Erasmus University Rotterdam
Master's Thesis
2016-2017

Does one size fit all when forecasting market risk on country indices with heterogeneous market classifications?

An analysis under current and future regulations of the Basel Committee on Banking Supervision.

DAVID H. ROSENBRAND

Student Number - 359210

Thesis Supervisor – J. Zhao

Date of Submission: August 21, 2017

ABSTRACT

For the preceding decade, the increasing level of globalization has induced investors to look for new international diversification opportunities within the environment of Emerging markets and, especially, Frontier markets. However, an extensive field of empirical literature has documented that these two market classification groups exhibit unique peculiarities in terms of return dynamics when compared to Developed markets. In response, this paper investigated whether a 'one-size fits all' approach is justified when forecasting market risk on country indices with heterogeneous market classifications. In order to achieve maximum relevance and to cope with both current and future regulations of the Basel Committee on Banking Supervision, this paper forecasted series of both Value at Risk (99%) and Expected Shortfall (97.5%) for 30 country indices. The study included a wide spectrum of GARCH family models, multiple error distributions, and the innovative hybrid approach of conditional EVT. Overall, the results provided strong evidence that especially the extreme and unique characteristics belonging to Frontier markets ask for a different approach when forecasting risk. Risk models that work well for both Developed and Emerging markets are shown to be a likely victim of severe risk underestimation in Frontier markets. Therefore, the major implication of this study would be that risk practitioners should acquire a deep understanding of each specific market before considering a 'one-size fits all' approach regarding market risk quantification.

Keywords: Financial Risk Measurement, Market Classifications, VaR, ES

JEL: G24, G31, C22, C52, C53

PREFACE

This master's thesis entitled "Does one size fit all when forecasting financial market risk on country indices with heterogeneous market classifications?" is the final project in obtaining my master's degree in *Economics and Business* from the Erasmus University Rotterdam. From December 2015 to August 2017, I have conducted research on the process of quantifying financial market risk on country indices with different market classifications. The concept of this research project originated in Singapore, where I followed a financial internship at a multinational firm that had risk exposure to many different countries. Once back home in the Netherlands, I spent many hours on gathering all necessary data and on learning how to write scripts in R for the required calculations and analyses. Though occasionally with some reluctance due to the complexity and the time-consuming character of the subject, the making of this thesis was enjoyable, interesting and, in particular, very educational to me.

Reading my thesis will give you a better understanding of the unique dynamics exhibited by different markets and, additionally, to which extent these dynamics influence the way market risk quantification should be performed. In fact, I provide evidence that countries with a Frontier market classification should be treated differently than Developed and Emerging markets, since neglecting their differences could lead to severe risk underestimation. Hence, the findings from my study are of particular interest to investors who want to invest in alternative markets and to risk practitioners who need to perform risk measurement on country indices with heterogeneous market classifications.

I would like to take this opportunity to gratefully and sincerely thank Mrs. J. Zhao for her understanding, flexibility and patience over the course of this project. Furthermore, I would like to express my gratitude to my family and my best friend Lars, as their unceasing encouragement, support and patience were undeniably the bedrock upon which this thesis has been built. Without them, it would have been impossible for me to be where I am now.

David H. Rosenbrand

Oud-Beijerland
August 21, 2017

NON-PLAGIARISM STATEMENT

By submitting this thesis the author declares to have written this thesis completely by himself/herself, and not to have used sources or resources other than the ones mentioned. All sources used, quotes and citations that were literally taken from publications, or that were in close accordance with the meaning of those publications, are indicated as such.

ABBREVIATIONS

ADF – Augmented Dickey Fuller

AIC – Akaike Information Criterion

AR – Autoregressive

ARCH – Autoregressive Conditional Heteroscedasticity

ARMA – Autoregressive Moving-Average

BCBS – Basel Committee on Banking Supervision

BIC – Bayesian Information Criterion

BM – Block Maxima

CC – Conditional Coverage

DUR – Duration

ES – Expected Shortfall

EGARCH – Exponential Generalized Autoregressive Conditional Heteroscedasticity

EVT – Extreme Value Theory

GARCH – Generalized Autoregressive Conditional Heteroscedasticity

GEV – Generalized Extreme Value

GJR GARCH – Glosten-Jagannathan-Runkle Generalized Autoregressive Conditional Heteroscedasticity

GPD – Generalized Pareto Distribution

IGARCH – Integrated Generalized Autoregressive Conditional Heteroscedasticity

IID – Independently and Identically Distributed

IND – Independence

JSU – Johnson System Unbounded

LM – Lagrange Multiplier

MA – Moving-Average

MENA – Middle East and North Africa

MSCI – Morgan Stanley Capital International

NAGARCH – Nonlinear Asymmetric Generalized Autoregressive Conditional Heteroscedasticity

POT – Peaks over Threshold

S&P – Standard and Poor's Financial Services

SEE - South Eastern Europa

TGARCH – Threshold Generalized Autoregressive Conditional Heteroscedasticity

UC – Unconditional Coverage

USA – United States of America

VaR – Value at Risk

WAEMU – West African Economic and Monetary Union

TABLE OF CONTENTS

(1) INTRODUCTION

1.1 The Framework of Market Risk Quantification

Over the course of financial history, various financial disasters have repeatedly emphasized the importance for financial institutions to have reliable risk assessment tools in order to quantify their risk exposure accurately. The most commonly reported measure of risk is the concept of Value at Risk (henceforth abbreviated as VaR), which can be defined as the maximum loss over a given time horizon at a given level of confidence (Jorion, 2001). It has been the industry-standard risk measure since the early 1990s, and was accepted by the Basel Committee on Banking Supervision (BCBS) as the basis for their regulation in 1996. Nonetheless, the reputation of VaR has been damaged severely due to the events sparked by the bankruptcy of Lehman Brothers in 2007, where the following global financial crisis painfully reminded banks and regulators alike that stress markets are clearly capable of producing losses far in excess of the amounts forecasted by VaR. By its definition, VaR ignores the statistical properties of the significant loss beyond the threshold. Therefore, the BCBS (2016) deemed it necessary to publish a complete revision of their market risk framework, in which they announced that banks will be enforced to replace the well-established VaR with a relatively new risk measure called Expected Shortfall (henceforth abbreviated as ES) by December 31, 2019. The latter measure can be defined as the expected value of loss given the exceedance, and according to many academics it should be able to tackle the dangerous drawback of VaR of not taking into account the severity of the incurred losses (Yamai & Yoshibita, 2005).

Nonetheless, whether it is Value at Risk or Expected Shortfall that needs to be estimated, in both contexts a precise prediction of the probability of an extreme movement in the value of a portfolio is a matter of utmost importance. As the field of financial econometrics has delivered a large quantity of sophisticated models, the responsibility of VaR/ES users lays in picking the one that is capable of making estimations that fit the future distribution of returns properly. In situations where VaR or ES is underestimated, a similar scenario could emerge as during the global financial crisis of 2007-2008, where the capital of banks was insufficiently constituted to provide coverage for the unexpected adverse events. Ultimately, the shortage was of such an extent that many financial institutions had to be bailed out by governments with billions of dollars, highlighting the significance of developing a precise VaR/ES model.

However, the process of calculating accurate VaR/ES estimations is hampered by the fact that financial returns tend to exhibit "non-standard" statistical properties such as non-normality. Starting with the introduction of the risk model called RiskMetrics™ in 1993, the powerhouse of Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models, with all its numerous extensions, has been a widely applied methodology for VaR/ES calculation. Its popularity stems from its proclaimed ability of capturing many of the salient stylized facts exhibited by financial return series. However, in spite of its popularity and claimed effectiveness, a noted downside of the GARCH family framework should be its naïve focus on the whole distribution, while VaR/ES is in essence only related to the tails of the probability distribution. So, instead of forcing a single distribution on the entire return series, one might be better off by using some kind of limit law to investigate the tails of the returns only. Therefore, Embrechts et al. (1997) and Reiss & Thomas (1997) both suggest the implementation of Extreme Value Theory (henceforth abbreviated as EVT). They highlight the method's capability of offering a parametric form for only the tail, which allows the user to model rare and extreme phenomena that lie outside the range of available observations. They argue that EVT is providing the means to obtain accurate risk estimates that are true to the empirical fat tailed behavior of the underlying distribution. However, McNeil & Frey (2000) argue that applying EVT directly upon raw financial returns would be inappropriate, as these are in practice not independently and identically distributed. To overcome this shortcoming, they propose an innovative hybrid method called conditional EVT, where EVT is combined with the GARCH family framework.

1.2 The Evolving Field of Diversification

In modern finance, it is a well-accepted belief that investors should not put "all of their eggs in one basket", meaning that they should strive for diversification instead, where wealth is allocated among different assets. The economist Harry Markowitz (1952) illustrates that when investors are forming such diversified portfolios, they should model return correlations of the underlying assets, as low return correlations would significantly reduce the risk of the overall portfolio without sacrificing the investors' presumed goal of return maximization.

A large body of literature on international portfolio diversification has documented that during the 1980s, the financial integration and interdependence among major industrialized markets increased greatly, leading to an erosion of their practicality within a diversification framework. Eventually, it led portfolio managers decide to revamp their asset allocation strategies and to start a search for new diversification opportunities in order to maintain maximum return per unit of risk.

Errunza (1983) was among one of the first academics that studied the potential attractiveness of Emerging markets within the framework of international portfolio diversification as a substitution to Developed markets. Emerging markets are the countries that have some characteristics of a Developed market, but do not meet the standards to be a Developed market; this includes both countries that may become Developed in the future as well as those that were Developed in the past. The study of Errunza (1983) reveals low return correlations between Emerging and Developed markets, which led to the believe that Emerging markets could offer great opportunities for portfolio risk reduction. However, over the last two decades it has become a key subject in asset allocation literature whether these claimed international diversification benefits of using the traditional Emerging markets are still feasible in a world of increasing globalization (Demirguc-Kunt & Levine, 1996; Bekaert & Harvey, 2003). This concern has only increased in magnitude after the global financial crisis of 2007-2008 had its devastating impact on both Developed and Emerging markets. At the end, these dangerous developments induced investors to seek for new 'save havens' that should be more resistant by nature against a possible new global financial crisis.

This urge of finding a new investment platform created a huge increase of interest in countries belonging to the Frontier market classification. These markets are officially defined as pre-Emerging markets, and they describe the smallest, least developed, less liquid markets that make up Emerging markets. Frontier markets' potential as an investment opportunity sparked in 2007 with the creation of two Frontier market indices by Standard and Poor's Financial Services (S&P) and one by Morgan Stanley Capital International (MSCI). Prayag et al. (2010) document that Frontier markets are providing international investors with the same benefits as those initially delivered by Emerging markets. There are substantial economic and investment rationales in considering the allocation of funds towards these markets, as they offer strong growth opportunities and have maintained consistent low return correlations with both Developed and Emerging markets (Speidell & Krohne, 2007; Berger et al., 2011).

1.3 The Research Question

All of the above suggests that a well-diversified portfolio should comprise assets originating from not only Developed markets, but also from Emerging and, especially, Frontier markets. However, when one wants to quantify the risk belonging to such diversified portfolios, it would be a misnomer to view the underlying markets as being completely homogeneous.

A significant part of literature has been devoted to the systematic differences between Developed markets and Emerging markets. In general, they document that Emerging markets are subject to different cultural, institutional, economic, and political circumstances. The environment of Emerging markets is best characterized by a lower liquidity, a higher level of country risk, a higher level of insider trading, a higher fluctuation of foreign exchange rates, more frequent changes in credit rating, and more frequent financial shocks. Concerning market risk quantification, the lower level of liquidity present in Emerging markets causes them to exhibit lower volatility in general, but the frequent financial shocks are giving rise to substantial extreme volatility (Božović & Totić, 2015). Similar findings are documented by Gençay & Selçuk (2004), who state that Emerging markets are more likely to exhibit multiple regime switches in short periods of time, which opens the possibility that the parameters of the return distribution change quite often. Furthermore, they argue that Emerging markets experience larger financial earthquakes than Developed markets, and that they should be labeled as 'markets with many fault lines'. Aggarwal et al. (1999) examined the events that cause major shifts in Emerging markets' volatility, and they find that, unlike Developed markets, large changes in volatility are often linked to country-specific events. Overall, the literature provides evidence that Emerging markets are in need of dynamic and flexible econometric models to capture the more complex and more volatile market conditions.

At the same time, the dynamics encountered in Frontier markets are presumed to be even more complex than the ones found in Emerging markets. Issues that plague many Emerging markets, such as political instability, weak regulation, and illiquidity, are amplified in many of the existing Frontier markets. By definition, the Frontier markets are a class described by small, illiquid, less accessible, and less known Emerging markets (Bello & Adedokun, 2011). Due to their illiquidity, the equity prices tend to respond more strongly to selling and trading pressures during flight-to-safety episodes where investors try to reduce their exposure quickly (Chan-Lau, 2011).

One would expect that the different characteristics belonging to these three market classification groups (Developed, Emerging, and Frontier) influence the properties of the returns in a substantial degree and, therefore, influence the relative forecasting accuracy of the various market risk quantification frameworks across the market groups as well. To examine whether the markets' unique peculiarities indeed ask different approach in terms of risk forecasting, this Master's Thesis seeks to answer the following research question:

"Does one size fit all when forecasting market risk on country indices with heterogeneous market classifications?"

