/10413/11309, University of KwaZulu-Natal, South Africa, (2013).

192. IMF (International Monetary Fund), World economic outlook, hopes, realities, risks, (2013), [online; cited April 2013]. Available from: http://www.imf.org/external/pubs/ft/weo/2013/01/pdf/text.pd.

193. Jansen M., Generalized cross validation in variable selection with and without shrinkage, *Université Libre de Bruxelles, Departments of Mathematics and Computer Science*, (2015).

194. Janssen R. and De Boeck P., Confirmatory analyses of componential test structure using multidimensional item response theory, *Multivariate Behavioral Research*, 34, (1999), 245-268.

195. Javed V. and Mantalos P., GARCH-type models and performance of information criteria Commun, *Stat. - Simul. Comput*, 42, (2013), 1917-1933.

196. Jawadi F. and Sousa R., Structural breaks and nonlinearity in US and UK public debts, *Applied Economics Letters*, 20, (2013), 653-57.

197. Jegadeeshwaran M. and Sangeetha V. M., Causal relationship among the stock markets: An empirical study on BRICS countries, *International Journal of Academic Research and Development*, 3, (2018), 100-105.

198. Ji Q., Liu B-Y., Cunado J. and Gupta R., Risk spillover between the US and the remaining G7 stock markets using time-varying copulas with Markov switching: Evidence from over a century of data, *The North American Journal of Economics and Finance, Elsevier*, 51, Article ID 100846, (2020).

199. Joe H., *Multivariate models and dependence concepts*, Chapman & Hall, London, (1997).

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

201. Kang S. H., McIver R. and Yoon S., Modelling time-varying correlations in volatility between BRICS and commodity markets, *Emerging Markets Finance & Trade*, 52, (2016), 1698-1723.

202. Karmakar M. and Paul S., Intraday risk management in International stock markets: A conditional EVT approach, *International Review of Financial Analysis*, http://dx.doi.org/10.1016/j.irfa.2015.11.008, (2015).

203. Karmakar M. and Shukla G. K., Managing extreme risk in some major stock markets: An extreme value approach, *International Review of Economics and Finance*, 35, (2015), 1-25.

204. Kassimatis K., Financial liberalization and stock market volatility in selected developing countries, *Applied Financial Economics*, 12, (2002), 389-394.

205. Kennedy P., *A guide to econometrics*, 5th ed. Cambridge (MA): MIT Press, (2003).

206. Kennedy R. E. and Sharma J. L., A comparative analysis of stock price behaviour of the Bombay, London and New York stock exchanges, *Journal of Financial and Quantitative Analysis*, 12, (1977), 391-413.

207. Khosravi A., Nahavandi S., Creigton D. and Atiya F., Lower upper bound estimation method for construction of neural network-based prediction intervals, *IEEE Transactions on Neural Networks*, 22, (2011), 337-346.

208. Kim E. H. and Singal V., Stock market openings: Experience of emerging economies, *Journal of Business*, 73, (2000), 25-66.

209. Kim H., Kim S., Shin H. and Heo J-H., Appropriate model selection methods for nonstationary generalized extreme value models, *Journal of Hydrology*, (2017), doi: http://dx.doi.org/10.1016/j.jhydrol.2017.02.005.

210. Kim S. Y., Huh D., Zhou Z., and Mun E. Y., A comparison of Bayesian to maximum likelihood estimation for latent growth models in the presence of a binary outcome, *International Journal of Behavioural Development*, (2020), 1–11.

211. King M., Sentana E. and Wadhwani S., Volatility and links between national stock markets, *Econometrica*, 62, (1994), 901-933.

212. King M. and Wadhwani S., Transmission of volatility between stock markets, *Review of Financial Studies*, 3, (1990), 5-33.

213. Konishi S. and Kitagawa G., Generalised information criteria in model selection, *Biometrika*, 83, (1996), 875-890.

214. K'onya L., Saving and investment rates in the BRICS countries, *The Journal of International Trade & Economic Development*, 24, (2015), 429-449.

215. Koutmos G., Asymmetries in the conditional mean and the conditional variance: Evidence from nine stock markets, *Journal of Economics and Business*, 50, (1998), 277-90.

216. Kuepper J., International diversification: Example portfolios, (2019), *The balance*, viewed 14 December 2020, from https://www.the balance.com/international-diversification-example-portfolios-4148204.

217. Kullback S. and Leibler R. A., On information and sufficiency, *The Annals of Mathematical Statistics*, 22, (1951), 79-86.

218. Kwiatkowski D., Phillips P. C. B., Schmidt P. and Shin Y., Testing the null hypothesis of stationarity against the alternative of a unit root, *Journal of Econometrics*, 54, (1992), 159-178.

219. Laurent S., G@RCH-site-Book63, from *http://www.core.ucl.ac.be/ laurent/ G@RCH/site/Book63.html*, (2010).

220. Lavine M., Sensitivity in Bayesian statistics: the prior and the likelihood, *JAm-Stat Assoc.*, 86, (1991), 396-399.

221. Leadbetter M., Lindgren G. and Rootzen H., *Extremes and related properties of random sequences and processes*, Springer series in statistics, Springer-Verlag, (1983).

222. Lee C., Chen M., and Sun E., Member states' pact and industry co-movements in the BRICS markets, *Applied Economics*, 49, (2017), 313-334.

223. Lee G. J. and Engle R. F., A permanent and transitory component model of stock return volatility, in Robert F. Engle and Halbert White (eds.), Cointegration Causality and Forecasting A Festschrift in Honor of Clive W. J. Granger. *Oxford: Oxford University Press*, (1999), 475-497.

224. Lee V. and Pai T. Y., REIT volatility prediction for skew-GED distribution of the GARCH model, *Expert Systems with Applications, Cardiff*, 37, (2010), 4737-4741.

225. Lee Y. H. and Pai T. Y., REIT volatility prediction for skew-GED distribution of the GARCH model, *Expert Systems with Applications*, 37, (2010), 4737-4741.

226. Levy H. and Sarnat M., International diversification of investment ponfolios, *American Economic Review*, 60, (1970), 668-675.

227. Li F., Cohen A. S., Kim S. and Cho S., Model selection methods for mixture dichotomous IRT models, *Applied Psychological Measurement*, 33, (2009), 353-373.

228. Lin W. L., Engle R. F. and Ito T., Do bulls and bears move across borders? International transmission of stock returns and volatility, *Review of Financial Studies*, 7, (1994), 507-538.

229. Liow K. H., Volatility spillover dynamics and relationship across G7 financial markets, North American, *Journal of Economics and Finance*, 33, (2015), 328-365.

230. Lipika B., Multivariate extreme value theory with an application to climate data in the Western Cape province, *Department of statistical sciences, University of Cape Town*, (2018).

231. Liu C. and Aitkin M., Bayes factors: prior sensitivity and model generalizability, *J Math Psychol*, 52, (2008), 362-375.

232. Longin F. M., From Value at Risk to stress testing: The extreme value approach, *Journal of Banking and Finance*, 24, (2000), 1097-1130.

233. Longin F. and Solnik B., Extreme correlation of international equity markets, *Journal of Finance*, 56, (2001), 649-676.

234. Longstaff F., The subprime credit crisis and contagion in financial markets, *Journal of Financial Economics*, 97, (2010), 436-450.

235. Lukas M. A., de Hoog F. R and Anderssen R. S., Efficient algorithms for robust generalized cross-validation spline smoothing, *Journal of Computational and A pplied Mathematics*, 235, (2010), 102-107.

236. Lyche T. and Schumaker L. L., Computation of smoothing and interpolating natural splines via local bases, *SIAM J, Numer. Anal, 10*, (1973), 1027-1038.

237. MacDonald A., Scarrott C. J., Lee D., Darlow B., Reale M. and Russell G., A flexible extreme value mixture model, *Comp. Statist. Data Anal.*, 55, (2011), 2137-2157.

238. MacKinnon J. G., Critical values for cointegration tests', in R. F. Engle and C. W. J. Granger (eds), Long-run Economic Relationships: Readings in Cointegration, *Oxford University Press, Oxford*, (1991), 267-276.

239. Makhwiting M. R., Lesaoana M. and Sigauke C., Modelling volatility and financial market risk of shares on the Johannesburg stock exchange, *African Journal of Business Management*, 6, (2012), 8065-8070.

240. Mandelbrot B., The variation of certain speculative prices, *The Journal of Business*, The University of Chicago Press, 36, (1963), 394-419.

241. Mandelbrot B., When can price be arbitraged efficiently? A limit to the validity of the random walk and martingale models, *Review of Economics and Statistics*, 53, (1971), 225-236.

242. Mandelbrot B., *Fractals and scaling in finance: discontinuity, concentration, Risk*, New York, N.Y. Springer Verlag, (1997).

243. Mandelbrot B., The stable Paretian income distribution, when the apparent exponent is near two, *The Journal of Business*, 36, (1963), 420-429.

244. Marimoutou V., Raggad B. and Trabelsi A., Extreme value theory and value at risk: application to oil market, *Energy Economics*, 31, (2009), 519-530.

245. McAleer M., Jimenez-Martin J. A. and Perez-Amaral T., Has the Basel accord improved risk management during the global financial crisis?, *The North American Journal of Economics and Finance*, 26, (2013), 250-265.

246. Maghyereh A. and Awartani B., Return and volatility spillovers between Dubai financial market and Abu Dhabi Stock Exchange in the UAE, *Applied Financial Economics*, 22, (2012), 837-848.

247. Makhwiting M. R., Lesaoana M. and Sigauke C., Modelling volatility and financial market risk of shares on the Johannesburg stock exchange, *African Journal of Business Management*, 6, (2012), 8065-8070.

248. Mandelbrot B., *Fractals and scaling in finance: discontinuity, concentration, risk*, New York, N.Y. Springer Verlag, (1997).

249. Mandelbrot B., The stable Paretian income distribution, when the apparent exponent is near two, *The Journal of Business*, 36, (1963), 420-429.

250. Mandelbrot B., The Variation of certain speculative prices, *The Journal of Business*, The University of Chicago Press, 36, (1963), 394-419.

251. Marimoutou V., Raggad B. and Trabelsi A., Extreme value theory and value at risk: application to oil market, *Energy Economics*, 31, (2009), 519-530.

252. Marques H. A., New M. B., Boock M. V., Barros H. P., Mallasen M. and Valenti W. C., Integrated freshwater prawn farming: state-of-the-art and future potential, *Rev. Fish. Sci. Aquac*, 24, (2016), 264-293.

253. Marques T. B., Do the political news impact financial markets? *Evidences from Brazil, Porto Alegre*, (2016).

254. McAleer M., Jimenez-Martin J. A. and Perez-Amaral T., Has the Basel accord improved risk management during the global financial crisis?, *The North American Journal of Economics and Finance*, 26, (2013), 250-265.

255. McLeod A. I. and Li W. K., Diagnostic checking ARMA time series models using squared residual autocorrelations, *Journal of Time Series Analysis*, 4, (1983), 269-273.

256. McNeil A. J. and Frey R., Estimation of tail related risk measure for heteroscedastic financial time series: An extreme value approach, *J. Empir. Financ.*, 7, (2000), 271-300.