To the best of my knowledge, the existing empirical literature on this subject is bearing some severe limitations, implying that it is not yet capable of providing a legitimate and sufficient answer. This study tries to overcome these limitations and seeks to provide the financial world with the crucial and relevant means for performing adequate and robust market risk quantification when investing internationally. In the next subsection, several limitations of existing literature are addressed along with the proposed solutions.

1.4 Limitations of Current Literature

Limitation I: The first crucial drawback present in the existing literature would be the tendency to conduct research on indices from countries that are classified as Developed markets, leaving Emerging markets, and especially Frontier markets, largely untouched. This traditional reluctance is not only displayed by the overwhelming majority of the academic community, but also by the BCBS, who tends to purely submit Consultative Documents that are tailored towards the spheres of Developed markets. The occasional study of academics or regulators that does take into account Emerging or Frontier markets tends to address only one index or two indices, making it very dangerous to generalize their findings as being representative for Emerging or Frontier markets as a whole. At the same time, all studies implement different risk models, time periods and backtesting procedures, which would make it naïve and unjustified to bundle their isolated results for individual countries as an answer for the complete market classification group. Therefore, the current literature is only capable of providing a modest step into the right direction of adequate market risk quantification when dealing with alternative markets.

- ✓ On the contrary, this study includes a large number of Developed, Emerging and Frontier market indices in order to deliver a uniform but extensive comparison. Each market class is represented by ten market indices that are designed by Morgan Stanley Capital International (MSCI). The possibility of performing such an extensive research on Frontier markets is relatively new as return data availability for Frontier markets just started in 2005 and only by now delivers enough observations to conduct an econometric justified study. In order to produce the most relevant information, the study includes thirty country indices that (i.) have the highest stock market capitalization within their market classification group, (ii.) have return data availability since 01.12.2005 in the financial database DataStream, and (iii.) have consistent market classifications according to MSCI over the whole sample period, e.g. Argentina is only a Frontier market since May 2009 and should therefore not be included.

<u>Limitation II:</u> Regardless of the market under study, a prominent problem exhibited by current literature is that it is focused almost exclusively on Value at Risk, which has been the dominant risk measure for many years. However, over the last decade, the limitations of VaR have prompted the necessity of implementing an alternative measure of risk: Expected Shortfall. Considering the fact that ES will be the new standard of the BCBS regulation framework by December 31, 2019, one could argue that neglecting the concept would result in at least partly irrelevant or outdated results and conclusions.

- ✓ This study aims at filling this empirical void by examining the forecasting accuracy of market risk quantification frameworks from the perspective of a large number of indices under both the current BCBS regulative framework of VaR (99%) and the future BCBS regulative framework of ES (97.5%).

<u>Limitation III:</u> The main concern regarding the forecasting of market risk is the choice of the appropriate estimation model, e.g. an ill-suited model may turn out to be catastrophic for the risk taking vendor. Moreover, as the BCBS has not hitherto recommended a particular methodology for calculating risk, the process of finding the most appropriate model is still a matter of utmost importance. However, the majority of existing literature does not use the full potential that the field of financial econometrics has to offer. They tend to incorporate purely parametric models that work fine in areas of the empirical distribution where there are many observations, but these are known to work rather poor to the extreme tails of the empirical distribution. At the same time, many studies fail to incorporate important empirical concepts such as fat tails, skewness, and the leverage effect.

- ✓ In order to achieve relevant results, this study includes various GARCH family models (GARCH, IGARCH, EGARCH, GJR-GARCH, TGARCH, NAGARCH), multiple error distributions (Normal, Student-t, JSU), and the promising methodology of Conditional Extreme Value Theory.

Limitation IV: Another limitation of current literature would be the tendency to only incorporate backtesting procedures that do not allow for a statistical comparison between risk models. However, ranking competing models is crucial if one wants to draw hard conclusions on which type of model produces the most accurate estimations. At the same time, the majority of current research neglects the issue of conditional efficiency in VaR series, meaning that they ignore the possibility of clustered violations. VaR models that yield clustered violations may induce solvency issues for the risk taking vendors. In the case of ES, which is a relatively new measure of risk, the field of backtesting procedure is still in its infancy and the amount of tests available is low and relatively unknown.

- ✓ In the case of VaR, this study implements an innovative backtesting procedure that helps to distinguish statistically sound models from their inappropriate rivals and allows for a statistical comparison in terms of forecasting performance. Regarding ES, this study implements a bootstrap backtesting procedure that also allows for a statistical ranking between the competing forecasting models.

Limitation V: The final issue of current literature would be the tendency to make certain simplifications in the calculation process in order to reduce the computational effort. They often re-estimate the parameters on a basis that is higher than daily, which implies that their results exhibit less realistic and relevant outcomes.

- ✓ This study re-estimates each parameter on a daily basis for all 36 models and 30 countries for both VaR and ES. Regarding the approach of conditional EVT, this study implements an innovative procedure that automatically calculates the threshold value. The old-fashioned way would be to look at daily graphs and subjectively choose the correct value, which would naturally be impossible to achieve time-wise. To cope with the required processing power, several servers from the company Amazon Web Services are rented.

This study also exhibits the advantage of sharing the completely self-written script (written in the R software environment). This script contains several powerful menus and detailed guidelines that allow the reader to replicate the complete study or to apply the same methodology on a different group of assets of own choice. Appendix 2 also provides a detailed user manual on how the script could be adjusted in order to add new risk models and error distribution. This should make it even feasible for users with very limited programming knowledge.

1.5 Main Results

In order to answer the research question of this study thoroughly, the results have been divided into two main sections, each with their own null hypothesis and alternative hypothesis. The first part of the research has been dedicated towards the current regulation framework of the BCBS and incorporated the forecasting of Value at Risk at the 99 percent level of confidence. The second part of the research embodied the future regulation framework of the BCBS and forecasted Expected Shortfall at the 97.5 percent confidence level.

HYPOTHESIS ONE AND TWO	
$H0_1$	The underlying market classification of a country index is *unrelated* to the relative forecasting performance of $VaR_{99\%}$ models.
HA_1	The underlying market classification of a country index is *related* to the relative forecasting performance of $VaR_{99\%}$ models.
$H0_2$	The underlying market classification of a country index is *unrelated* to the relative forecasting performance of $ES_{97.5\%}$ models.
HA_2	The underlying market classification of a country index is *related* to the relative forecasting performance of $ES_{97.5\%}$ models.

The results from this study indicate that both null hypotheses need to be rejected in favor of their alternative hypotheses. The most striking results were observed within the environment of Frontier markets, where the extreme and unique peculiarities urged for a completely distinctive approach when performing financial risk management. This discrepancy was shown to be most prominent under the future BCBS regulation framework of Expected Shortfall, which suggests that risk practitioners should be even more careful when considering a homogeneous approach of risk management in the future. Overall, in terms of the research question, one could argue that the forecasting accuracy of market risk quantification frameworks differs substantially across the three specified market classification groups when forecasting both Value at Risk (99%) and Expected Shortfall (97.5%). Therefore, the major implication of this study would be that risk practitioners should acquire a deep understanding of markets' unique characteristics before considering a 'one-size fits all' approach of financial risk measurement.

1.6 Thesis Outline

The remainder of this thesis is divided into seven main chapters. First of all, Chapter 2 discusses the relevant literature and provides an overview of their main findings. Secondly, Chapter 3 outlines the key concepts of the theoretical framework. Next, Chapter 4 presents a detailed overview of the implemented data and explains the chosen methodology. Subsequently, Chapter 5 covers the empirical results for both Value at Risk (99%) and Expected Shortfall (97.5), respectively, and discusses the way these results should be interpreted. Next, Chapter 6 draws a conclusion, forms the main implications for risk practitioners, and provides suggestions for further research. Subsequently, Chapter 7 provides an overview of all the cited literature. Lastly, Chapter 8 embodies the Appendix. The first Appendix provides a list with all country indices that have a MSCI market classification, whereas the second Appendix is devoted to providing and explaining the performed script.

(2) LITERATURE REVIEW

The field of empirical literature on market risk quantification knows a very long history and has already been able to address many of the demanding challenges that risk practitioners tend to face on a daily basis. However, despite the fact that there has already been so much research conducted on the subject, there is still no universally accepted method on how to forecast risk. That is, to this day no risk model has been found that documents superior performance across all assets. In the light of international portfolio diversification, this raises the prominent issue in present-day risk quantification whether different country indices should be treated uniformly. That is, "Is the best risk model for country A also the best risk model for country B?". Over the last two decades, a significant part of literature has documented strong systematic differences in (return) dynamics between Developed and Emerging markets (see, amongst others, Aggarwal et al., 1999; Gençay & Selçuk, 2004; Božović & Totić, 2015) and between Developed and Frontier markets (see, amongst others, Bello & Adedokun, 2011; Chan-Lau, 2011). Therefore, it is critical to examine whether these unique characteristics of markets also demand a unique approach of risk measurement.

First of all, this chapter provides a profound review of the most important works on financial market risk measurement within the context of both Emerging and Frontier markets. It presents their main findings, along with a critical evaluation on the quality of their methodologies and conclusions. Subsequently, this chapter discusses why the current literature as a whole is not yet capable of providing a sufficient answer on whether a homogeneous approach of risk measurement is appropriate when dealing with countries that have heterogeneous market classifications. Finally, this chapter provides a discussion on why the paper of Dimitrakopoulos et al. (2010) is a step in the right direction. It explains its main findings, along with a critical evaluation on the limitations, deficiencies, and gaps left by the paper that need to be addressed in order to cope with present-day risk measurement.

#1 Gençay & Selçuk (2004)

This paper investigates the relative performance of VaR models for nine different Emerging markets. Their backtesting results indicate that the approach of conditional Extreme Value Theory is an indispensable part of risk measurement when forecasting VaR in Emerging markets. First of all, they state that Emerging markets are more likely to exhibit multiple regime switches in short periods of time than Developed markets, which suggests that one should use risk models that can capture return distributions with frequent changes in the parameters. Furthermore, using estimated tail indices, they argue that Emerging markets tend to experience larger financial earthquakes than Developed markets. This makes them to support the believe that risk practitioners should only incorporate risk models that are capable of capturing more extreme dynamics. However, before treating their conclusions as solid evidence, one should consider two important limitations exhibited by the chosen methodology. The first limitation would be that this study incorporates only the VaR violation ratio for the purpose of comparing VaR models. Although this approach gives some information about the quality of the VaR model, it does not provide any strong statistical information on the relative performance of VaR models. Therefore, one should also incorporate a loss function that statistically measures the difference in forecasting performance between each specified risk model. Another important weakness of this study would be that they use different sample sizes for each country, e.g. Taiwan (January 1, 1973 – December 29, 2000) and Mexico (June 6, 1995 – December 29, 2000). As a result, some countries can include significantly more data for the purpose of forecasting than other countries. However, a large quantity of empirical literature (e.g., Frey & Michaud, 1997; Angelidis et al., 2004) shows that the length of the sample influences the estimation quality of the risk model significantly. Therefore, it would be highly preferred to maintain equal sample sizes for each country.

#2 Maghyereh & Al-Zoubi (2006)

This paper investigates the relative performance of various VaR models for the largest capitalization markets in the Middle East and North Africa (MENA). Their sample consists of two Emerging markets (Egypt and Turkey) and five Frontier markets (Bahrain, Jordan, Morocco, Oman and Saudi Arabia) with return data spanning from January 1996 to December 2003. The backtesting results indicate that the method of EVT delivers significantly better results for the majority of the considered MENA markets than the more conventional methods such as a GARCH-type model. They illustrate that the usage of non-EVT methods to estimate financial market risk in MENA countries could lead to faulty estimation of risk during periods of high disturbance. Overall, they argue that the dynamics exhibited by MENA countries are explicitly different than those observed for Developed markets. Nevertheless, this study bears one crucial limitation that needs to be addressed. Whereas the previous discussed paper from Gençay & Selçuk (2004) incorporates only the violation ratio test, this paper makes the mistake of using only the conditional coverage test to compare VaR models. However, the conditional coverage test is only capable of testing the statistical adequacy of individual VaR models. This means that the test should not be used to statistically rank the performance of different VaR models. Instead, one should incorporate a loss function that statistically measures the difference in forecasting performance between each specified risk model. The consequence of this limitation would be that one should treat their conclusions on the relative performances with caution.

#3 Žiković (2007) | #4 Žiković & Aktan (2009)

The first paper of Žiković from 2007 tests a wide range of popular VaR models on one Emerging market (Turkey) and four Frontier markets (Bulgaria, Croatia and Romania) with return data spanning from 2000 to 2007. Based on the backtesting results, he claims that VaR models that work well for Developed markets are not necessarily well suited for measuring financial market risk in other markets. He argues that risk managers have to start thinking outside the frames of Developed markets or else they may find themselves in serious trouble, dealing with losses that they have not been expecting. Subsequently, Žiković & Aktan (2009) extent this research by investigating the relative performance of VaR models for the Turkish (Emerging) market and the Croatian (Frontier) prior to and during the global financial crisis of 2007-2009. They alter the approach of the original work of Žiković (2007) by adding the framework of conditional EVT in order to generate VaR for higher quantiles. Their backtesting results indicate that during times of turbulence all tested VaR models except cEVT tend to severely underestimate the true level of risk. They argue that only advanced VaR models such as cEVT can adequately measure financial market risk in the dynamic environment of Emerging and Frontier markets. Unfortunately, both works bear the same crucial limitation of Maghyereh & Al-Zoubi (2006) of not incorporating a loss function. Therefore, also these ranking results should be interpreted with caution.

#5 Vee et al. (2014)

This paper discusses the relative performance of conditional EVT models to forecast series of VaR on six Frontier markets (Croatia, Kazakhstan, Mauritius, Pakistan, Sri Lanka and Tunisia). They illustrate that the dynamic approach of cEVT performs generally well for all considered countries. They find the conditional EVT model with an IGARCH volatility specification to work satisfactorily for the far majority of the countries. Overall, they strongly argue that the Frontier markets are better described by risk models that look specifically at the tails of a distribution. Regarding further research, they advise to take a larger set of indices to ascertain any commonalities among Frontier markets. Nonetheless, it should be noted that this study includes only the approach of conditional EVT models in their backtesting routine, meaning that they explicitly exclude the framework of GARCH. In the paper it is stated that this decision is based on the work of Gençay & Selçuk (2004). As a reminder, this study argues that cEVT is more adequate. However, this exclusive focus on cEVT knows important weaknesses that need to be addressed. First of all, as already mentioned in this literature review, the paper of Gençay and Selçuk does not include any statistical test that supports their findings in a justified manner. Next, the study of Gençay and Selçuk is tailored towards Emerging markets, whereas this study is focused on Frontier markets. Lastly, and most importantly, the decision to include only conditional EVT models limits the capability of detecting systematic risk overestimation. It could be that the best performing risk model is actually a model that is a victim of systematic risk overestimation. Instead, in order to signal potential excessive allocation of capital, one should also include risk models that are more likely to be a victim of risk underestimation (e.g. models from the standard GARCH framework).

#6 Božović & Totić (2016)

The last paper compares the forecasting ability of various VaR models on six country indices from the South Eastern Europa (SEE) region, which includes one Emerging market (Hungary) and five Frontier markets (Bulgaria, Croatia, Romania, Slovenia and Serbia). The backtesting results indicate that only the cEVT-based models are capable of capturing risk during turbulent periods of time. Therefore, this paper draws the conclusion that capturing extreme tail risk is of critical importance when forecasting risk in markets other than Developed markets. However, this paper makes some disputable decisions regarding the way they backtest the results of VaR and ES. The first weakness would be that it includes only the conditional coverage test for the purpose of backtesting VaR. As already mentioned in the review of several other works, this test should not be used to rank the performance of different VaR models. The second weakness of the paper would be that it uses a violation ratio to rank ES. However, this approach does not provide any information that allows a statistical comparison between ES models. Moreover, the authors acknowledge this flaw by stating "Nevertheless, it is very informative". Although the violation ratio could be used as a guidance, the consequence of only using this statistic would be that their ranking results should be taken with much caution. A proper solution would be to incorporate a statistical test besides the violation ratio that allows for a statistical comparison, such as the Bootstrap ES test of McNeil & Frey (2000).

2.1 The State of Current Literature and Their Typical Limitations

The general finding of the discussed papers would be that they all support the believe that Emerging and/or Frontier markets exhibit dynamics that are different and more extreme than those observed for Developed markets. In particular, they argue that advanced risk models such as cEVT are necessary in order to achieve adequate financial market risk measurement in Emerging and/or Frontier markets. However, the gap left by these papers would be that none of them performs a direct examination on whether risk models perform statistically different across countries with heterogeneous market classifications. That is, they do not include countries from all market classification groups in order to compare the forecasting performance of risk model statistically. This examination is crucial for risk practitioners who invest in different markets and need to know whether a 'one-size fits all' approach of risk measurement is sufficient. Unfortunately, apart from the fact that all the discussed papers exhibit crucial weaknesses, they all incorporate different model specifications and testing methodologies. This would make it naïve to use their isolated testing results for the purpose of comparing market classifications. Nevertheless, there is one paper that is a step in the right direction and offers a stable foundation for the rest of this study.

Dimitrakopoulos et al. (2010)

This study compares the forecasting accuracy of various VaR approaches with data from four Developed markets and sixteen Emerging markets, spanning over America, Asia and Europe. The study uses return series for the period between June 1995 and September 2003. Regarding the considered risk models, they choose to incorporate the symmetric GARCH model, the asymmetric EGARCH model, and the framework of conditional EVT, all with various distribution functions. Overall, their backtesting results indicate that the most successful VaR models are common for both market groups. That is, VaR models seem to perform uniformly among the examined groups of Emerging and Developed markets, despite their unique characteristics. The only discrepancy in terms of forecasting performance is observed during periods of financial turmoil. In times of crisis, the Emerging markets seem to suffer more severely from a diminishing effect on the forecasting performance of VaR models. However, also this paper bears some critical drawbacks and leaves some important gaps that need to be addressed before one could translate their conclusions in terms of actual present-day risk management.

The first limitation would be that the paper includes only the conditional coverage test for the purpose of backtesting VaR. As already mentioned in the review of several other works, this test is not designed to discriminate between the forecasting accuracy of different risk models or to pick the 'best' risk model among 'good' risk models. As a result, the paper's conclusion of uniformity between market is based on at least weak evidence and should therefore be taken with much caution. A proper solution would be to incorporate a loss function that measures the degree of risk overestimation/underestimation of each 'good' VaR model. This would allow for a statistical ranking and a stronger conclusion.

The second limitation would be that the authors do not include a single country that bears a Frontier market classification. However, for the preceding decade, the increasing level of globalization and the global financial crisis of 2007-2009 have induced investors to seek for new diversification opportunities in the environment of Frontier markets. Therefore, in order to be relevant for present-day risk measurement, it would be better to also include a number of country indices that bear this market classification. At the same time, the paper includes only four Developed markets (Germany, Japan, the UK and the USA). However, to ascertain any commonalities among the classification group of Developed markets, it would be advisable to take a larger set of country indices.

The third limitation would be the paper's lack of direct practicality for present-day risk practitioners. Although the paper has been published in 2010, the authors have included return data only spanning till 2003. As the last two decades have shown a significant increase in globalization, it would be dangerous to assume that their findings and conclusions are still as relevant when looking at the changed relation between Developed and Emerging markets.

The last limitation would be that the paper only addresses the concept of VaR, while the concept of ES will be the new standard risk measure in the financial industry by December 31, 2019. Looking at the discussed papers in this chapter, it shows that this decision to only consider the measure of VaR is a prominent problem. Only the paper of Božović & Totić (2016) includes the concept of ES, however, this study bears the weakness that it does not include a proper backtest for ES. The fact that the work of Dimitrakopoulos et al. (2010) documents that Emerging markets behave differently during periods of financial turmoil raises the expectation that the far tails of their return distribution are far from equal. In order to be relevant for present-day risk management, future studies should incorporate ES to expose potential differences in the far ends of the tails of different markets.

The aim of this thesis is to fill these empirical voids of current literature by analyzing and comparing the forecasting accuracy of standard and more sophisticated market risk quantification frameworks from the perspective of the three market classification groups under both the current BCBS regulative framework of VaR (99%) and the future BCBS regulative framework of ES (97.5%). To ascertain any commonalities among the market classification groups, each market is represented by ten country indices. Moreover, this thesis seeks to solve the prominent problem of weak backtesting procedures in current literature by incorporating a backtesting procedure for both VaR and ES that allows for a statistical comparison between and within each market classification group. Overall, this thesis aims to provide robust analyses and crucial knowledge to risk practitioners who need to perform risk measurement on country indices with heterogeneous market classifications.

(3) THEORETICAL FRAMEWORK

This Chapter outlines the key concepts of the theoretical framework implemented in this study. First of all, the concepts of two major risk measures are introduced: Value at Risk and Expected Shortfall. A special emphasis is put on the interpretation of the mathematical terms. Next, the existence of various stylized facts in financial time series is highlighted along with their complex effects on forecasting VaR and ES. On top of that, a detailed review of the most common GARCH family models is provided together with an overview of several probability density functions. At last, the innovative framework of Extreme Value Theory is defined along with its unique application to forecasting VaR and ES.

3.1 Value at Risk

3.1.1 Brief History of Value at Risk

The origins of Value at Risk can be traced back to as far as the 1950s, where portfolio theorists Markowitz (1952) and Roy (1952) independently published the basic underlying mathematics. However, due to the limited availability of processing power in that period of time, the published risk measures were still only largely theoretical (e.g., Tobin, 1958; Treynor, 1961; Sharp, 1964). As the 1970s turned to the 1980s, markets were becoming more volatile and the sources of market risk were proliferating. Financial disasters and sophisticated innovations repeatedly emphasized the importance for banks to develop reliable risk assessment tools in order to quantify their risk exposure accurately (Holton, 2002). By that time, the resources necessary to calculate VaR were also becoming more available. Processing power had become inexpensive and data vendors such as Reuters and Bloomberg were starting to make large quantities of historical price data available. The complex environment induced financial institutions to develop and implement sophisticated VaR measures, however, for a long period of time, these tools remained primarily known to professionals within those institutions (Shams et al., 2012). The breakthrough of VaR to the financial industry as a whole could be largely attributed to the commercial bank J.P. Morgan. They developed their own VaR service called RiskMetrics™, which was rolled out to the public with fanfare and without any charge in October 1994 (Guldimann, 2000). The timing of introduction was perfect as it was during a time of deep global concerns about derivative instruments and leverage. Ultimately, VaR was recognized and accepted by the Basel Committee on Banking Supervision to be the basis for their regulative framework in 1996, making it the dominant measure of financial risk for both financial institutions and financial regulators (Engle & Manganelli, 2004).

3.1.2 Defining Value at Risk

The concept of Value at Risk seeks to give a relatively simple measure of financial risk by answering the question 'How bad can things get?' (Dowd, 2005). Financial institutions find a certain level of comfort in the idea of compressing all the Greek letters for all risk factors associated with a financial asset into one simple number (Hull, 2015). Jorion (2001) defined VaR as "the worst expected loss over a given horizon under normal market conditions at a given level of confidence". Formally, the function of VaR can then be mathematically expressed as

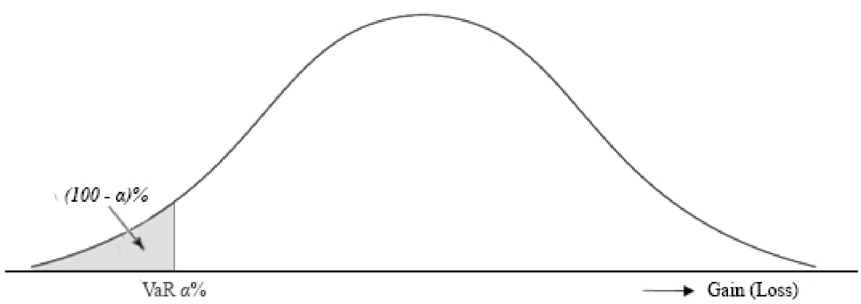
$$VaR_{\alpha}(X) = \min\{x \mid P(X > x) \leq 1 - \alpha\} \quad [3.1]$$

Equation [3.1] illustrates that VaR, given some confidence level α, is given by the smallest number x such that the probability that the loss X exceeds x is no larger than $1 - \alpha$. For example, imagine a time series of daily returns for an asset called 'Portfolio A'. If we would plot all these returns, we could probably end up with a bell-shaped curve as the one displayed in Figure [3.1]. If one wants to determine $VaR_{95\%}$, he should make a cut right between the bottom 5 percent and the top 95 percent. In other words, $VaR_{95\%}$ is a representation of the 5 percent quantile of the return distribution, as 5 percent of the outcomes are worse and 95 percent of the outcomes are better (Engle, 2001). In more formal terms, a 1-day $VaR_{95\%}$ of -1.645 percent would mean that in 95 out of 100 days the user does not expect to lose more than 1.645 percent on that specific asset. Based on this number, investors can get a more accurate overview of how risky the asset is and adjust their risk appetite accordingly.

The choice of confidence level α depends highly on the management's relation to risk (Gustafsson & Lundberg, 2009). In general, a more risk averse user of VaR will prefer a higher confidence level. The most

Figure 3.1 – VaR α% of Portfolio A

This figure illustrates VaR from the probability distribution of the change in the value of Portfolio A, where the confidence level of VaR is set to α%. Gains in the portfolio value are positive; losses are negative.