257. Mensi W., Hammoudeh S., Yoon S. and Nguyen D. K., Asymmetric linkages between BRICS stock returns and country risk ratings: evidence from "dynamic panel threshold models" *Review of International Economics*, 24, (2016), 1-19.

258. Mikosch T., How to model multivariate extremes if one must?, *Statistica Neerlandica*, 59, (2005), 324-338.

259. Miralles-Marceloa J. L., Miralles-Quirósa J. L. and Miralles-Quirós M. M., Intraday linkages between the Spanish and the US stock markets: evidence of an overreaction effect, *Applied Economics*, 42, (2010), 23-35.

260. Muthén L. and Muthén B., Multilevel modeling with latent variables using Mplus, *Unpublished manuscript*, [Google Scholar], (2007).

261. Myung J., Model selection methods, Ohio State University, *Amsterdam Workshop on Model Selection*, (2004).

262. Mzamane T. P., GARCH Modelling of volatility in the Johannesburg stock exchange index, *MSc thesis*, http://researchspace.ukzn.ac.za/xmlui/handle/ 10413/ 10232, (2013).

263. Nam J. H., Yuhn K. H. and Kim S. B., What happened to pacific-basin emerging markets after the 1997 financial crisis?, *Applied Financial Economics*, 18, (2008), 39-58.

264. Nassif A., Feijo C. A. and Araújo E., Structural change, catching up and falling behind in the BRICS: A comparative analysis based on trade patterns and Thirlwall's law, *PSL Quarterly Review*, 69, (2016), 373-421.

265. Nelson D., Asymptotic filtering theory for multivariate ARCH models, *Journal of Econometrics*, 71, (1996), 1-47.

266. Nelson D. B., Conditional heteroscedasticity in asset returns: A new approach, *Econometrica*, 59, (1991), 347-370.

267. Nishii R., Asymptotic properties of criteria for selection of variables in multiple regression, *Annals of Statistics*, 12, (1984), 758-765.

268. Niyitegeka O. and Tewari D. D., Volatility clustering at the Johannesburg stock exchange: Investigation and analysis, Mediterranean *Journal of Social Sciences*, 4, (2013), 621-626.

269. Northrop P. J. and Coleman C. L., Improved threshold diagnostic plots for extreme value analyses, *Extremes*, 17, (2014), 289-303.

270. Nowotarski J. and Weron R., Computing electricity spot price prediction intervals using quantile regression and forecast averaging, Institute of Organization and Management, *Wrocław University of Technology, Wyb. Wyspia nskiego*, 27, 50-370 Wroclaw, Poland, Springerlink.com.

271. Odit M. P., Dookhan K. and Marylin. J. C., The impact of risk management and portfolio diversification on the Mauritian banking sector, *International Journal of Management & Information Systems* – Second Quarter, 15, (2011).

272. Ogum G., An analysis of asymmetry in the conditional mean returns: Evidence from three sub-Saharan Africa emerging equity markets, *African Finance Journal*, 4, (2002), 78-82.

273. Ole B. N., Exponentially decreasing distributions for the logarithm of particle size (1977), Proceedings of the Royal Society of London, Series A, Mathematical and Physical Science, *The Royal Society*, 353, (1674), 401-409.

274. Ole E. B., Mikosch T. and Resnick S. I., Lévy Processes: Theory and applications, *Birkhäuser*, (2013).

275. O'Donnell C. and Rayner V. Imposing stationarity constraints on the parameters of ARCH and GARCH models, *Advances in Econometrics*, 23, (2009), 545-566.

276. O'Sullivan F., Discussion of "some aspects of the spline smoothing approach to non-parametric regression curve fitting" by B.W. Silverman, *J. Roy, Statist. Soc. Ser. B*, 47, (1985), 39-40.

277. Peng X. Y., The downward trend of China's foreign exchange reserves has been formed, *Caixin* [online; cited October 2015], from: http://opinion.caixin.com/2015-10-09/100861118.html (in Chinese), (2015).

278. Pescetto G. M. and Appiah-Kusi J., Volatility and volatility spill-overs in emerging markets: The case of the African stock markets, *Ekonomia*, 2, (1998), 171-186.

279. Pickands J. III, Statistical inference using extreme order statistics, *Ann. Statist.*, 3, (1975), 119-131.

280. Pollmann M., Are you sure you are using the correct model? Model selection and averaging of impulse responses, *Marble series: Quantitative methods in Business and Economics*, 1, (2015), 1-45.

281. Poon S., Rockinger M. and Tawn J., New extreme-value dependence measures and finance applications, *CEPR Discussion Paper No. 2762 (London: Centre for Economic Policy Research)*, (2001).

282. Poon S., Rockinger M. and Tawn J., Extreme value dependence in financial markets: diagnostics, models, and financial implications, *Review of Financial Studies*, 17, (2004), 581-610.

283. Poon S-H., Rockinger M. and Tawn J., Modelling extreme-value dependence in international stock markets, *Statistica Sinica*, 13, (2003), 929-953.

284. Pourahmadi M., Construction of skew-normal random variables: Are they linear combination of normals and half-normals?, *J. Stat. Theory Appl.*, 3, (2007), 314-328.

285. Prescott P. and Walden A. T., Maximum likelihood estimation of the parameters of the generalized extreme-value distribution, *Biometrika*, 67, (1980), 723-724.

286. Pretorius E., Economic determinants of emerging market interdependence, *Emerging Markets Review*, 3, (2002), 84-105.

287. Quan H., Dipti Srinivasan D. and Khosravi A., Particle swarm optimization for construction of neural network-based prediction intervals, *Neurocomputing*, 127, (2014), 172-180.

288. Qiao Z., Liew V. K. and Wong W., Examining the impact of the U.S. IT stock market on other IT stock markets, *Handbook of Quantitative Finance and Risk Management*, Springer Science+Business.

289. Qiao Z. and Wong W., Revisiting volume vs. GARCH effects using univariate and bivariate GARCH models: Evidence from U.S. stock markets, C.-F. Lee et al. (eds.), *Handbook of Quantitative Finance and Risk Management*, Springer Science+Business Media, LLC, (2010).

290. Rana M., Koprinska I., Khosravi A. and Agelidis V. G., Prediction intervals for electricity load forecasting using neural networks, *Conference: Neural Networks (IJCNN)*, (2013).

291. Rao C. R. and Wu Y., On model selection, *IMS Lecture Notes - Monograph Series*, 38, (2001).

292. Reinsch C. H., Smoothing by spline functions, *Numer. Math*, 10, (1967), 177-183.

293. Reinsch C. H., Smoothing by spline functions II, *Numer. Math*, 16, (1971), 451-454.

294. Reisen H., Pensions, savings and capital flows from ageing to emerging markets, *Department Centre Studies, OECD, Paris*, (2000).

295. Reiss R. D. and Thomas M., *Statistical analysis of extreme values with applications to insurance, Finance, Hydrology and Other Fields*, Springer-Verlag, (2007).

296. Rejeb A. B. and Boughrara A., The relationship between financial liberalization and stock market volatility: the mediating role of financial crises, *Journal of Economic Policy Reform*, 17, (2014), 46-70.

297. Resnick I., *Extreme Values, Regular Variation, and Point Processes*, Springer, New York, 1987

298. Ribatet M. A., A user's guide to the POT package (Version 1.0), 12:69-71, *url = http://cran.r-project.org/*, (2006).

299. Ribatet M and Dutang C., POT: Generalized Pareto distribution and peaks over threshold, *Version: 1.1-5*, (2016).

300. Robert C. P., Error and inference: an outsider stand on a frequentist Philosophy, *Theory Dec*, 74, (2013), 447-461

301. Romero A. A. and Kasibhatla K. M., Volatility dynamics and volatility forecasts of equity returns in BRIC Countries, *Journal of business & economic studies*, 19, (2013).

302. Ross M., Relation of implicit theories to the construction of personal histories, *Psychol Rev 96*, (1989), 341-357.

303. Rossi R., Unit roots Tests, *Fin. Econometrics*, 10, (2014), 1-40.

304. Samiev S., GARCH (1,1) with exogenous covariate for EUR/SEK exchange rate volatility: On the effects of global volatility shock on volatility, *Digitala Vetenskapliga Arkivet*, (2012), https://www.diva-portal.org/smash/record.jsf?pid=diva2

305. Samuel R. A., Sigauke C. and Bere A., Modelling equity risk and extremal dependence: A survey of four African stock markets, *University of Venda*, from https://univendspace.univen.ac.za/handle/11602/1356.

306. Santos P. A., Alves I. F. and Hammoudeh S., High quantiles estimation with quasi-port and dpot: An application to value-at-risk for financial variables, *The North American Journal of Economics and Finance*, 26, (2013), 487-496.

307. Sashi S., Cracks in BRICs: A sectoral financial balances analysis and implications for macroeconomic policy, *Asociatia Generala a Economistilor din Romania - AGER, Autumn*, 3, (2016), 53-78.

308. Scarrott C., Hu Y. and Akbar A., 'evmix': Extreme value mixture modelling, Threshold Estimation and Boundary Corrected Kernel Density Estimation Version 2.12, *R Documentation*, (2019).

309. Scarrott C. and MacDonald A., A review of extreme value threshold estimation and uncertainty quantification, *REVSTAT Statistical Journal*, 10, (2012), 33-60.

310. Schloerke B., Crowley J., Cook D., Briatte F., Marbach M., Thoen E. and Elberg A., GGally: Extension to ggplot2, *R package version 1.3.2*, URL https://CRAN.R-project.org/package=GGally, (2018).

311. Segal T., Diversification, viewed 16 December 2020, from https://www.investopedia.com/terms/d/diversification.asp, *Investopedia*, (2020).

312. Shannon C. E. and Weaver W., The mathematical theory of communication, *Urbana, Illinois: University of Illinois Press.* ISBN 0-252-72548-4. LCCN 49-11922, (1949).

313. Shibata R., Consistency of model selection and parameter estimation, *Journal of Applied Probability, Essays in Time Series and Allied Processes*, 23, (1986), 127-141.

314. Shibata R., Statistical aspects of model selection, *IIASA International Institute for Applied Systems Analysis*, (1989).

315. Shono H., Is model selection using Akaike's information criterion appropriate for catch per unit effort standardization in large samples?, *Fisheries science*, 71, (2005), 978-986.

316. Sibuya M., Bivariate extreme statistics, I, *Annals of the Institute of Statistical Mathematics*, 11, (1960), 195-210.

317. Sigal M. J and Chalmers R. P., Play It Again: Teaching Statistics with Monte Carlo simulation, *Journal of statistics education*, 24, (2016), 136-156, http://dx.doi.org/10.1080/10691898.2016.1246953.

318. Sigauke C., Modelling electricity demand in South Africa, *PhD Thesis*, (2014), in https://scholar.ufs.ac.za/handle/11660/1569.

319. Sigauke C., Volatility modeling of the JSE all share index and risk estimation using the Bayesian and Frequentist approaches *Economics, Management, and Financial Markets*, 11, (2016), 33-48.