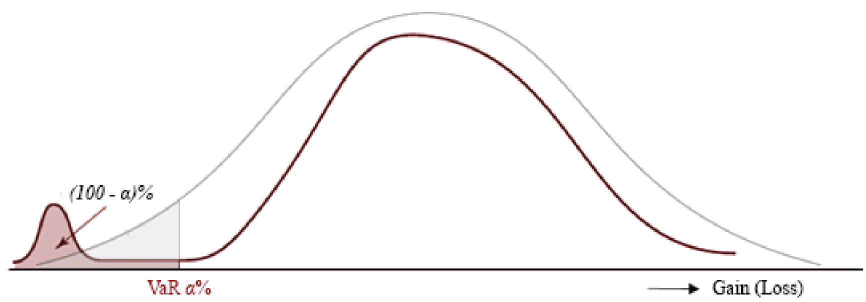
common confidence levels are between 95 percent and 99 percent, although they can vary in academic literature between 90 percent and 99.9 percent (Hendricks, 1996). The Basel Committee on Banking Supervision requires the use of 99 percent when determining minimum capital levels for commercial banks.

3.1.3 Limitations of Value at Risk

Although widely applied in practice for more than two decades, several limitations of Value at Risk as a regulative risk measure have been identified. The first problem would be the existence of tail risk, which implies that VaR does not provide any information about the severity of the potential loss (Dowd, 2005). VaR is simply a single quantile of the return distribution and therefore only describes the minimum potential loss of an adverse market outcome. It does not say anything about how much more money one can lose beyond the VaR estimate. Such a weakness is troublesome for regulators as very large negative market movements are the biggest threat for financial institutions, and even given high confidence levels, VaR would not be able to reveal these. This severe limitation has been visualized in Figure [3.2], where the probability distributions of the change in value of the two different assets are shown. The assets have the exact same value for VaR $\alpha\%$, but the red line belonging to Portfolio B has a much riskier tail than the grey line belonging to Portfolio A, meaning that the potential losses are much larger.

FIGURE 3.2 – VaR $\alpha\%$ of Portfolio A against VaR $\alpha\%$ of Portfolio B

This figure illustrates VaR from the probability distributions of the change in the value of Portfolio A (grey line) from Figure 1 and a new asset called 'Portfolio B' (red line), where the confidence level of VaR is set to $\alpha\%$. Gains in the portfolio value are positive; losses are negative.



The second limitation of VaR would be that it does not qualify as a coherent measure of risk on the grounds of not satisfying the property of sub-additivity (Artzner et al., 1999). In the context of financial risk, the property of sub-additivity means that the VaR of a portfolio as a whole can never exceed the sum of the VaR of its mutually exclusive sub-portfolios. Acerbi & Tasche (2002) and Dowd (2005) show that argue that sub-additivity is a fundamental requirement of any 'good' risk measure. The idea behind this statement is intuitive; diversification cannot make the risk greater, but it might make the risk smaller. When sub-additivty is violated, it could lead to a situation where financial institutions that want to have exposure to two portfolios would be better off by opening a separate account for each portfolio, as the risk-based margin requirement would be lower than if they were held in the same account.

3.2 Expected Shortfall

3.2.1 Brief History of Expected Shortfall

In 1996, the BCBS adopted VaR as their standard risk measure to determine banks' regulative capital requirements. From that moment on, VaR became the standard method for financial risk management. However, while the usage of VaR rose significantly, many academics claimed to have discovered severe conceptual drawbacks. These serious drawbacks were eventually acknowledged by the BCBS (2011) in their review of academic literature on risk management. In their search for possible alternative risk measures, they identified Expected Shortfall as an acceptable option. This was supported by many academic researchers such as Artzner et al. (1999) and Kerkhof & Melenberg (2004), who classified ES as a viable risk measurement tool that satisfied the sub-additivity property and simultaneously captured the tail risk. In May 2012, the BCBS (2012) issued a notion to replace VaR with ES, as the global financial crisis of 2008 painfully reminded banks and regulators alike that stress markets were capable of producing losses far in excess of amounts forecasted by VaR. The BCBS (2016) did state in one of their most recent reports that "ES will help to ensure a more prudent capture of tail risk and capital adequacy during periods of significant financial market stress" and they require banks to report under the new standards by December 31, 2019. The BCBS proposes the usage of 97.5% confidence level as the new standard, as $ES_{97.5\%}$ should yield a similar magnitude of risk as $VaR_{99\%}$ under the standard normal distribution.

3.2.2 Defining Expected Shortfall

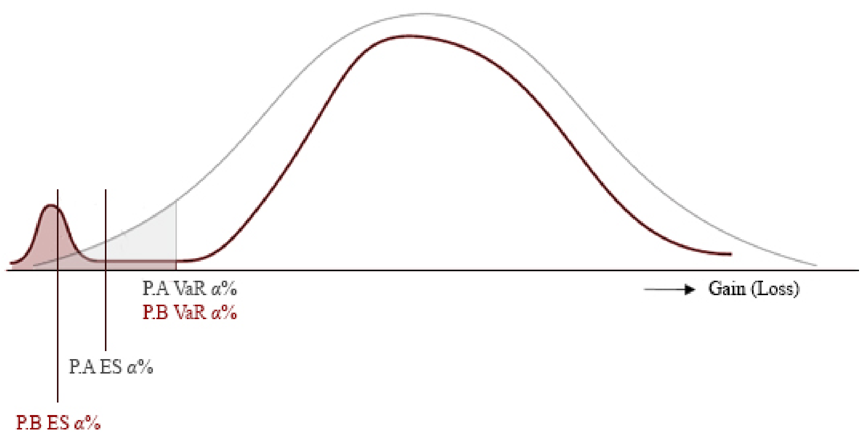
Artzner et al. (1999) proposed Expected Shortfall (also called 'Conditional VaR') in order to alleviate the inherent problems existing in Value at Risk. Where VaR asked the question 'How bad can things get?', the ES measure asks 'If things go bad and VaR is exceeded, how much can we expect to lose?' instead. Hull (2015) defines ES as "the expected loss during time T conditional on the loss being greater than the X^{th} percentile of the loss distribution". In other words, ES at confidence level α is the conditional expectation of loss given that the loss is beyond the VaR level. Mathematically speaking, ES can then be defined as

$$ES_\alpha(X) = E[X|X > VaR_\alpha(X)] \quad [3.3]$$

This equation is very appealing as it takes into account the magnitude of potential losses beyond the VaR threshold as well as their probability of occurring. A risk measure with such properties is clearly useful for risk managers and can be easily understood in boardrooms (Johansson & Engblom, 2015). This advantage is made visual in Figure [3.3], where again the probability distributions of the change in the value of Portfolio A and Portfolio B are shown, along with the fictional values of VaR $\alpha\%$ and ES $\alpha\%$. As already explained in Section 3.1.3, the portfolios have the exact same value for VaR $\alpha\%$, despite the higher risk in the tail of Portfolio B. To overcome this problem, the framework of Expected Shortfall looks at the tail and takes into account the observations that lie beyond the α^{th} percentile of the loss distribution. The figure shows that the value for ES $\alpha\%$ for Portfolio B lies below the one for Portfolio A, meaning that it is indeed capable of visualizing the higher risk that belongs to Portfolio B. At the same time, the concept of ES satisfies the principle of sub-additivity, which makes it a coherent risk measure.

FIGURE 3.3 – VaR $\alpha\%$ and ES $\alpha\%$ of Portfolio A and B

This figure illustrates VaR and ES from the probability distributions of the change in the value of Portfolio A (abbreviated as P.A. - grey line) and Portfolio B (abbreviated as P.B. - red line), where the confidence level of VaR and ES is set to $\alpha\%$. Gains in the Portfolio value are positive; losses are negative.



3.3 Forecasting Value at Risk and Expected Shortfall

In the context of financial risk management, the main challenge would be to calculate an accurate one-day-ahead forecast for Value at Risk and Expected Shortfall, respectively. When one wants to estimate the one-day-ahead VaR, this could be done by implementing the following equation:

$$VaR_{\alpha,t+1|t} = \hat{\mu}_{t+1|t} + \hat{\sigma}_{t+1|t}F(1-\alpha;\theta^t) \quad [3.4]$$

Equation [3.4] illustrates that the one-day-ahead VaR of a single asset depends on at least three key factors; (i.) the conditional forecast of the mean at time $t+1$, given the information available at time t, (ii.) the conditional forecast of the standard deviation at time $t+1$, given the information available at time t, and (iii.) the assumed probability density function, given the estimated parameters θ at time t. Following Equation [3.4], the one-day-ahead ES can then be defined as follows:

$$ES_{\alpha,t+1|t} = \frac{1}{1-\alpha}\int_{0}^{1-\alpha} VaR_{u,t+1|t}\, du \qquad [3.5]$$

The above equation illustrates that ES is calculated by taking the average of $VaR_{u,t+1|t}$ for all confidence levels u in the interval [0, (1-α)].

3.4 Stylized Facts

As revealed by an offhand examination of articles published by financial journals, the viewpoint of many market analysts tends to be an event-based approach in which they attempt to rationalize market movements by relating it to an economic or political driven news event (Cont, 2001). From this point of view, one could come to the belief that price series obtained from different assets or markets will exhibit completely different properties as they are not necessarily influenced by the same type of news events or information sets. After all, why should the properties of Apple Inc. shares be similar to those of crude oil futures or the Euro/Dollar exchange rate? Nevertheless, an extensive body of empirical research carried out since the 1950s indicates that this is actually the case if one examines their properties from a statistical point of view. These statistical properties that are common across a wide range of instruments, markets and time periods are usually referred to as stylized facts (McNeil et al., 2005). Unfortunately, Berkowitz & O'Brien (2002) and Pérignon and Smith (2010) find that the most prominent commercial banks from the USA still rely on models that ignore stylized facts and perform quite poorly in forecasting volatility changes. This ignorance could eventually lead to dangerous situations such as clustering of Value at Risk violations and a severe chance on bankruptcy.

3.4.1 Stylized Fact 1: Leptokurtic Distribution

Mandelbrot (1963) was the first one to document the stylized fact that empirical distributions belonging to financial time series tend to exhibit fat tails, meaning non-Gaussian, sharp-peaked and heavy-tailed distributions. These type of distributions are often classified as leptokurtic distributions and they enable the accommodation of the likelihood of large positive and negative shocks (Cont, 2001). Financial time series that are affected by such behavior tend to have much more extreme events than the normal distribution can predict, which can cause serious problems in the framework of VaR and ES. Gokcan (2000) illustrates that falsely assuming normality can lead to situations where risk estimates are systematically overestimated on low confidence levels and systematically underestimated on high confidence levels. Rachev et al. (2005) support these findings and show the importance of incorporating models that satisfy the tail assumption of financial returns. However, although the presence of fat tails warrants the use of more sophisticated distributions, the normal distribution is still dominant in finance due to its simplicity and analytical tractability (Bidarkota & Dupoyet, 2004).

3.4.2 Stylized Fact 2: Return Asymmetry

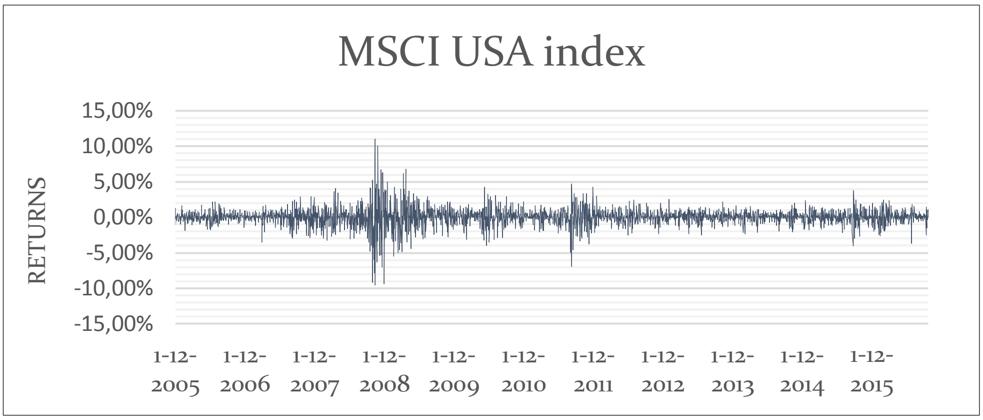
An early paper from Fama (1965) addresses the stylized fact that financial time series tend to exhibit an asymmetry between the upside and downside potential in returns. This phenomenon is usually referred to as skewness, where a distribution with negative skewness has a longer tail in the lower-return side and where a distribution with positive skewness has a longer tail on the higher-return side of the curve. The work by Cheng et al. (2000) states that the very largest movements in financial time series are usually decreases, rather than increases – that is, these financial assets are more prone to melt down than to melt up. This statement is supported by Cont (2001), who observes large drawdowns but not equally large upward movements in the S&P500, and by Engle & Patton (2001), who report that equity returns are usually substantially negatively skewed. Damodaran (1985) explains this behavior by the asymmetrical reaction of investors to good news and bad news. A crucial implication of skewness within the context of risk modeling would be that one should find a distribution that models this problem more adequately than the normal distribution does, as the normal distribution assumes that there is no significant asymmetry in financial returns.