320. Sigauke C., Forecasting medium-term electricity demand in a South African electric power supply system, *The Journal of Energy in Southern Africa*, 28, (2017), 54-67, DOI:http://dx.doi.org/10.17159/2413-3051/2017/v28i4a2428.

321. Sigauke C., Maposa D., Mudimu E. and Nyamugure P., Volatility modeling using {ARIMA-GARCH models in a hyperinflationary economic environment: The {Zimbabwean experience, *SIAM In Peer-reviewed Proceedings of the Annual Conference of the South African Statistical Association*, (2010).

322. Silverman B. W., Density estimation for statistics and data analysis, *Chapman and Hall*, London, (1986).

323. Singla N., Jain K. and Sharma S. K., Goodness of fit tests and power comparisons for weighted gamma distribution, *REVSTAT – Statistical Journal*, 14, (2016), 29-48.

324. Skrepnek G. H., The contrast and convergence of Bayesian and Frequentist

statistical approaches in Pharmacoeconomic analysis, *PharmacoEconomics*, 25, (2007), 649-64.

325. Smith R. L., Maximum likelihood estimation in a class of non-regular cases, *Biometrika*, 72, (1985), 67-90.

326. Smith R. L., Threshold methods for sample extremes: Statistical extremes and applications *(J. Tiago de Oliveira, ed.)*, (1984), 621-638.

327. Smith R. L., Extreme value analysis of environmental time series: An application to trend detection in ground-level Ozone, *Statistical Science*, 4, (1989), 367-377.

328. Smith R. L., Statistics of extremes, with applications in environment, insurance, and finance, *Monographs on Statistics and Applied Probability*, (2003), 54-60.

329. Sohel-Azad A. S. M., Efficiency, cointegration and contagion in equity markets: Evidence from China, Japan and South Korea, *Asian Economic Journal*, 23, (2009), 93-118.

330. Solnik B., An equilibrium model of the international capital market, *Journal of Economic Theory*, 8, (1974), 500–24.

331. Solnik B., Why do not diversify internationally rather than domestically *Financial Analyst Journal*, 30, (1974), 48-54.

332. Southworth H. and Heffernan J. E., texmex: Statistical modelling of extreme values, *R package version 2.1.*, (2013).

333. Southworth H., Heffernan J. E. and Metcalfe P. D., texmex: Statistical modelling of extreme values, *R package version 2.4*, (2016).

334. Speigelhalter D. J., Abrams K. R. and Myles J. P., *Bayesian approaches to clinical trials and health-care evaluation*, West Sussex (UK): John Wiley & Sons, (2004).

335. Starica C., Multivariate extremes for models with constant conditional correlations, *Journal of Empirical Finance*, 6, (1999), 515-53.

336. Stavros D. and Evdokia X., Autoregressive Conditional Heteroskedasticity (ARCH) Models: A Review, Published in: *Quality Technology and Quantitative Management*, 1, (2004), 271-324.

337. Stephens M. A., Tests based on EDF statistics. In Goodness-of-Fit Techniques (R. B. D'Agostino and M. A. Stephens, Eds.), *Marcel Dekker*, New York, (1986), 97-193.

338. Stephenson A., Functions for extreme value distributions, *Package 'evd'*, viewed 01 December 2020, from https://cran.r-project.org/web/packages/evd/evd.pdf, (2018).

339. Stulz R. M., A model of international asset pricing, *Journal of Financial Economics*, 9, (1981), 383-406.

340. Suliman O., The large country effect, contagion and spillover effects in the GCC, *Applied Economics Letters*, 18, (2011), 285-94.

341. Susmel R. and Engle R. F., Hourly volatility spillovers between international equity markets, *Journal of International Money and Finance*, 13, (1994), 3-25.

342. Taleb N., *Fooled by Randomness: The hidden role of chance in the markets and life*, New York, N.Y. Random House, (2007).

343. Tancredi A., Anderson C. W. and O'Hagan A., Accounting for threshold uncertainty in extreme value estimation, *Extremes*, 9, (2006), 87-106, doi:10.1007/s10687-006-0009-8.

344. Tawn J., Bivariate extreme value theory: models and estimation, *Biometrika*, 75, (1988), 397-415.

345. Tawn J. A., An extreme value theory model for dependent observations, *J. Hydrology*, 101, (1988), 227-250.

346. Taylor S. J. and Poon S. H., Stock returns and volatility: An empirical study of the UK stock market, *Journal of Banking and Finance*, 16, (1992), 37-59.

347. Theil H., *Economic forecasts and policy*, 2nd Edition, North-Holland, Amsterdam, (1961).

348. Tian H., The BRICS and the G20, *China & world economy*, 24, (2016), 111-126.

349. Timmermann A. G., Forecast combinations, In: Elliott G, Granger C. W., Timmermann A. (eds) *Handbook of economic forecasting, Elsevier*, Amsterdam, (2006), 135-19.

350. Tinyakova V. T., The new approaches in econometric research of financial markets. Distributed volatility, *Review of Applied Socio-Economic Research*, 4, (2012), 247-255. https://www.tralac.org/resources/our-resources/6198-brics-legal-texts-and-policy-documents.html.

351. Todorovic P. and Zelenhasic E., A stochastic model for flood analysis, *Water resources research*, 6, (1970), 1641-1648.

352. Uppal J. Y., Measures of extreme loss risk - An assessment of performance during the global financial crisis, *Journal of Accounting and Finance*, 13, (2013).

353. Vo X. V. and Daly K. J., European equity markets integration—implications for US investors, *Research in International Business and Finance*, 19, (2005), 155-170.

354. Wahba G., Bayesian "confidence intervals" for the cross-validated smoothing spline, *J. Roy. Statist. Soc. Ser. B*, 45, (1983), 133-150.

355. Wahba G., *Spline models for observational data*, Philadelphia Society for Industrial and Applied Mathematics, (1990).

356. Wang A., From Xiamen to Johannesburg: The role of the BRICS in global governance, *Summit Studies BRICS Research Group, BRICS Information Centre, University of Toronto*, viewed 17 June 2020, from http://www.brics.utoronto .ca/analysis/xiamen-johannesburg-event.html, (2018).

357. Wang W., Van Gelder P. H. A. J. M., Vrijling J. K. and Ma J., Testing and modelling autoregressive conditional heteroskedasticity of streamflow processes. *Nonlinear process. Geophys. European Geosciences Union*, 12, (2005), 55-66.

358. Warshaw E., Extreme dependence and risk spillovers across north American equity markets, *The North American Journal of Economics and Finance, Elsevier*, 47, (2019), 237-251.

359. Weissman I., Estimation of parameters and large quantiles based on the k largestobservations, *Journal of the American Statistical Association*, 73, (1978), 812-815.

360. Wickham H., ggplot2 elegant graphics for data analysis, Springer, (2016).

361. Wilson D. and Puroshothaman R., Dreaming with BRICs: The path to 2050, Global Economic Paper, n. 99, *Goldman Sachs*, New York, (2003).

362. Wójtowicz T. and Gurgul H., Long memory of volatility measures in time series, viewed 20 July 2020, in *ResearchGate*, (2009).

363. Worthington A. and Higgs H., Transmission of equity returns and volatility in Asian developed and emerging markets: a multivariate GARCH analysis, *Int J Finan Econ*, 9, (2004), 71-80.

364. Xia Q., Liang R. and Liu J., A Bayesian analysis of autoregressive models with exogenous variables and power-transformed and threshold GARCH errors, *Communications in Statistics—Theory and Methods*, 44, (2015), 1967-1980.

365. Yang H., Extreme value mixture modelling with simulation study and applications in finance and insurance, viewed 28 September 2021, in https://ir.canterbury.ac.nz/handle/10092/8538, *UC research repository*, (2013), 1-113.

366. Yavas B. F., Benefits of international portfolio diversification, *Graziadio Business Report*, 10, (2007), viewed 14 December 2020, from https://gbr.pepperdine.edu/2010/08/benefits-of-international-portfolio-diversification/, (2020).

367. Yu J. and Hassan M. K., Global and regional integration of the Middle East and North African (MENA) stock markets, *The Quarterly Review of Economics and Finance*, 48, (2006), 482-504.

368. Zakoian J. M., Threshold heteroscedastic models, *Journal of Economic Research*, 4, (2012), 247-255.

369. Zhang J., Zhang D., Wang J. and Zhang Y., Volatility spill-overs between equity and bond markets: Evidence from G7 and BRICS, *Romanian Journal of Economic Forecasting*, 4, (2013).

370. Zonouzi S. J. M., Mansourfar G. and Azar F. B., Benefits of international portfolio diversification implication of the Middle Eastern oil-producing countries, *International Journal of Islamic and Middle Eastern Finance and Management*, 7, (2014), 457-472.

Appendix

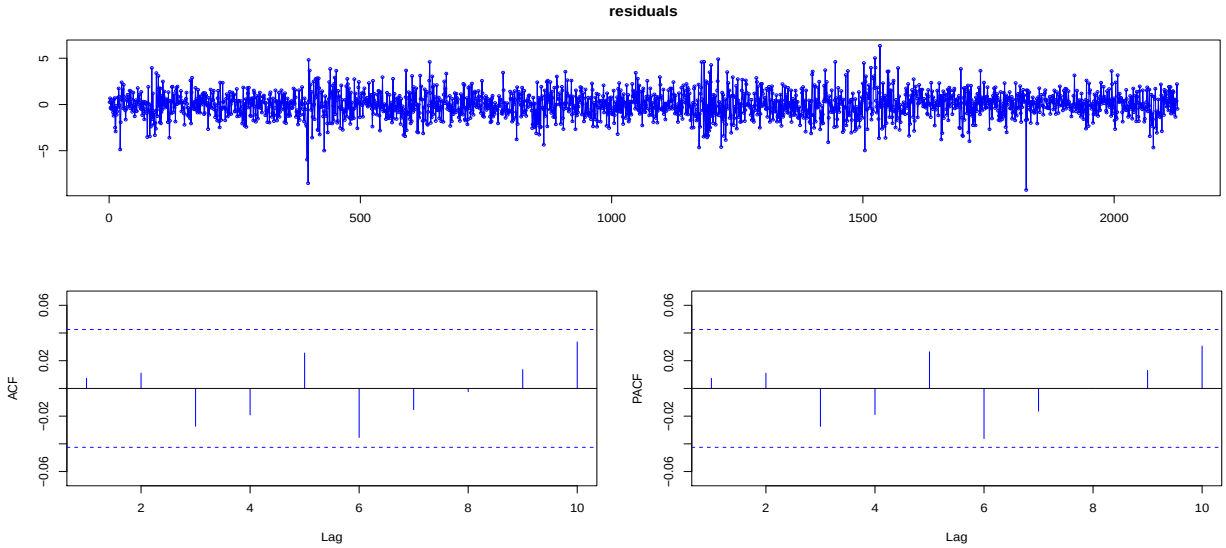


Figure 7.1: ACF and PACF of residuals for Bovespa index (Brazil).