3.4.3 Stylized Fact 3: Volatility Clustering

As early as the 1960s, Mandelbrot (1963) observed a certain pattern in the volatility of financial time series, which he summarizes as: "Large changes tend to be followed by large changes – of either sign – and small changes tend to be followed by small changes". This stylized fact, which is often referred to as volatility clustering, states that return series belonging to financial assets reveal significant positive and slowly decaying autocorrelation (Fama, 1965). According to this behavior, today's volatility is a good predictor of volatility in the next period. What this means in practice is that when a market suffers a volatile shock, more volatility should be expected. Figure [3.4] shows a time series plot of daily returns for the MSCI USA index. From this plot, it is apparent that the amplitude of the return is changing over time, i.e. periods of low (high) volatility are followed by periods of low (high) volatility. Consequently, academics have designed several methods that try to quantify and model this phenomenon, from which the ARCH and GARCH family models are the most common ones.

FIGURE 3.4 – Volatility Clustering

A plot of the daily log return series for the MSCI USA index, including observations from 01.12.2005 to 22.09.2016.



3.4.4 Stylized Fact 4: Leverage Effect

The last stylized fact that should be addressed is 'the leverage effect'. Works by Simon (2003), Giot (2005) and Hibbert et al. (2008) show that future volatility tends to be higher following a negative shock in returns than following a positive shock of equal magnitude, i.e. returns and volatility are often significantly negatively correlated. Black (1976) and Christie (1982) argue that as asset prices decline, companies become mechanically more leveraged since the relative value of their debt rises relative to that of their equity. As a result, the equity of a firm is more exposed to the firm's total risk, which makes their stock riskier and hence more volatile. In response, academics have developed sophisticated risk models that incorporate a possible asymmetric response in volatility.

3.5 ARMA-GARCH models

The forecasting formulas of VaR and ES of Section 3.3 both consist of two parameters: the conditional mean ($\hat{\mu}_{t+1|t}$) and the conditional variance $\hat{\sigma}_{t+1|t}$. The conditional mean is modelled by the ARMA model, whereas the conditional variance is modelled by the GARCH model and its numerous extensions.

3.5.1 Autoregressive Moving Average (ARMA)

The Autoregressive Moving-Average (ARMA) process is designed for modeling the mean of a time series. The model combines an autoregressive (AR) part with a moving-average (MA) part in order to capture the time dependencies that are often exhibited by financial return series and is given by:

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \qquad [3.6]$$

The first part of Equation [3.6] represents the AR process, where φ_i depicts the parameters of the autoregressive component of order p, indicating that the current value of a time series depends linearly on its own previous values and an error term. The second part of Equation [3.6] introduces the MA process, where θ_j denotes the parameters of the moving average component of order q, indicating that the current value of a time series depends on the current value and the lagged value(s) of the error terms (ε_t).

3.5.2 Autoregressive Conditional Heteroscedasticity (ARCH)

The ARMA process of the previous section operates under the assumption that the volatility of returns is constant over time. However, many academics such as Mandelbrot (1963) and Akgiray (1989) have illustrated that in the context of financial time series the volatility is rarely a constant, i.e. the volatility of returns tends to change over time. Therefore, Engle (1982) introduced the Autoregressive Conditional Heteroscedasticity (ARCH) model, which accommodates the time-varying concept of the volatility process. Before Engle's publication, the most popular tool for calculating volatility was the rolling standard deviation (Engle, 2001). This simplistic model assumes that the variance of tomorrow's return is an equally weighted average of the squared residuals from the last n days. This specification of giving equal weights seems to be unattractive as it neglects the presumably higher relevance of more recent events. Therefore, Engle (1982) proposed the ARCH process where weights are parameters to be estimated, i.e. where the data determines the best weights. The mathematical framework of the ARCH(q) process is as follows:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 \qquad [3.7]$$

Equation [3.7] represents the general form of an ARCH(q) model, where $\omega > 0$ and $\alpha_i \geq 0$, and where the sum of the autoregressive parameters should be less than one in order for the process to be stationary (Poon & Granger, 2005). The process lets the conditional variance be a function of the squared error terms of the previous q days, where the lag order q determines how long a shock persists (Bera & Higgings, 1993).

3.5.3 Generalized Autoregressive Conditional Heteroscedasticity (GARCH)

Bollerslev et al. (1992) disclose an important limitation of the ARCH process by documenting that the stylized fact of volatility clustering typically causes the lag order q to be of a high order. This tendency induces a non-parsimonious conditional variance model where non-negative constraints on α_i from Equation [3.7] are more likely to be violated (Abdalla, 2012). In order to avoid the potential problems of long-lag structures and negative coefficients, Bollerslev (1986) proposed a generalized version of the ARCH model, better known as the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model. It treats variance as a persistent phenomenon, while still allowing for a more parsimonious description of data (Engle, 2001). The general form of the GARCH(p, q) model is given by:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \qquad [3.8]$$

It is necessary to impose certain restrictions on Equation [3.8] such as $\omega > 0$, $\alpha_i \geq 0$, $\beta_j \geq 0$, and $\sum_{i=1}^{\max(p,q)} (\alpha_i + \beta_i) < 1$ (Tsay, 2010). The last condition guarantees that there is stationarity in variance (Brooks, 2008). When Equation [3.7] and [3.8] are compared, it shows that Bollerslev (1986) has generalized the ARCH model by including lagged values of the conditional variance σ_{t-j}^2. In Equation [3.8], the parameter α_i measures the volatility response to movements in the market and parameter β_j denotes the persistence of shocks to the conditional variance. The relative sizes of these parameters determine the observed volatility dynamics obtained from the financial time series. In a scenario where β_j is relatively large to α_i, the volatility will appear to be persistent as it will remain at around the same level for a longer period of time, while if the reverse is true, then the volatility will appear to be more spiky as it reacts more quickly to market movements (Dowd, 2005).

3.5.4 Integrated GARCH (IGARCH)

A large body of studies dealing with financial time series have shown that the sum of the parameter estimates α_i and β_i are typically very close to unity. This provided Engle & Bollerslev (1986) with the empirical motivation to propose the Integrated GARCH(p, q) model. The process is given by the standard GARCH framework from Equation [3.8], but adds the parameter condition that $\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j = 1$.

3.5.5 Exponential GARCH (EGARCH)

It should be noted that the GARCH and IGARCH specifications ignore information about the direction of asset returns: i.e. only the magnitude is taken into consideration. As they incorporate only a symmetric response to market news, both of the processes are often referred to as 'symmetric models'. However, there is very convincing evidence that the direction of returns has a significant effect on the volatility as well. This phenomenon of asymmetric response belongs to the concept of the leverage effect, which suggests that negative returns tend to be followed by periods of higher volatility than positive returns of equal size (Engle, 2001). Alexander (2001) documents that when symmetric GARCH and IGARCH processes are applied on time series showing a significant leverage effect, the problem of having very spiky conditional volatilities will play up, i.e. they will show a large reaction (large α_i) and a low persistence (low β_i). This implies that symmetric models will often give a too low (large) estimate of the conditional volatility after a price drop (increase). In order to capture the possible leverage effect exhibited by financial time series, a family of 'asymmetric' GARCH models has been developed. One of the earliest and most popular asymmetric models is the Exponential GARCH (EGARCH) model introduced by Nelson (1991). Another important drawback of the GARCH and IGARCH models would be the need to impose estimation constraints on the coefficients in order to ensure positive conditional variance. The EGARCH model tackles this problem by implementing a log transformation on the conditional variance, which guarantees non-negativity without the need to impose any additional non-negative constraints. The EGARCH(p, q) model is given by

$$\ln(\sigma_t^2) = \omega + \sum_{i=1}^{q} \alpha_i g(Z_{t-i}) + \sum_{j=1}^{p} \beta_j \ln(\sigma_{t-j}^2) \quad [3.9]$$

Nelson (1991) noted that "to accommodate the asymmetric relation between stock returns and volatility changes (…), the value of $g(Z_t)$ must be a function of both the magnitude and the sign of Z_t". This leads to the following equation:

$$g(Z_t) = \underbrace{\gamma_1 Z_t}_{sign\ effect} + \underbrace{\gamma_2[|Z_t| - E|Z_t|]}_{magnitude\ effect} \quad [3.10]$$

In the case of financial time series, the estimated parameter γ_1 is usually found to be negative, which reflects the leverage effect.

3.5.6 Glosten-Jagannathan-Runkle GARCH (GJR-GARCH)

Glosten et al. (1993) proposed the GJR-GARCH process as an alternative to the EGARCH process. Similar to the latter, the GJR-GARCH process attempts to capture the asymmetric effect of different signs of past residuals on the conditional variance. The non-negative constraints are similar to those of the standard GARCH model. The GJR-GARCH(p, q) process takes the following form:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} (\alpha_i \varepsilon_{t-i}^2) + \sum_{i=1}^{q} (\gamma_i d(\varepsilon_{t-i} < u) \varepsilon_{t-i}^2) + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \quad [3.11]$$

The only difference is the extra constraint that the sums of α_i and γ_i must be equal or above zero. The $d(\cdot)$ in Equation [3.11] denotes the indicator function, which is structured as follows

$$d(\varepsilon_{t-i} < u) = \begin{cases} 1, & \varepsilon_{t-i} < u \\ 0, & \varepsilon_{t-i} \geq u \end{cases} \quad [3.12]$$

If the lagged residuals are below the threshold value u, which is in the case of the standard GJR-GARCH model set to zero, the indicator function takes the value one, and zero otherwise. A significant estimate for the asymmetry parameter γ_i reveals the existence of an asymmetric reaction to news. Regarding financial time series, the estimated asymmetry parameters are usually found to be positive. This implies that positive news contributes nothing but $\alpha_i \varepsilon_{t-i}^2$ to the conditional variance, whereas negative news has a larger contribution denoted by $(\alpha_i + \gamma_i)\varepsilon_{t-i}^2$.

3.5.7 Threshold GARCH (TGARCH)

The Threshold GARCH (TGARCH) model Developed by Zakoian (1994) makes one fundamental change to the framework of GJR-GARCH. In contrast to modeling the conditional variance directly using lagged squared error terms, the TGARCH model parameterizes the conditional standard deviation as a function of lagged absolute values of the error terms. The TGARCH(p, q) process can be expressed as:

$$\sigma_t = \omega + \sum_{i=1}^{q} (\alpha_i |\varepsilon_{t-i}|) + \sum_{i=1}^{q} (\gamma_i d(\varepsilon_{t-i} < u)|\varepsilon_{t-1}|) + \sum_{j=1}^{p} \beta_j \sigma_{t-j} \quad [3.13]$$

The $d(\cdot)$ in Equation [3.13] denotes the indicator function, which is structured as follows

$$d(\varepsilon_{t-i} < u) = \begin{cases} 1, & \varepsilon_{t-i} < u \\ 0, & \varepsilon_{t-i} \geq u \end{cases} \quad [3.14]$$

3.5.8 Nonlinear Asymmetric GARCH (NAGARCH)

In order to deal with situations where asset prices and volatilities are negatively and asymmetrically correlated, Engle & Ng (1993) introduced the Nonlinear Asymmetric GARCH (NAGARCH) models, also known as the NGARCH model. The general form of the NAGARCH(p, q) model is as follows:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i (\varepsilon_{t-1} - \theta \sigma_{t-1})^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-1}^2 \quad [3.15]$$

The above equation illustrates that the asymmetry parameter θ is added to the standard GARCH(p, q) framework in order to control for the direction of the shock. Regarding financial time series, the estimated asymmetry parameter is usually found to be positive. If the parameter is indeed positive, it reflects the leverage effect, as by design of the equation negative returns will then increase future volatility by a larger amount than positive returns of the exact same magnitude.

3.5.9 Probability Density Functions

In his original work on the ARCH model, Engle (1982) assumes the random variable z_t to be normally distributed. This assumption of normality is popular due to its conceptual simplicity and the fact that it is included in the lion's share of statistical software. Nonetheless, many academics such as Mandelbrot (1963), Fama (1965) and Christoffersen & Diebold (2000) illustrate that financial returns tend to be leptokurtic, meaning that time series of financial returns are likely to exhibit fatter tails than assumed by the normal distribution. A serious implication of falsely assuming normality in a risk measure framework would be a systematical underestimation (overestimation) of risk at (high) low levels of significance as the non-normal characteristics of financial returns are not captured (Gokcan, 2000). This urgency of using another distribution than the normal distribution is deeply supported by Shephard (2013). He highlights the believe that when GARCH models are used for market risk management, it is critical to optimize the precision of estimated volatility process parameters, which can be achieved by applying a better approximation to the conditional distribution of standardized returns.

In response to these findings, Bollerslev (1987) proposed the application of the Student-t distribution to account for the high degree of leptokurtosis exhibit by financial time series. It was argued that a standard GARCH$(1,1)$ model with a Student-t distribution would offer a better fit than when the normal distribution was implemented, even in situations where the real underlying distribution of the residuals is unknown.

However, a fundamental drawback from using the Student-t distribution is that it does not account for possible skewness in the financial return series, i.e. asymmetry around the mean. When the used model fails to incorporate negative (positive) skewness, it will fail to account for a longer left (right) tail. A model that is capable of capturing the potential asymmetry of the error distribution is the reparametrized Johnson SU (JSU) distribution (Rigby & Stasinopoulos, 2005). The distribution has two shape parameters that allow for a wide range of skewness and kurtosis levels. In fact, Alexander et al. (2011) argue that the JSU distribution is most relevant for financial applications due to its capability of fitting data with a skewed leptokurtic distribution. Simonato (2012) employs the skewed and leptokurtic JSU distribution in the specifications of the GARCH process, from which he concludes that large negative shocks commonly affecting financial time series are adequately captured by the JSU distribution.