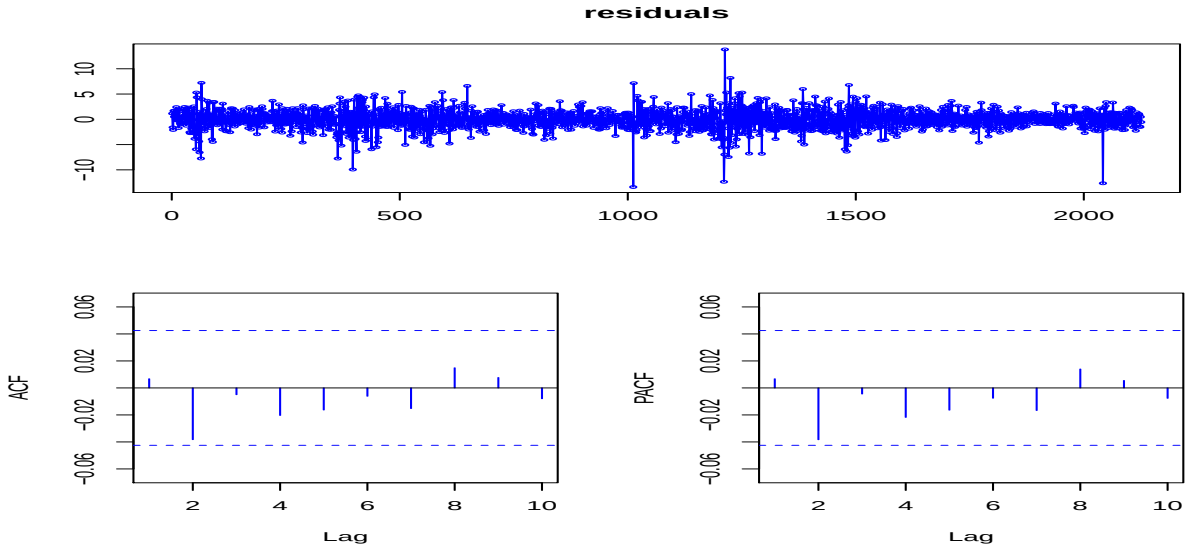


Figure 7.2: ACF and PACF of residuals for IMOEX index (Russia).

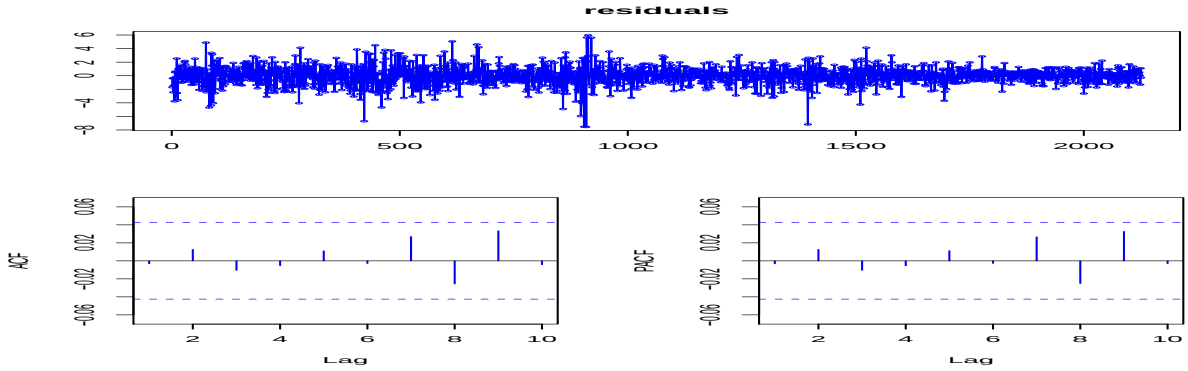


Figure 7.3: ACF and PACF of residuals for NIFTY index (India).

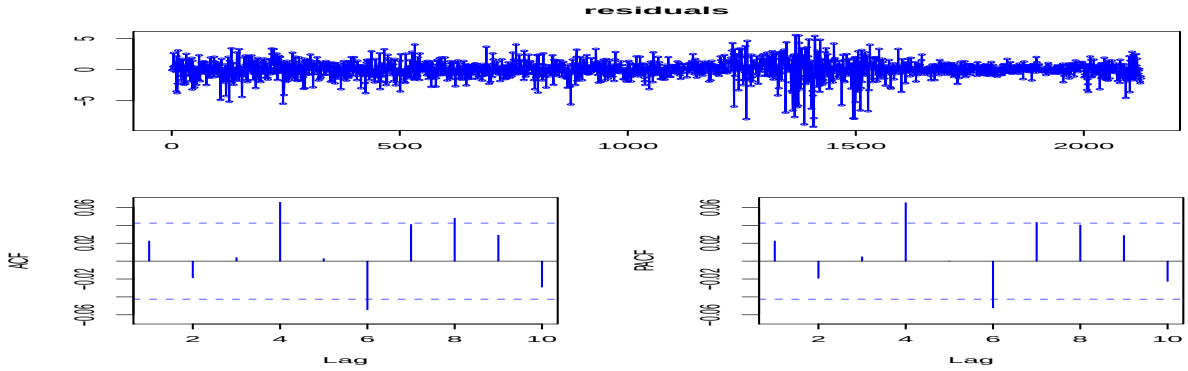


Figure 7.4: ACF and PACF of residuals for SHCOMP index (China).

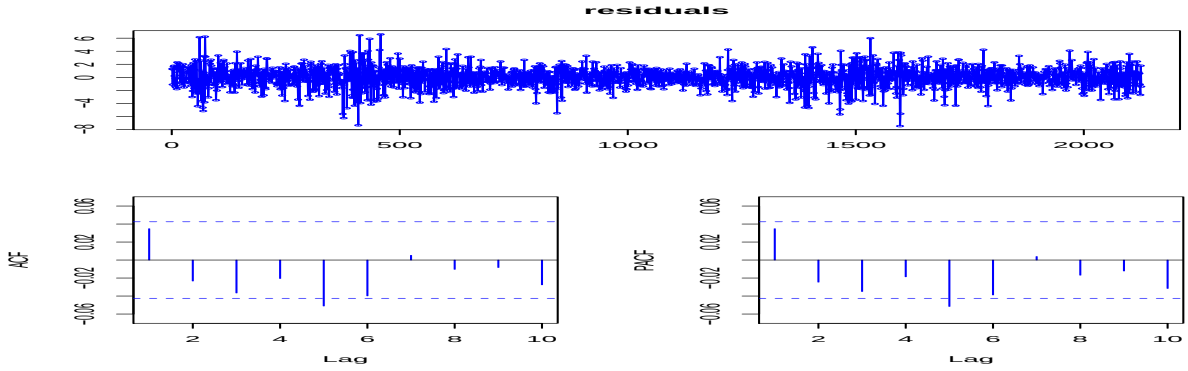


Figure 7.5: ACF and PACF of residuals for JALSH index (South Africa).

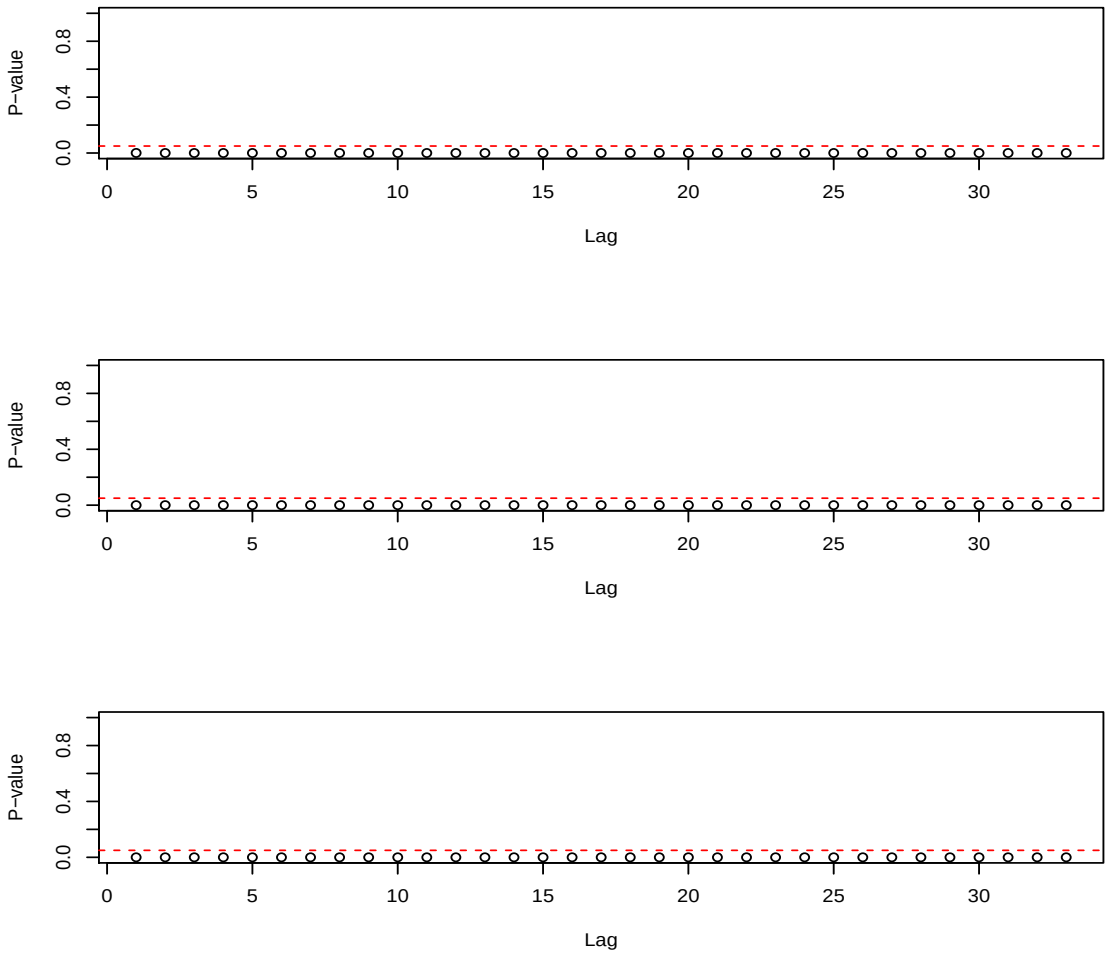


Figure 7.6: McLeod-Li test statistic for the returns of Bovespa (top panel), IMOEX (middle panel) and NIFTY (bottom panel) indices

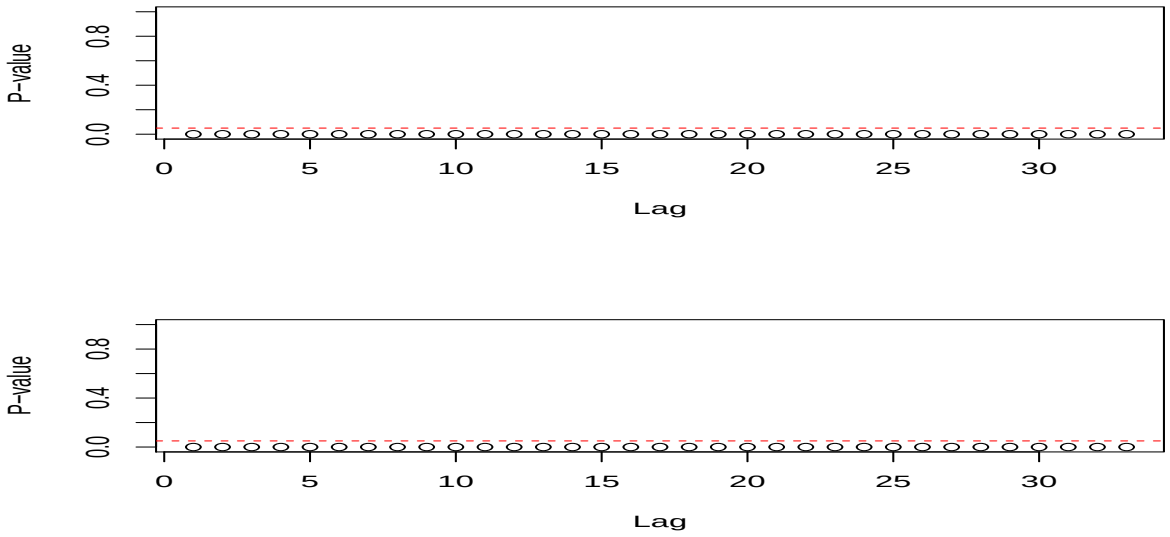


Figure 7.7: McLeod-Li test statistic for the returns of SHCOMP (top panel) and JALSH indices (bottom panel)

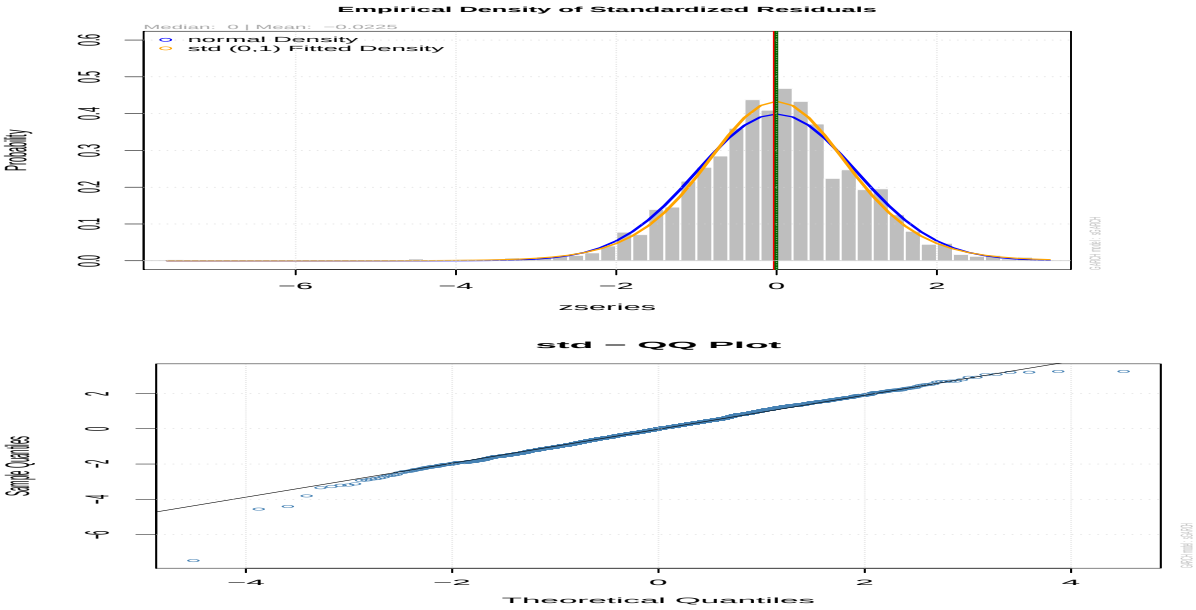


Figure 7.8: (a) Top panel: Return series with student's t and normal densities (b) Bottom panel: Student's t QQ plot for Bovespa index (Brazil).

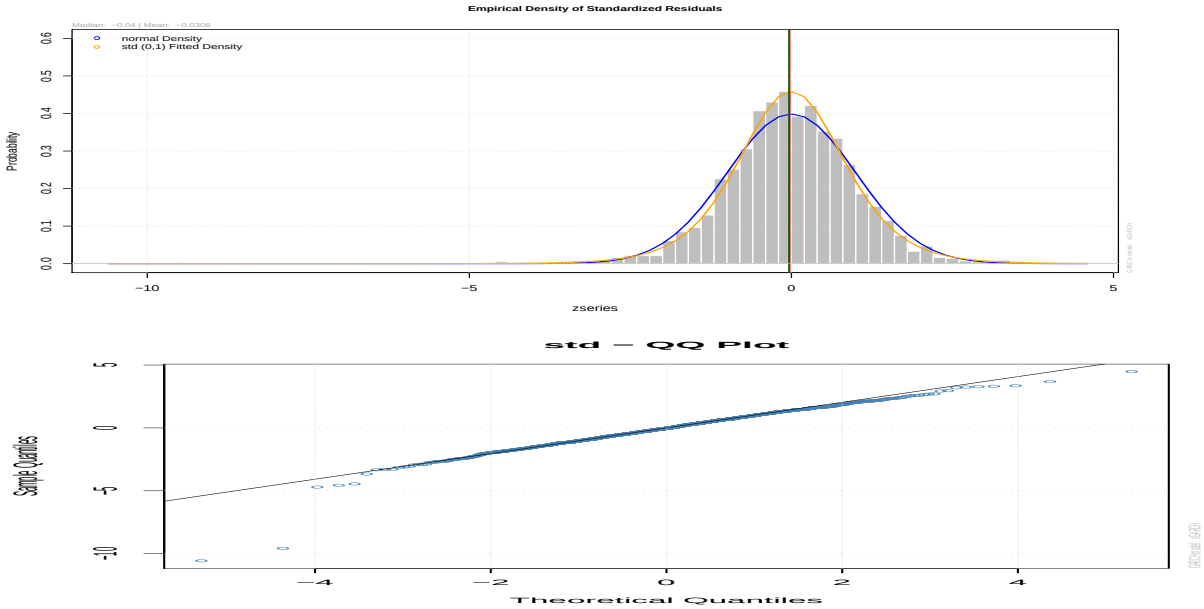


Figure 7.9: (a) Top panel: Return series with Student's t and normal densities (b) Bottom panel: Student's t QQ plot for IMOEX index (Russia).

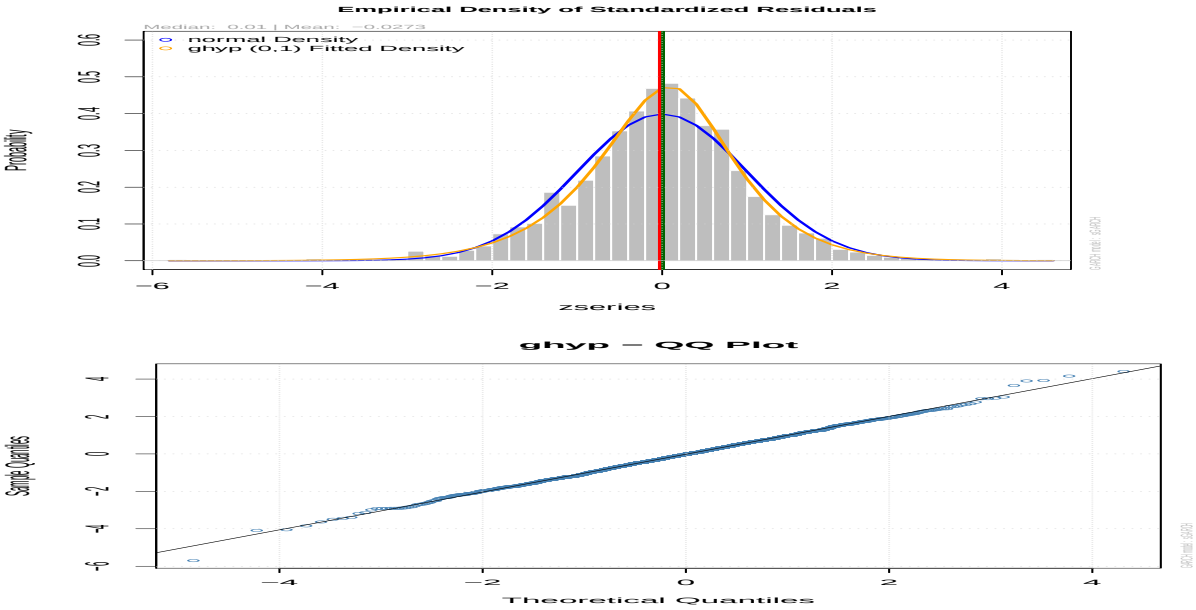


Figure 7.10: (a) Top panel: Return series with GHYP and normal densities (b) Bottom panel: GHYP QQ plot for NIFTY index (India).

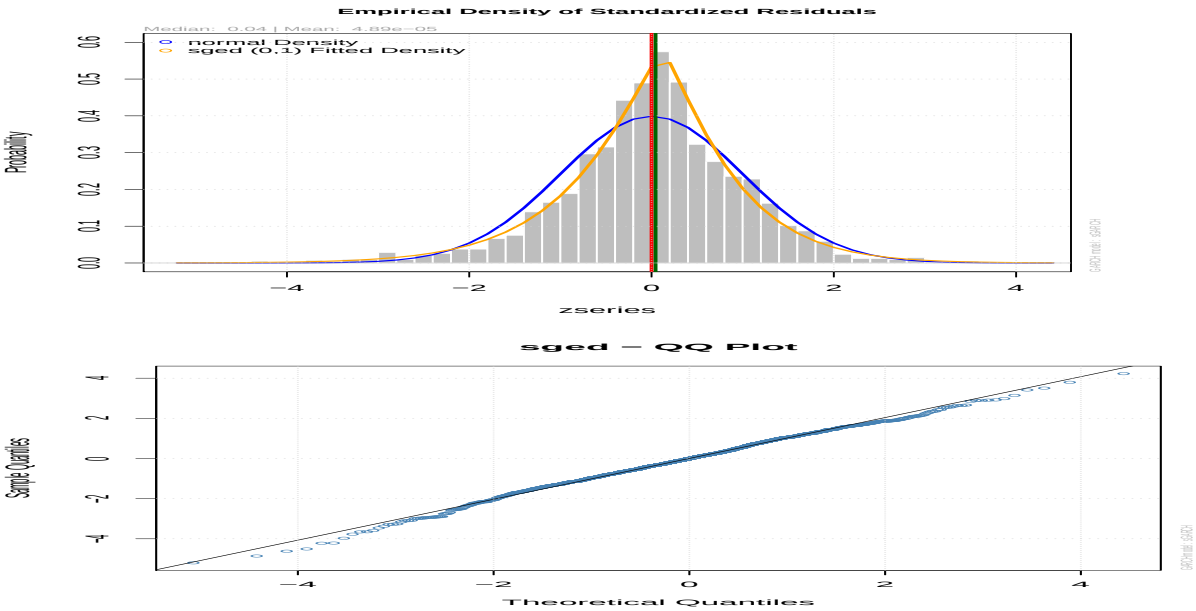


Figure 7.11: (a) Top panel: Return series with skew-GED and normal densities (b) Bottom panel: Skew-GED QQ plot for SHCOMP index (China).