3.6 Extreme Value Theory

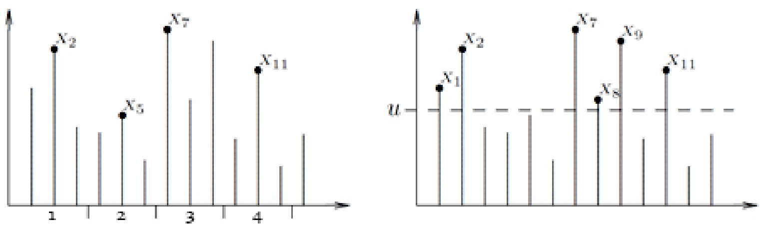
When it comes to measuring the risk of a financial asset, a typical question one would like to have answered is 'If things go wrong, how wrong can they go?'. In general, large and rare losses are the ones that can cause significant instabilities in financial markets worldwide. By way of explanation, it is the upper tail of the loss distribution that usually gives the most interesting outcomes. This has motivated the search for applicable methodologies that are capable of coping with rare events that have heavy financial consequences. Unfortunately, the traditional parametric approaches using the Gaussian distribution or the Student-t distribution often fail to adequately estimate conditional expectations that are far out in the tail. The field of finance is in need of an approach that is capable of extracting more information from the large losses that are observed, and that allows a better prediction of large and rare losses, possibly even larger than the ones within the range of available observations. Extreme Value Theory (EVT) has become a well-established theoretical framework in a wide array of fields, reaching from modern science to astronomy. Over the last two decades, numerous research studies have followed the example of Embrechts et al. (1997), McNeil (1997) and Embrechts et al. (1999) of analyzing the potential benefits of using EVT in the unique field of finance. It is a common finding that EVT is capable of providing a firm theoretical foundation on which the tail behavior of financial time series could be modelled.

The majority of risk models focus on modeling the entire distribution, where observations at the center of the distribution dominate the estimation process, given the relative scarcity of extreme observations (Danielsson, 2011). As a result, these models can give a good approximation of the distribution of data for common events, but provide inaccurate estimates for the rare events on which the tails are built upon. EVT, on the other hand, focuses explicitly on modeling the tails regions of the distribution by using solely extreme values rather than the whole dataset.

There are two main approaches within the framework of EVT to model extreme events: Block Maxima (BM) and Peaks-over-Threshold (POT). The fundamental difference between these two methods is how extreme events are identified and how the principal distribution is used. Let us first consider a random variable X_t that represents the daily losses of a financial asset. In the first approach, the data points of variable X_t are divided into equal blocks (e.g. months), from which the maxima of all blocks are modelled followed the Generalized Extreme Value (GEV) distribution. In the left panel of Figure [3.5], the observations X_2, X_5, X_7, X_{11} represent the block maxima for four successive periods of three observations each. The second approach focuses on the data points of variable X_t that exceed a certain threshold value u. In the right panel of Figure [3.5], the observations X_1, X_2, X_7, X_8, X_9 and X_{11} are all classified as extreme events as they exceed the selected threshold u. After determining the extreme events, all outcomes are fitted to the Generalized Pareto Distribution (GPD).

FIGURE 3.5 – Illustrative Example for the Block Maxima and Peaks-over-Threshold Approach

The left plot represents the Block Maxima approach where the sample of variable X_t is divided into four successive periods of three observations. The maxima of each block are highlighted. The right plot illustrates the Peaks-over-Threshold approach where data points from variable X_t that exceed a certain threshold value u are highlighted.



The BM method is typical for datasets that are affected by seasonality, such as hydrological and climatologic data, but is somewhat wasteful of data if other extreme events are available in the subsamples. The POT method, on the other hand, uses data more efficiently and hence is not as dependent on the requirement for large datasets as the BM method. For this reason, the approach of using a threshold u is more useful in the context of measuring financial risk where observations are scarce (McNeil et al., 2005).

3.6.1 Peaks-over-Threshold

The procedure of Block Maxima is rather wasteful on data and a relatively large dataset is needed in order to achieve accurate estimation. The Peaks-over-Threshold, on the other hand, is based on all large observations that exceed a high threshold and hence makes better use on data on extreme values (Danielsson, 2011). This makes the POT-model in respect to financial applications the more modern and generally preferred model.

$$(X_i)_{i \in T} = X_1, X_2, \dots, X_T \qquad [3.16]$$

Equation [3.16] illustrates a sequence of independent and identically distributed (i.i.d.) random variables. Following the results from Balkema & De Haan (1974) and Pickands (1975), the distributions of exceedances over a high threshold are well approximated by the Generalized Pareto Distribution. The limiting distribution $H(\cdot)$ of $(X_i)_{i \in T}$ beyond some high threshold takes the following form:

$$G_{\xi,ß}(x) = \begin{cases} 1 - (1 + \xi \left(\frac{x}{ß}\right)^{-\left(\frac{1}{\xi}\right)}, & \xi \neq 0 \\ 1 - \exp\left(\frac{x}{ß}\right), & \xi = 0 \end{cases} \qquad [3.17]$$

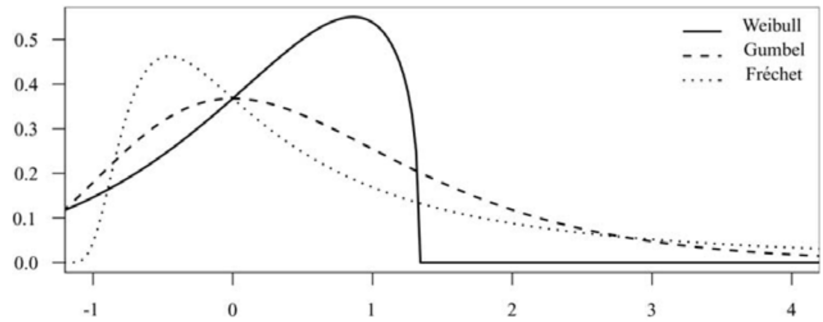
In the framework of GPD, one needs to estimate both the shape (ξ) and scale (ß) parameter. The $G_\xi(\cdot)$ becomes the Weibull if $\xi < 0$, the Gumbel if $\xi = 0$ and the Fréchet $\xi > 0$. Figure [3.6] shows the probability density functions belonging to the standard Weibull, Gumbel and Fréchet distributions. The Weibull has a finite endpoint, whereas the Gumbel and Fréchet have an infinite endpoint. Furthermore, the Gumbel shows an exponentially declining tail whereas the tail of the Fréchet declines by a power law. As a result, the Fréchet is more appropriate for financial applications as it suits heavy tailed distributions better.

3.6.2 Conditional Extreme Value Theory

The unconditional Extreme Value Theory approach explained in the previous section assumes that the data under study is independently and identically distributed, which is clearly not the case for most financial return series. In order to overcome this drawback, McNeil & Frey (2000) propose a conditional Extreme

FIGURE 3.6 – Extreme Value Distributions

The Extreme Value Theory states that, regardless of the overall shape of the distribution, the tails of all distributions fall into one of the following categories: Weibull, Gumbel, or Fréchet. This figure provides a graph with the three distinctive forms.



Value Theory (cEVT) approach that combines GARCH volatility forecasting with EVT tail estimation. The conditional EVT approach involves a multistage procedure for estimating VaR and ES:

Stage 1: In order to achieve essentially i.i.d. series, one must first filter the return observations by fitting an ARMA-GARCH family model to the return data by quasi-maximum likelihood. That is, maximize the log-likelihood function of the sample by assuming a distribution of innovations. Consequently, the $\mu_{t+1|t}$ and $\sigma_{t+1|t}$ should be estimated from the fitted model and the standardized residuals z_t should be extracted.

Stage 2: The standardized residuals from Stage 1 are considered to be realizations of a white noise process, which enables an estimation of the tails by the Peaks-over-Threshold method of EVT. Next, the desired quantiles are to be computed.

Stage 3: Next, the GPD quantiles of Stage 2 should be used in conjunction with the dynamic one-day-ahead forecasts from Stage 1 in order to obtain both VaR and ES estimates:

$$VaR_{\alpha,t+1|t} = \hat{\mu}_{t+1|t} + \hat{\sigma}_{t+1|t}\left[u + \frac{ß}{\xi}\left[\left(\frac{1-\alpha}{F(u)}\right)^{-\xi} - 1\right]\right] \quad [3.18]$$

$$ES_{\alpha,t+1|t} = \frac{1}{1-\alpha}\int_{0}^{1-\alpha} VaR_{u,t+1|t}\, du \quad [3.19]$$

The advantage of combining GARCH and EVT lies in the ability to capture conditional heteroscedasticity in the data through the GARCH framework, while simultaneously being able to model the extreme tail behavior through the POT method of EVT.

(4) DATA AND METHODOLOGY

This Chapter presents a detailed overview of the implemented data along with the chosen methodology of testing. First of all, the data on the different markets are introduced together with their descriptive statistics. Inside the subsection of autocorrelation, the selection procedures of the ARMA and GARCH orders are explained and implemented upon the thirty market indices. Secondly, the advanced technique for calculating the threshold value within the framework of Extreme Value Theory is described. Next, the choice of the window structure regarding estimating and backtesting is explained. Finally, the backtests for Value at Risk and Expected Shortfall are provided and the chosen backtesting procedures are clarified.

4.1 Descriptive Statistics

In order to investigate the return dynamics of the different markets thoroughly, the sample of the research comprises return series of stock indices for 30 selected markets: 10 are classified as Developed markets (United States of America, Japan, Hong Kong, France, United Kingdom, Germany, Canada, Australia, Netherlands, Italy), 10 are classified as Emerging markets (China, India, South Korea, Brazil, Mexico, Malaysia, Indonesia, Poland, Colombia, Peru), and 10 are classified as Frontier markets (Nigeria, Kazakhstan, Croatia, Romania, Kenya, Tunisia, Lebanon, Mauritius, Slovenia, Estonia). The value-weighted Morgan Stanley Capital International (MSCI) national stock market indices are used as proxies for the markets' equity portfolios, and their market classifications for the specific indices are leading. These indices are widely recognized as one of the most comprehensive and reliable measures that aim to cover the performance of the most active stocks in their respective stock markets. The selection of the chosen indices has been made carefully, as these markets (i) have the highest stock market capitalization within their market classification group, (ii) have return data availability since 01.12.2005 in the financial database DataStream, and (iii) have consistent market classifications according to MSCI over the whole sample period, e.g. Argentina is only a Frontier market since May 2009 and should therefore not be included. All the available MSCI stock markets are provided in Appendix [1], along with their MSCI market classifications. The list shows that this study covers almost 50 percent of all available MSCI stock markets. However, it should be noted that this list also includes markets that could not be included in this study, due to either a lack of data or intermediate changes in market classifications within the chosen period of time.

The price series are obtained from the financial database DataStream and span the period from 30.11.2005 till 22.09.2016, amounting to a total of 2822 daily observations per index. The daily stock index returns are calculated as $r_t = \ln(P_t) - \ln(P_{t-1})$, where $\ln(P_t)$ denotes the log of the stock index price on day t. For this study, the time series of returns need to be divided into two periods, one period for the purpose of parameter estimation in both the ARMA-GARCH and cEVT frameworks (i.e. in-sample) and one period for the purpose of backtesting the forecasted series of VaR and ES (i.e. out-of-sample). The length of both periods are found to play an important role, but there is till this point no clear method for deciding the optimal number of observations. Hendricks (1996) and Danielsson (2002) argue that an increase in the length of the 'in sample' period tends to generate more accurate forecasts. However, Frey & Michaud (1997) and Angelidis et al. (2004) show that the 'in-sample' size should not be too large as otherwise the GARCH models can lose their capability of capturing structural changes in trading behavior. Overall, it seems that there is a kind of trade-off between a long 'in-sample' that is necessary to obtain reliable estimates and a prolonged extension that does not reflect the current market sentiment. McNeil & Frey (2000) state that within the framework of ARMA-GARCH and cEVT a length of around 1000 observations is often appropriate. Another hard challenge is choosing the optimal length for the 'out-of-sample'. Nordbo et al. (2012) investigate the power of backtesting procedures and discovered that the minimum length for the 'out-of-sample' equals 1000 observations, a finding supported by Pfaff (2012). In order to adapt to all these empirical findings and to create an optimal structure for this study, a length of 1321 observations is chosen for the 'in-sample' period and 1500 observations for the 'out-of-sample' period.