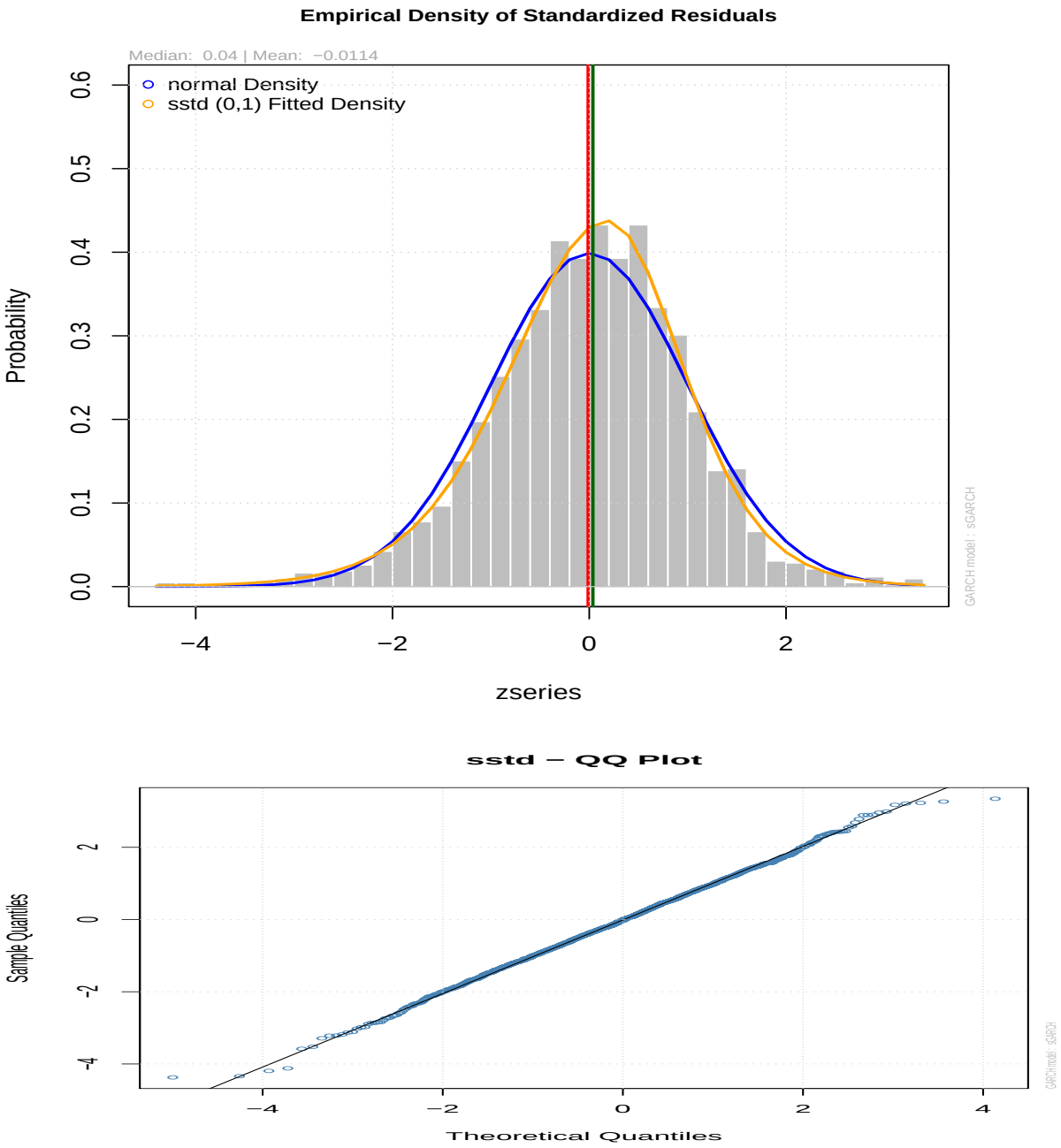


Figure 7.12: (a) Top panel: Return series with skew-student's t and normal densities (b) Bottom panel: Skew-student's t QQ plot for JALSH index (South Africa).

Volatility Persistence Diagrams

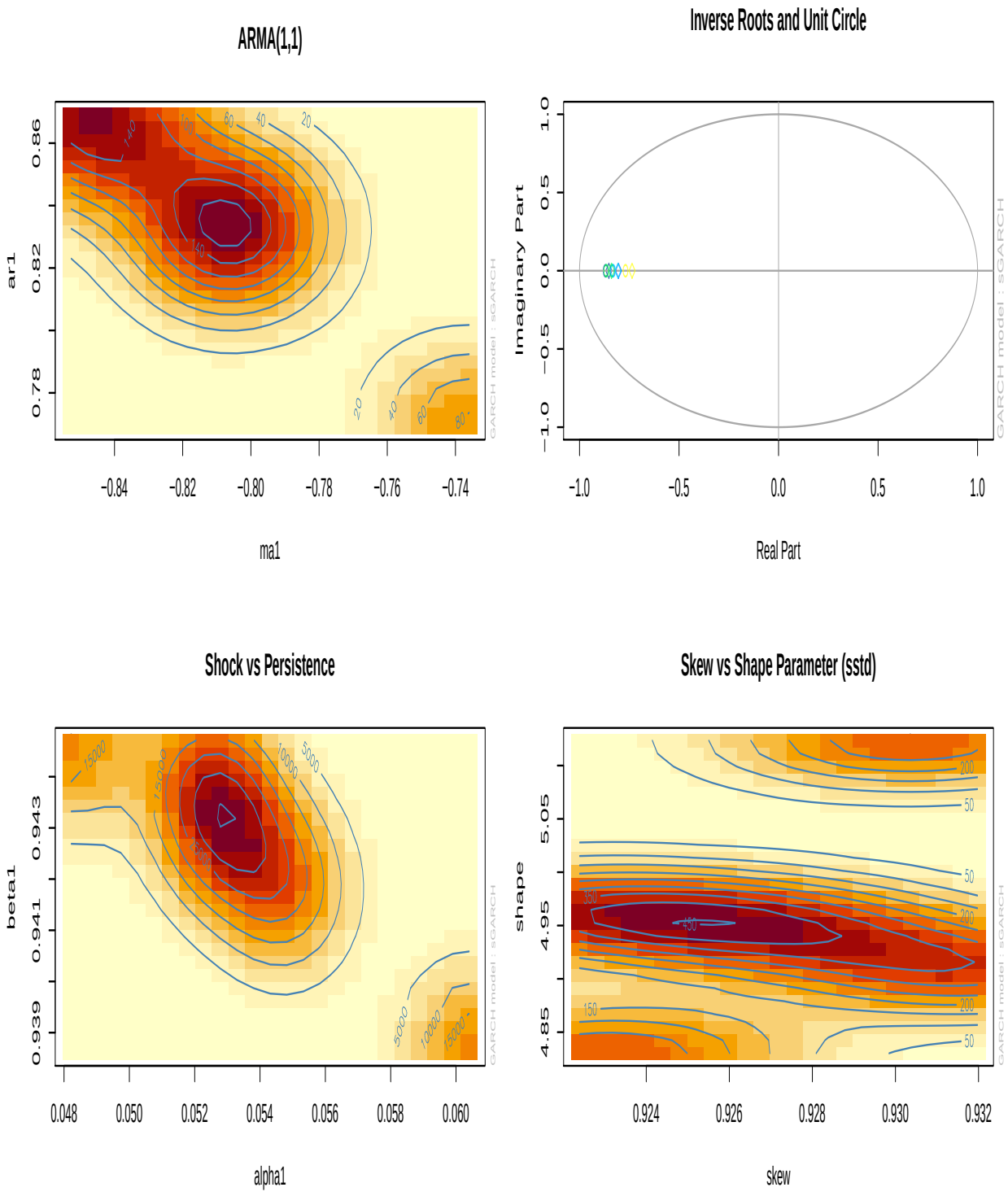


Figure 7.13: Volatility persistence plot: Chinese SHCOMP index.

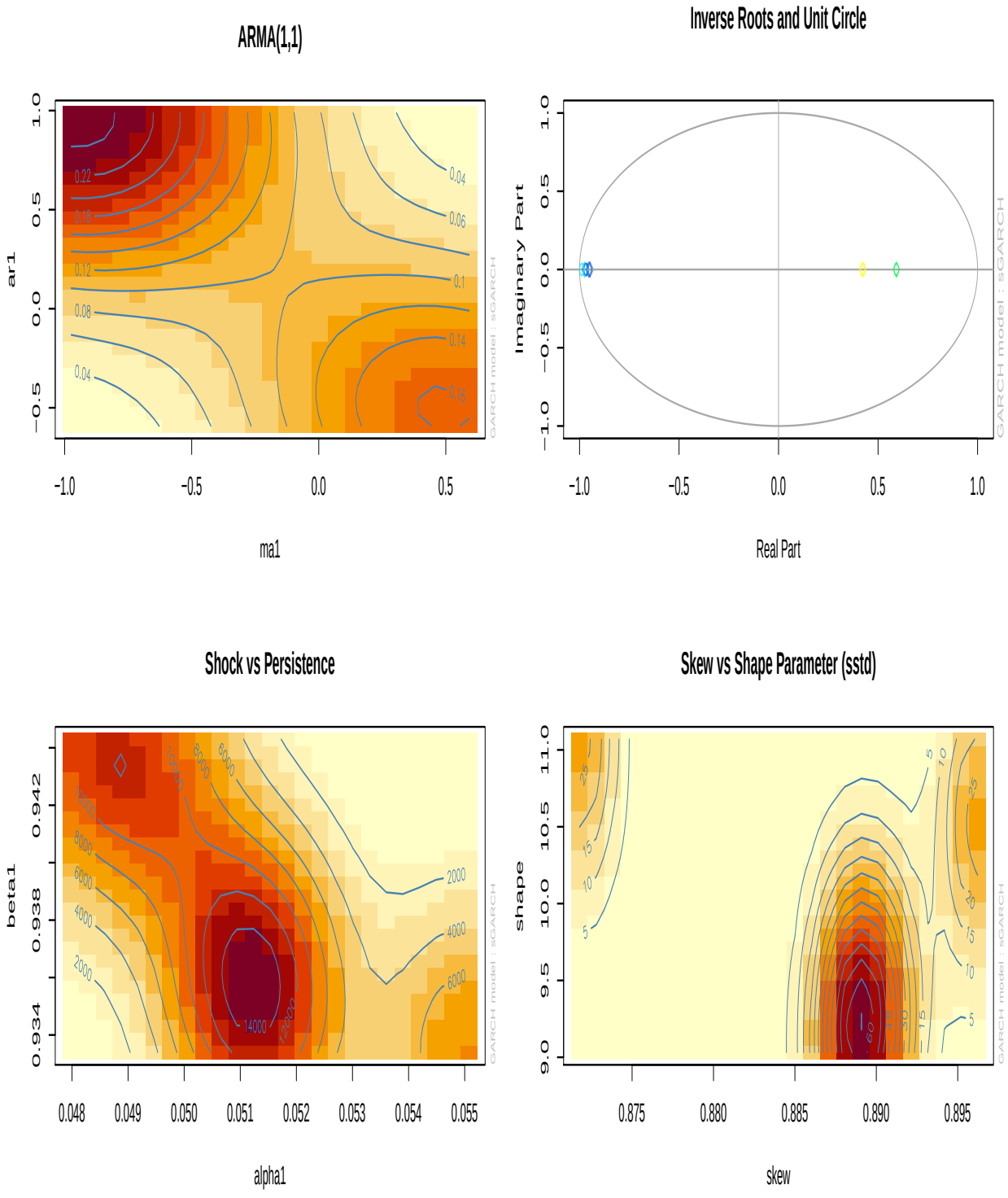


Figure 7.14: Volatility persistence plot: South African JALSH index.

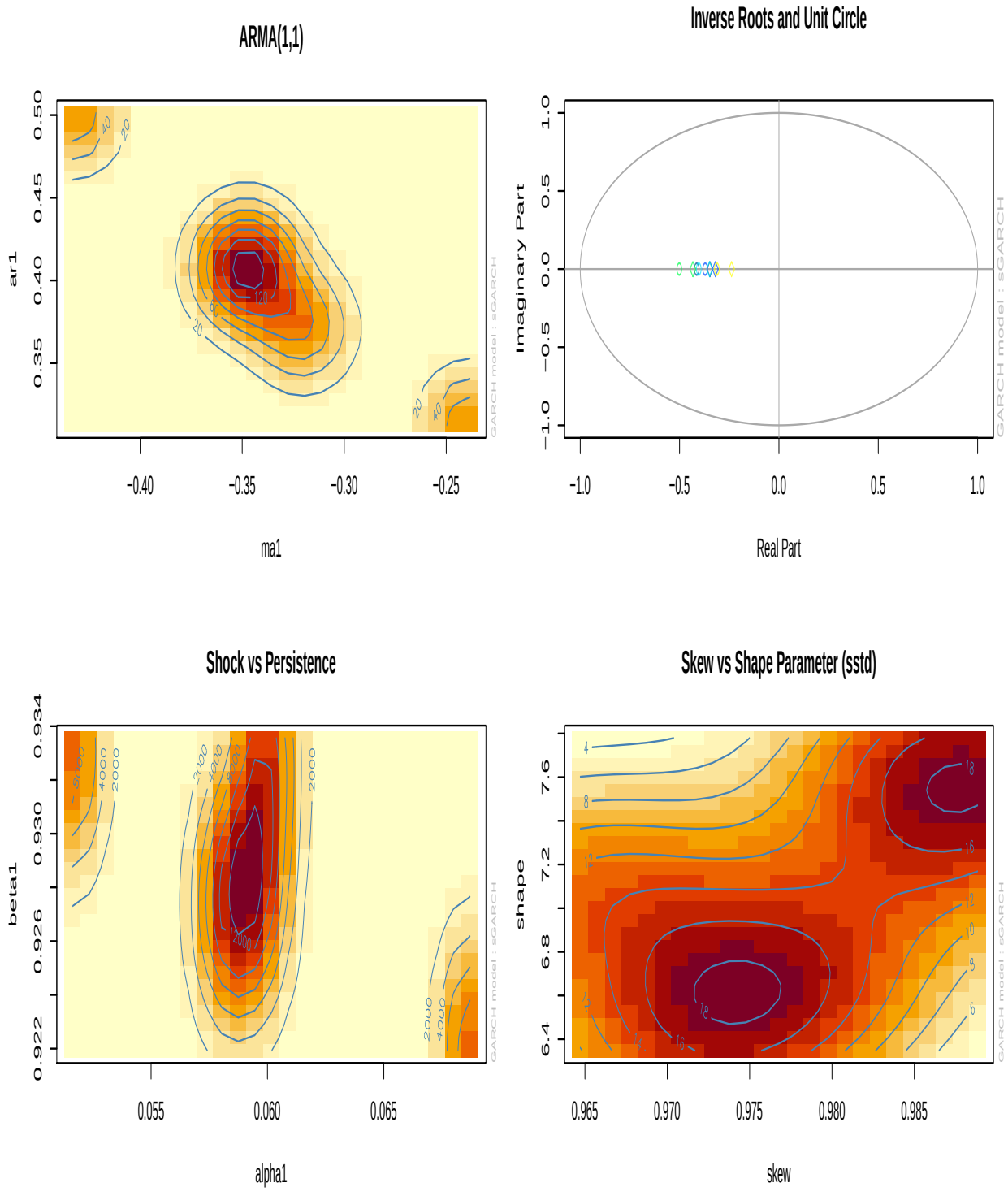


Figure 7.15: Volatility persistence plot: Russian IMOEX index.

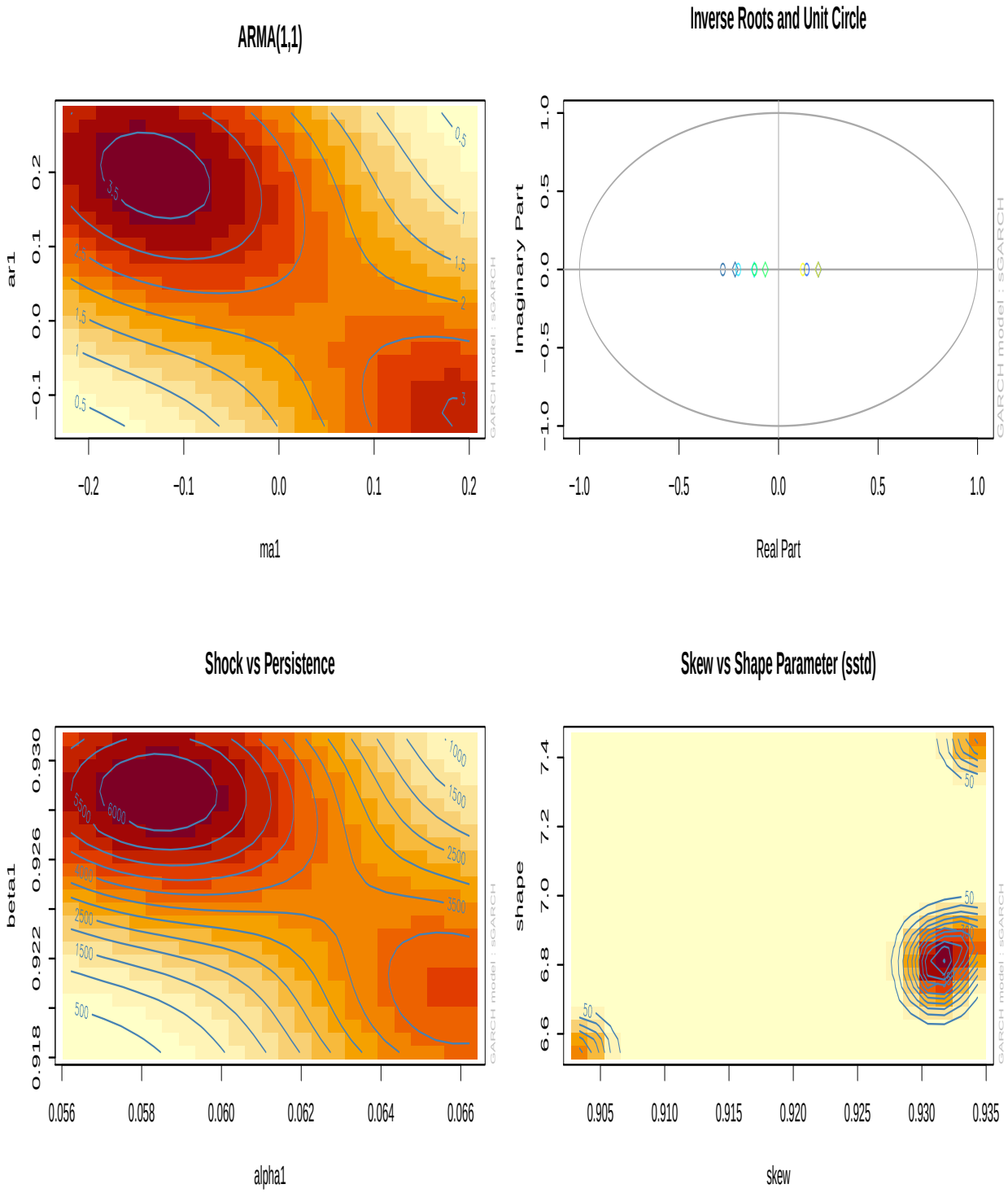


Figure 7.16: Volatility persistence plot: Indian NIFTY index.

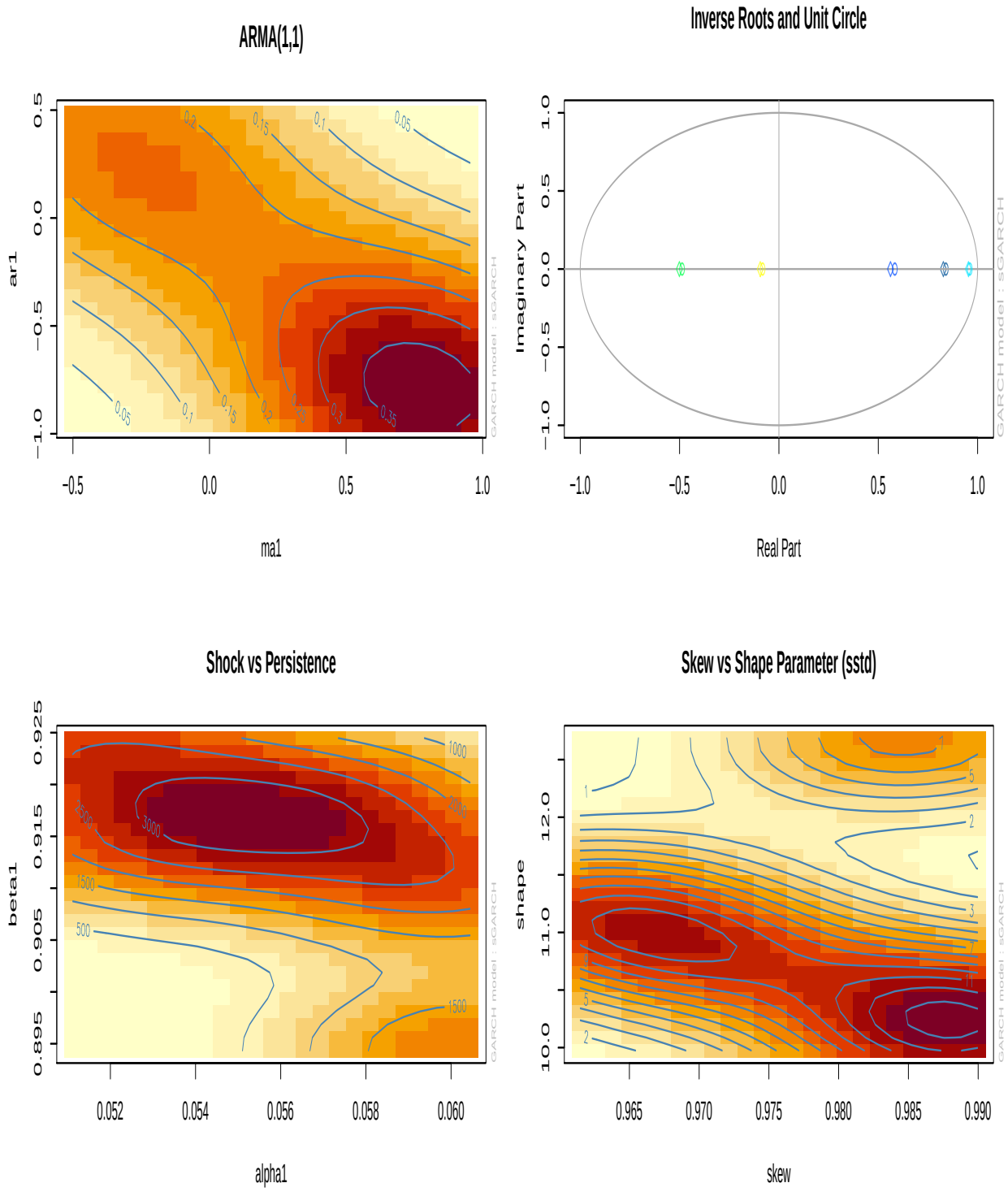


Figure 7.17: Volatility persistence plot: Brazilian Bovespa index.