4.1.1 Normality

A popular stylized fact for financial time series data is that the empirical return distribution often differs from the normal distribution: they tend to exhibit a leptokurtic distribution with high peaks and fat tails. Table [4.1] reports the kurtosis coefficients, the skewness coefficients, and the Jarque-Bera statistics. These statistics help to establish an initial understanding on the form of the return distribution of the used data. The first measure, which is the kurtosis coefficient, refers to the degree of peakedness of a distribution. A kurtosis coefficient exceeding the value of three implies a distribution with a high, small peak around the mean with fat tails. In this situation, the probability of extreme losses is higher than that observed for a normal distribution. According to the significant sample kurtosis estimates at the 1 percent confidence level, one could say that all the daily returns are far from being normally distributed. The lowest kurtosis estimates are 7.769 (Poland) and 8.129 (Japan), while the highest estimates are 48.030 (Nigeria) and 26.055 (Romania). A salient detail would be that high levels of kurtosis are more common across Frontier markets. This could suggest that these markets have lower returns in general, but simultaneously do face more extreme outliers.

The second measure denotes the skewness coefficient, which refers to the lack of symmetry in the return distribution. A skewed distribution occurs when one half of the return distribution does not mirror the other half. According to the sample skewness estimates, the vast majority of markets tend to have a significant asymmetric distribution as no less than 20 out of the 30 considered stock market indices show significantly negative skewness coefficients at the five percent confidence level. Another interesting result would be that all five of the significantly positive skewness coefficients belong to Frontier markets.

TABLE 4.1 – Normality Statistics

The kurtosis statistics and the skewness statistics show the coefficients of the kurtosis and skewness tests for the composite hypothesis of normality (Shapiro et al., 1968), along with the p-values between square brackets, which are based on 2000 replications in the Monte Carlo simulation. The J-B statistic is the Jarque-Bera (1980) test for normality ($X^2(2)$ distributed).

		Kurtosis		**Skewness**		**Jarque-Bera (JB)**	
		Statistic	*p*-value	Statistic	*p*-value	Statistic	*p*-value
DEVELOPED MARKETS	USA	13.950	[0.000]	-0.356	[0.000]	14152.869	[0.000]
	Japan	8.129	[0.000]	-0.182	[0.000]	3107.324	[0.000]
	Hong Kong	10.963	[0.000]	-0.204	[0.000]	7472.650	[0.000]
	France	8.884	[0.000]	-0.052	[0.252]	4070.076	[0.000]
	United Kingdom	12.256	[0.000]	-0.238	[0.000]	10096.004	[0.000]
	Germany	8.680	[0.000]	-0.092	[0.045]	3796.488	[0.000]
	Canada	12.841	[0.000]	-0.723	[0.000]	11628.078	[0.000]
	Australia	10.990	[0.000]	-0.768	[0.000]	7781.768	[0.000]
	Netherlands	9.901	[0.000]	-0.165	[0.001]	5610.507	[0.000]
	Italy	8.713	[0.000]	-0.215	[0.000]	3857.568	[0.000]
EMERGING MARKETS	China	9.967	[0.000]	-0.021	[0.660]	5705.266	[0.000]
	India	11.668	[0.000]	0.071	[0.128]	8834.466	[0.000]
	South-Korea	23.873	[0.000]	-0.173	[0.000]	51222.815	[0.000]
	Brazil	10.395	[0.000]	-0.210	[0.000]	6448.798	[0.000]
	Mexico	10.242	[0.000]	-0.057	[0.206]	6165.912	[0.000]
	Malaysia	10.583	[0.000]	-0.445	[0.000]	6851.078	[0.000]
	Indonesia	9.840	[0.000]	-0.284	[0.000]	5537.199	[0.000]
	Poland	7.679	[0.000]	-0.243	[0.000]	2600.679	[0.000]
	Colombia	12.667	[0.000]	-0.305	[0.000]	11050.240	[0.000]
	Peru	9.740	[0.000]	-0.339	[0.000]	5393.371	[0.000]
FRONTIER MARKETS	Nigeria	48.030	[0.000]	-2.444	[0.000]	241149.333	[0.000]
	Kazakhstan	10.900	[0.000]	0.238	[0.000]	7361.998	[0.000]
	Croatia	11.982	[0.000]	-0.172	[0.000]	9496.661	[0.000]
	Romania	26.055	[0.000]	-1.468	[0.000]	63490.918	[0.000]
	Kenya	13.454	[0.000]	0.109	[0.016]	12851.690	[0.000]
	Tunisia	9.186	[0.000]	0.124	[0.010]	4504.841	[0.000]
	Lebanon	23.099	[0.000]	-0.033	[0.470]	47483.914	[0.000]
	Mauritius	16.955	[0.000]	0.231	[0.000]	22913.933	[0.000]
	Slovenia	9.005	[0.000]	-0.328	[0.000]	4289.767	[0.000]
	Estonia	8.388	[0.000]	0.109	[0.018]	3417.365	[0.000]

At last, the Jarque-Bera (1980) test is performed, which is a goodness-of-fit test that examines whether the daily returns have the skewness and kurtosis levels that match a normal distribution. According to the test statistics and the corresponding p-values, all markets reject the null hypothesis of normality at the one percent confidence level. These results are in line with the significant skewness and kurtosis coefficients. Overall, the statistics suggest that one should consider the implementation of non-normal distributions.

4.1.2 Autocorrelation

Before implementing ARMA and/or GARCH processes, the financial time series data needs to be tested for autocorrelation in the logarithmic returns and squared logarithmic returns. The presence of autocorrelation in logarithmic returns is tested by the Ljung & Box (1978) Q-statistic, whereas the presence of autocorrelation in the squared logarithmic returns is tested by implementing the Ljung-Box Q-statistic along with the Lagrange Multiplier (LM) test from Engle (1982), as is suggested by Zikovic (2007). If autocorrelation in the logarithmic returns is detected, it can be removed by fitting an ARMA(p, q) framework to the time series data. On the other hand, if autocorrelation is detected in the squared log returns, the ARCH effect should be removed by fitting a GARCH(p, q) model to the AMRA filtered data.

The Ljung-Box Q statistics for the logarithmic returns and the squared logarithmic returns are presented for up to 8 lags in Table [4.2], followed by their p-values. The null hypothesis of this test states that the first m autocorrelation coefficients of the time series are jointly zero, where m stands for the number of lags. The Q[8] statistics indicate that there is significant autocorrelation in the logarithmic returns for almost all markets at a 1 percent significance level, which means that ARMA processes should be considered. The Q²[8] statistics indicate on their part that there exists significant autocorrelation in the squared log returns for all markets at a 1 percent significance level. This supports the popular stylized fact on returns about

TABLE 4.2 – Autocorrelation Statistics

The values for Q(8) and Q²(8) denote the Ljung & Box (1980) Q-statistics on the first 8 lags of the sample autocorrelation function of logarithmic returns and squared logarithmic returns testing for autocorrelation and heteroscedasticity, respectively (distributed as $X^2(2)$). The value for ARCH(1-8) is the Lagrange Multiplier (LM) test of order 8 (Engle, 1982). The numbers in square brackets indicate exact significance levels. The number of lags is set to 8, following Tsay (2010) who sets lags equal to ln(T).

		Q[8]		**Q²[8]**		**ARCH[1-8]**	
		Q-statistic	*p*-value	Q-statistic	*p*-value	LM-statistic	*p*-value
DEVELOPED MARKETS	USA	48.639	[0.000]	2032.997	[0.000]	821.961	[0.000]
	Japan	57.535	[0.000]	1261.721	[0.000]	776.098	[0.000]
	Hong Kong	9.613	[0.293]	1634.443	[0.000]	547.129	[0.000]
	France	28.338	[0.000]	986.684	[0.000]	698.500	[0.000]
	United Kingdom	47.954	[0.000]	1321.717	[0.000]	434.958	[0.000]
	Germany	13.906	[0.084]	940.807	[0.000]	576.553	[0.000]
	Canada	81.470	[0.000]	2308.799	[0.000]	433.479	[0.000]
	Australia	19.509	[0.012]	2459.304	[0.000]	534.874	[0.000]
	Netherlands	33.705	[0.000]	1288.543	[0.000]	564.222	[0.000]
	Italy	29.962	[0.000]	619.929	[0.000]	297.530	[0.000]
EMERGING MARKETS	China	18.530	[0.018]	1760.827	[0.000]	705.289	[0.000]
	India	21.850	[0.005]	373.038	[0.000]	204.755	[0.000]
	South-Korea	19.458	[0.013]	695.347	[0.000]	501.637	[0.000]
	Brazil	15.615	[0.048]	1952.750	[0.000]	884.131	[0.000]
	Mexico	48.069	[0.000]	1810.837	[0.000]	713.611	[0.000]
	Malaysia	38.378	[0.000]	219.389	[0.000]	139.997	[0.000]
	Indonesia	47.821	[0.000]	575.281	[0.000]	276.786	[0.000]
	Poland	15.384	[0.052]	912.737	[0.000]	460.026	[0.000]
	Colombia	51.012	[0.000]	1613.057	[0.000]	677.959	[0.000]
	Peru	17.894	[0.022]	818.977	[0.000]	411.725	[0.000]
FRONTIER MARKETS	Nigeria	364.242	[0.000]	19.533	[0.012]	17.545	[0.025]
	Kazakhstan	11.946	[0.154]	927.797	[0.000]	424.056	[0.000]
	Croatia	70.567	[0.000]	1351.144	[0.000]	577.961	[0.000]
	Romania	15.753	[0.046]	39.681	[0.000]	30.895	[0.000]
	Kenya	397.568	[0.000]	1521.465	[0.000]	709.644	[0.000]
	Tunisia	37.017	[0.000]	225.449	[0.000]	147.714	[0.000]
	Lebanon	61.905	[0.000]	273.666	[0.000]	193.558	[0.000]
	Mauritius	72.221	[0.000]	602.744	[0.000]	396.308	[0.000]
	Slovenia	41.684	[0.000]	1383.523	[0.000]	637.843	[0.000]
	Estonia	14.908	[0.061]	688.621	[0.000]	321.443	[0.000]

positive dependence between squared returns on nearby days, and justifies the use of GARCH-type models for forecasting volatility. However, it is common to confirm possible ARCH effects by incorporating the Lagrange Multiplier (LM) test. This test considers the null hypothesis of no ARCH errors versus the alternative hypothesis that the conditional error variance is given by an ARCH process. Overall, the ARCH[1-8] test statistics from Table [4.2] show clear evidence of a time varying phenomenon in the conditional volatility at the 1 percent significance level for all markets.

A crucial next step is to identify the optimal orders for ARMA(p, q) and GARCH(p, q) such that a parsimonious model can be built that is capable of capturing as much variation in the data as possible. In fact, Brorsen and Yang (1993) argue that one of the challenges associated with GARCH is identifying the proper lag structure. In practice, academics tend to follow the recommendation of Bollerslev (1986) by working solely with first orders. He argues that these model specifications are strong enough and capable of delivering relatively accurate results. Nonetheless, this study will let the data determine the appropriate lag structure, despite the added computational effort. This study will include both AIC and BIC. These selection criterion do not only reward goodness of fit, but also include a penalty term in the form of an increasing function of the number of estimated parameters. If AIC and BIC give conflicting results, the BIC is used as the primary selection criterion, as BIC will always select a more parsimonious model (Cappiello et al., 2006). In this study, for both ARMA(p, q) and GARCH(p, q), the orders $p = 0,..,4$ and $q = 0, ..., 4$ are considered, making up for a total selection of 625 models per index. Table [4.3] shows the most appropriate order structures of ARMA and GARCH for each individual market. Concerning the ARMA models, the orders vary between 0 and 2. For the GARCH models, the most popular structure is indeed the (1,1) as suggested by Bollerslev (1986). However, there is a noteworthy quantity of stock market indices that seem to demand a different lag structure, highlighting the importance of a decent order selection procedure.

TABLE 4.3 – Optimal ARMA-GARCH Order Structures

The p and q orders for ARMA denote the optimal orders for the AR and MA processes, respectively. The p and q orders for GARCH denote the optimal orders for the ARCH and GARCH processes, respectively. For both ARMA(p, q) and GARCH(p, q), the orders $p = 0,..,4$ and $q = 0, ..., 4$ are considered, making up for a total selection of 625 models. The optimal orders are selected by choosing the lowest value for BIC (Schwarz, 1978).

DEVELOPED MARKETS					EMERGING MARKETS					FRONTIER MARKETS				
	ARMA		GARCH			ARMA		GARCH			ARMA		GARCH	
	p	q	p	q		p	q	p	q		p	q	p	q
USA	0	1	2	1	China	0	1	1	1	Nigeria	1	0	1	1
Japan	0	1	1	1	India	1	0	1	1	Kazakhstan	0	1	1	1
Hong Kong	0	0	1	1	South-Korea	1	1	1	1	Croatia	1	0	1	2
France	0	0	2	1	Brazil	0	2	1	1	Romania	1	0	1	2
United Kingdom	0	0	1	1	Mexico	1	0	1	2	Kenya	0	1	1	2
Germany	0	0	1	1	Malaysia	0	2	1	4	Tunisia	0	2	1	4
Canada	2	2	1	1	Indonesia	0	0	1	1	Lebanon	0	0	1	1
Australia	0	0	1	1	Poland	0	0	1	1	Mauritius	0	1	1	2
Netherlands	0	0	1	1	Colombia	1	0	1	1	Slovenia	0	1	1	1
Italy	0	0	1	1	Peru	1	1	1	1	Estonia	0	0	1	1

4.1.3 Stationarity

"Past returns do not necessarily reflect future performance". This warning can be found everywhere on brochures describing various investments and funds. However, the most basic requirement of any statistical analysis of financial time series data is the existence of some statistical properties which remains stable over time, otherwise it is pointless to try to identify them. In the context of this study, the application of processes such as ARMA and GARCH is conditioned by stationarity of the financial time series. Therefore, each market has been tested by the Augmented Dickey Fuller (ADF) test (Dickey & Fuller, 1981) test for the presence of unit root, which would imply that the particular time series is not stationarity. Looking at the results from Table [4.4], it is clear that the null hypothesis of unit root is rejected in favor of the stationary alternative for all markets, which effectively means that our data can be used for modeling.