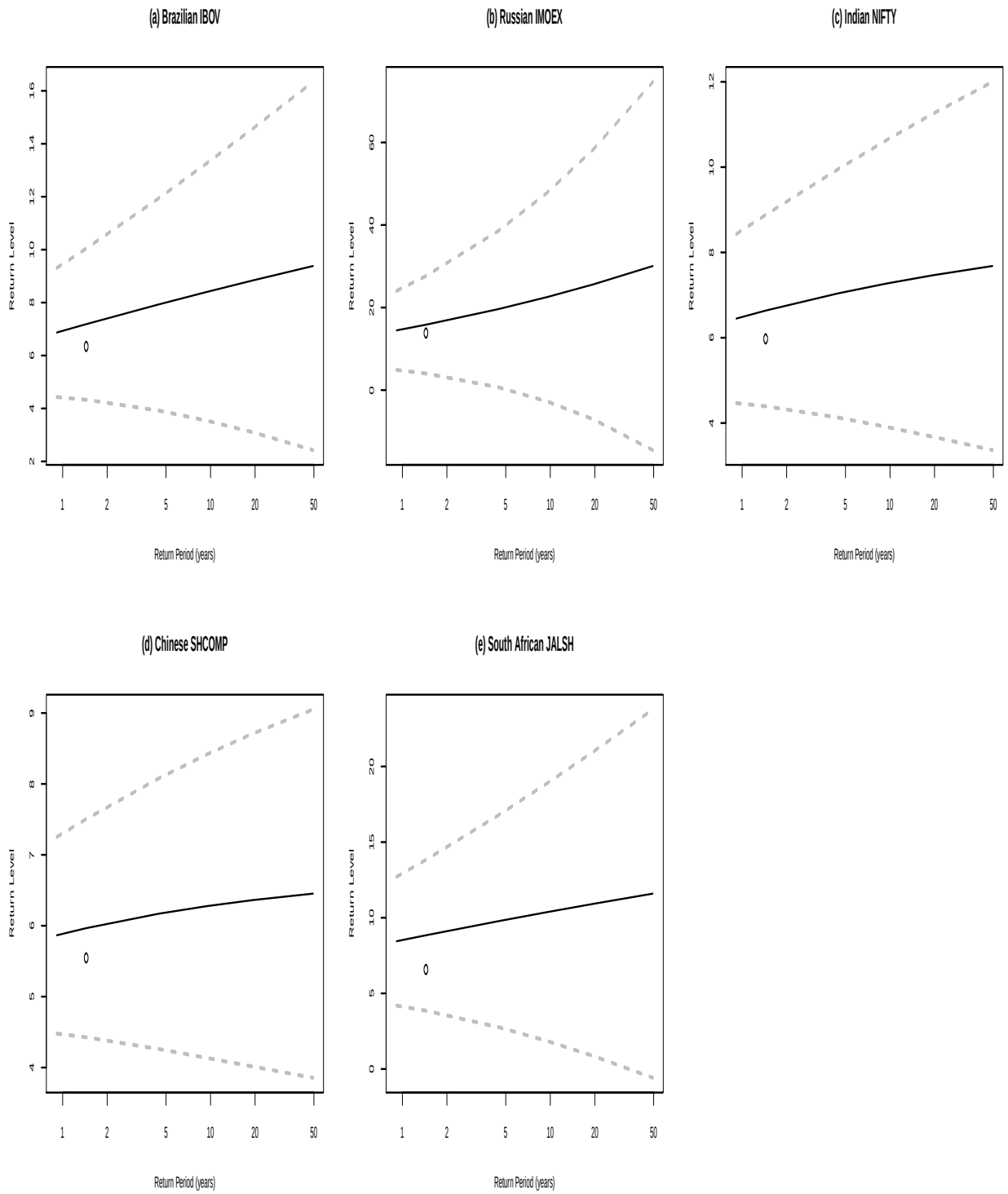


Figure 7.18: The return levels confidence intervals.

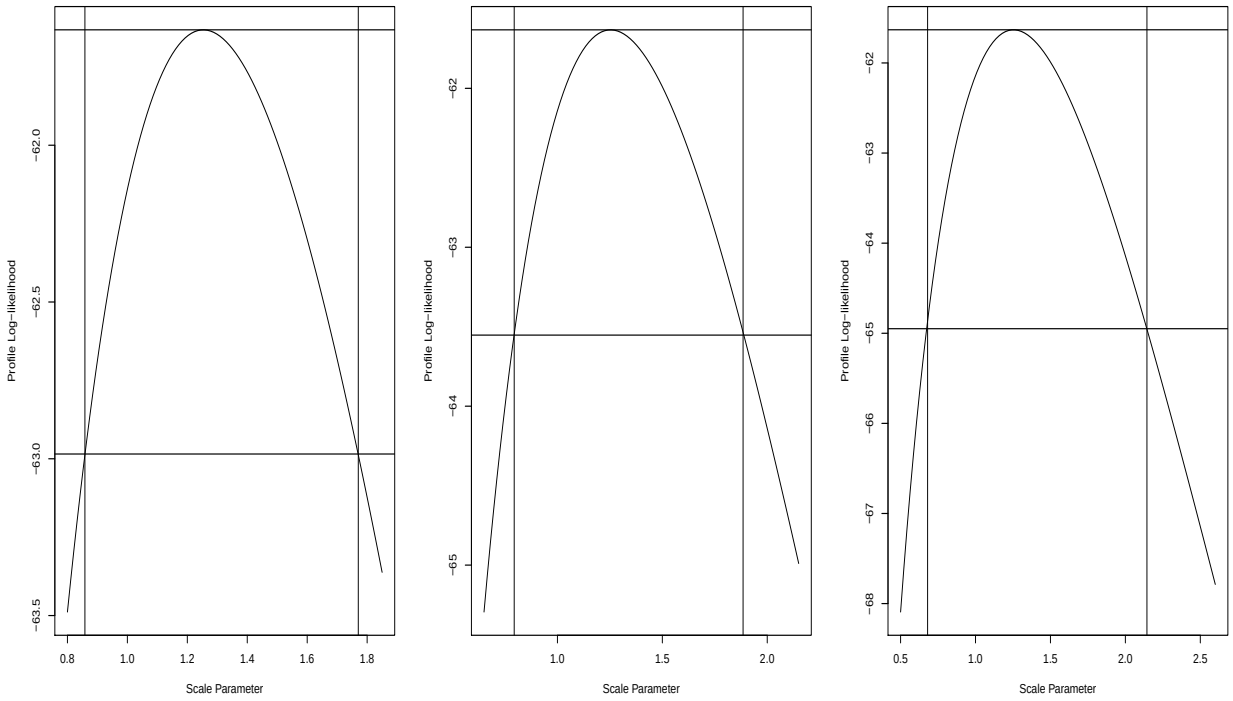


Figure 7.19: Profile log-likelihood intervals of IMOEX scale parameter σ.

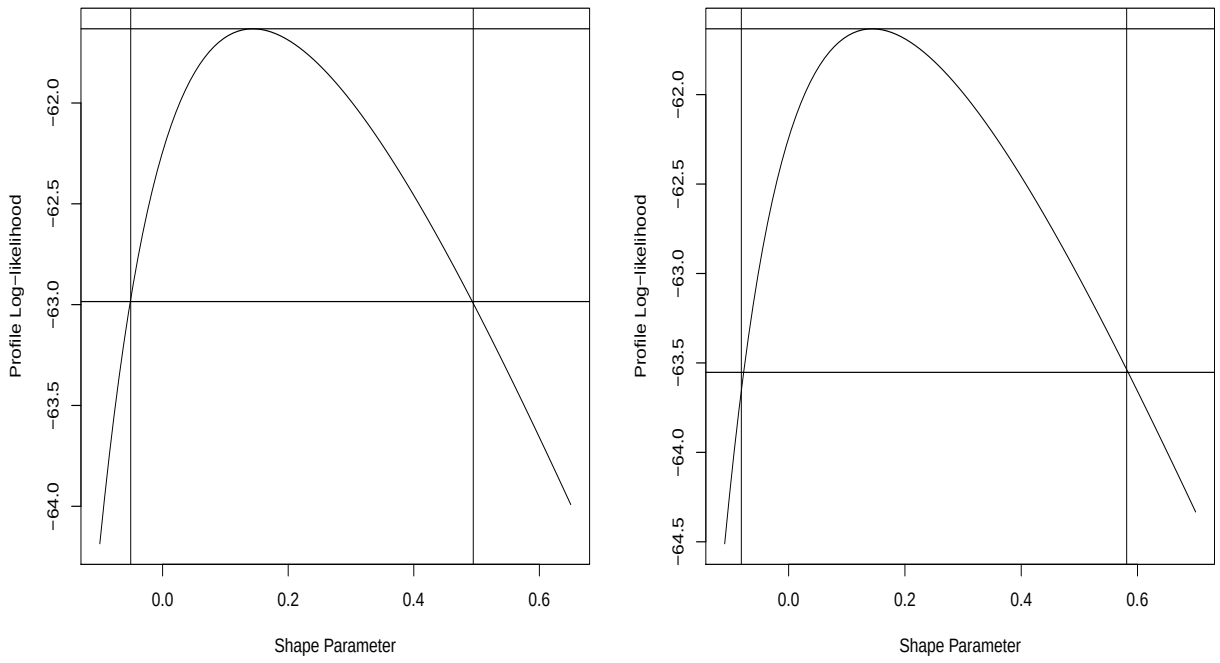


Figure 7.20: Profile log-likelihood intervals of IMOEX shape parameter ξ.