4.2 EVT Threshold Choice

The choice of the threshold u is of high importance. Unfortunately, there is some difficulty in choosing the appropriate threshold in the Peaks-over-Threshold approach. Coles et al. (2001) argue that the selection of the threshold is always a trade-off between bias and variance. A low threshold value means more observations to use when fitting a GPD distribution to the data, however, too low a value for the threshold gives the risk of including observations that are not far enough in the tail for the POT approach to be valid, which in turn could lead to biased parameter estimations (Kjelsson, 2013). Contrarily, choosing too high a value for the threshold will generate fewer excesses to estimate the shape and scale parameter upon, which may lead to high variance and greater estimation uncertainty.

Over the years, many diagnostic threshold procedures have been proposed, such as the mean residual life plot and the parameter stability plot (Coles et al., 2001). Despite the popularity of these methods in the majority of existing literature, it is worth mentioning that an associated drawback of these methods is that inspecting graphs in order to select the correct threshold value suffers from substantial subjective elements. In other words, different practitioners could choose different thresholds while using the exact same underlying time series. Another drawback of the traditional methods is their incapability of adapting to changing markets. Unfortunately, it would simply be impossible timewise to re-estimate the threshold value at each step in time of every individual asset by inspecting plots. A solution for this matter of contention is found in a relatively unknown and underused method called the 'double bootstrap approach' of Danielsson et al. (2001), which is best described as a computationally intensive procedure that automates the search for the appropriate threshold value. The automatic algorithm enables a frequent re-estimation of the threshold value and simultaneously ends any doubts regarding the subjectivity of the results. This study rented several serves from Amazon Web Services in order to cope with the required processing powers. This opened up the possibility of having daily re-estimations of the threshold value for all 18 EVT models in all 30 markets.

TABLE 4.4 – Stationarity Statistics

The provided ADF statistics are from the Augmented Dickey and Fuller (1981) test. The lag length of the ADF is set to $(T-1)^{1/3}$, where T is the length of the total sample. In this study, the number of lags is set equal to 14. The numbers in square brackets indicate the exact significance levels.

DEVELOPED MARKETS			EMERGING MARKETS			FRONTIER MARKETS		
	ADF-statistic	*p*-value		ADF-statistic	*p*-value		ADF-statistic	*p*-value
USA	-18.882	[0.000]	China	-17.194	[0.000]	Nigeria	-16.546	[0.000]
Japan	-19.617	[0.000]	India	-16.437	[0.000]	Kazakhstan	-18.079	[0.000]
Hong Kong	-17.645	[0.000]	South-Korea	-18.116	[0.000]	Croatia	-17.175	[0.000]
France	-19.008	[0.000]	Brazil	-19.167	[0.000]	Romania	-15.909	[0.000]
United Kingdom	-18.995	[0.000]	Mexico	-18.906	[0.000]	Kenya	-17.600	[0.000]
Germany	-18.640	[0.000]	Malaysia	-17.141	[0.000]	Tunisia	-18.137	[0.000]
Canada	-18.734	[0.000]	Indonesia	-18.393	[0.000]	Lebanon	-17.101	[0.000]
Australia	-18.439	[0.000]	Poland	-17.992	[0.000]	Mauritius	-15.336	[0.000]
Netherlands	-17.967	[0.000]	Colombia	-17.967	[0.000]	Slovenia	-18.144	[0.000]
Italy	-18.184	[0.000]	Peru	-17.796	[0.000]	Estonia	-17.487	[0.000]

4.3 Window Selection

Following Section 4.1, the full sample T (2821 observations) will be divided into an 'in-sample' period of length S (1321 observations) and an 'out-of-sample' period of length H (1500 observations). The first step of the estimation process would be to estimate the model parameters of ARMA-GARCH and CEVT over the 'in-sample' period. The next step has proven to be more difficult due to the fact that there is a clear dichotomy in literature on the correct method of forecasting. Marcellino et al. (2006) state that the recursive method of forecasting is usually employed in backtesting procedures, which is an approach that anchors the starting point (i.e. 'in-sample' period) and continually adds one observation each day. However, this would imply that the length of the 'in-sample' period would eventually increase to a level that is no longer optimal according to various empirical literature (see e.g. Frey & Michaud, 1997; Angelidis et al., 2004). To overcome this problem, this study will implement the more statisticall sound method of rolling window, which fixes the length of the 'in-sample' period and rolls up one observation at a time.

4.4 Backtesting Value at Risk

4.4.1 Conditional Coverage Test

The Conditional Coverage (CC) test of Christoffersen (1998) is a joint test that combines the Unconditioanl Coverage (UC) test of Kupiec (1995) and the Independence (IND) test of Christoffersen (1998). The UC test evaluates whether the specified risk model captures a number of VaR violations that is consistent with the chosen level of confidence. For example, if daily VaR forecasts are computed at a 99 percent level of confidence for the interval of one year, one would expect to see 3 violations on average. In order to test for the statistical validity of each risk model, a likelihood ratio (LR) is conducted

$$LR_{UC} = -2\ln[\frac{(1-p)^{n_0}(p)^{n_1}}{(1-\hat{p})^{n_0}(\hat{p})^{n_1}}] \qquad [4.1]$$

where n_0 and n_1 denote the number of non-violations and violations, respectively, and p and $\hat{p}$ represent the violation rate from the chosen confidence level and the observed violation rate, respectively. The test is chi-square distributed with one degree of freedom. If the LR_{UC} exceeds the critical value, the risk model is deemed to be inaccurate. By design, the test penalizes both underestimation and overestimation of risk, i.e. it rejects models that have either a too high or a too low amount of violations.

An important shortcoming of the UC test would be that it ignores whether violations appear in clusters. Repeated severe capital losses could lead to solvency issues and eventually bankruptcy for the financial instiution. This phenemon of multiple VaR exceedances in a row is often referred to as violation clustering. The IND test of Christoffersen (1998) tackles this problem by verifying if the violations are independent of one another. The test implements the following likelihood ratio (LR) statistic

$$LR_{IND} = -2\ln[\frac{(1-\hat{p})^{n_{00}+n_{10}}(\hat{p})^{n_{01}+n_{11}}}{(1-\hat{p}_0)^{n_{00}}\hat{p}_0^{n_{01}}(1-\hat{p}_1)^{n_{10}}\hat{p}_1^{n_{11}}}] \qquad [4.2]$$

where $\hat{p}_0$ denotes the probability of tomorrow being a day with a violation, conditional on today being a day with violation, and $\hat{p}_1$ denotes the probability that tomorrow has a violation, given that today has no violation. The test is asymptotically chi-square distributed with one degree of freedom. If the LR_{ind} exceeds the critical value, the null hypothesis is rejected and the model is thought to have independence problems.

As mentioned, Christoffersen (1998) then created a joint test by combining the UC test and the IND test into the CC test.

$$LR_{cc} = LR_{uc} + LR_{ind} \qquad [4.3]$$

The test is asymptotically chi-square distributed and has two degrees of freedom. Again, if the LR_{cc} exceeds the critical value, the null hypothesis will be rejected. If this is the case, the model either overestimates risk or underestimates risk and/or generates too many clustered violations. An important side note comes from Christoffersen (2003), who argues that the number of observations, and even more so the number of violations, may in practice be too small for this test to be reliable. This can especially be the case for high confidence levels. To overcome this problem, he recommends doing a Monte Carlo simulation to obtain more reliable p-values.

4.4.3 Duration Test

Although the CC test is capable of providing a parsimonious procedure for model evaluation, it should be noted that this test assumes first-order Markov property. This means that only today's violation matters for tomorrow's violation. The consequence is that if the violation sequence exhibits a dependence structure other than first-order Markov dependence, the test would fail to detect it. In order to overcome this drawback, Christoffersen & Pelletier (2004) suggest a Duration test that focusses on the duration of violations rather than the sequence of violations. The motivation behind this approach is that if there is dependence in the hit sequence, there would be an excess of relatively short no-hit durations and an excess of relatively long no-hit durations. Christoffersen & Pelletier (2004) explain this as "if the one-day-ahead VaR is correctly specified for coverage rate, p, then, every day, the conditional expected duration until the next violation should be a constant $1/p$ days". The Duration test focusses on the duration between two sequential VaR violations. Under the null-hypothesis of a correctly specified model, the no-hit duration should have no memory.

4.4.4 "González-Rivera, Lee, and Mishra" – Loss Function

All the approaches mentioned so far are highly relevant when judging the statistical adequacy of the individual Value at Risk series, however, they fail to discriminate between models on the basis of their predictive accuracy. In other words, they fail to answer the question whether there is any statistical difference between the forecasting performance of the different models. To overcome this problem, González-Rivera et al. (2004) designed an asymmetric VaR loss function, which enables the comparison of the performance of various volatility models on the basis of a statistical loss function. The asymmetric VaR loss function of González-Rivera et al. (2004) is defined as

$$\ell\left(r_{t+1}, VaR^{\tau}_{j,t+1|t}\right) = T_0^{-1}\rho_\tau\left(r_{t+1} - VaR^{\tau}_{j,t+1|t}\right), t = 1,2,\dots,T_0 \qquad [4.4]$$

where $j = 1,2,\dots,m$ is the model indicator, r_{t+1} is the return at time $t+1$, $VaR^{\tau}_{j,t+1|t}$ denotes the predicted VaR at time $t+1$ given the information up to time t, $\rho_\tau = z(\tau - \mathrm{I}_{-\infty,0}(z))$ is the τ-th quantile loss function, and T_0 is the length of the out-of-sample. By design, the asymmetric VaR loss function penalizes observations below the τ-th quantile level more heavily than observations above. Models can then be compared based upon their loss function value: low values speak in favor of a model, while high values do the opposite.

4.4.5 ‘Two-Stage Backtesting’ Procedure

In order to select the best model among the competing forecasting methods, this study implements a ‘two-stage backtesting’ procedure. The first stage will consist of two statistical tests that are used to examine the statistical accuracy of the risk models: the CC test and the DUR test. There are three ways in which a model could end up being rejected: (i.) In the scenario of too many violations, which would cause too low allocation of risk-taking capital in the portfolio, (ii.) In the scenario of too few violations, which would cause opportunity costs and inefficient allocation of capital, and (iii.) In het scenario of clustered violations, which would expose the risk taking vendor to a high threat of a default. The models that fail to pass the first stage are said to lack statistical soundess and should therefore not be incorporated into practice. The second stage of the backtesting procedure looks at the remaining models and investigates whether there are statistical differences between these models in terms of forecasting accuracy. In order to distinguish accurate models from loose models, this stage incorporates the asymmetric loss function of González-Rivera et al. (2004).

4.5 Backtesting Expected Shortfall

As a result of the complexity and novelty of Expected Shortfall, the available literature on its backtesting framework is still relatively small when compared to that of Value at Risk. Nonetheless, a backtest that is deemed to be sufficient and accurate is the Expected Shortfall Bootstrap test of McNeil & Frey (2000), which allows for a statistical ranking between competing forecasting models. It looks at the difference between the next-day return r_{t+1} and the ES estimate at time t, $ES_{\alpha,t+1|t}(r_{t+1})$, conditional on that the realized return r_{t+1} exceeds the VaR estimate, $VaR_{\alpha,t+1|t}(r_{t+1})$. Based on the financial time series and the forecasting estimates of ES, the corresponding residuals can be constructed on days when VaR violations take place. In other words, these residuals measure the discrepancy between the realized losses and the ES estimates on days when VaR violations took place. Following the paper of McNeil & Frey (2000), these are called “exceedance residuals”:

$$x_{t+1} = \frac{r_{t+1} - ES_{\alpha,t+1|t}(r_{t+1})}{\sigma_{t+1}} \qquad [4.5]$$

The ES Bootstrap test incorporates the null hypothesis that μ_{t+1}, σ_{t+1} and the ES are estimated correctly, which would mean that these “exceedance residuals” behave like an i.i.d. sample with mean zero. In order to compare different models in their forecasting ability, the backtest delivers a value between zero and one, where high values speak in favor of a model and low values do the opposite.

(5) EMPIRICAL RESULTS

This Chapter covers the empirical results for both Value at Risk (99%) and Expected Shortfall (97.5%). Section 5.1.1 treats the results for the conditional coverage test and the duration test, respectively, and distinguishes statistically sound VaR models from their inappropriate rivals. Section 5.1.2 implements the second stage of the VaR backtesting procedure, which ranks the appropriate models by their scores in the applied loss function and tests "Hypothesis One". Section 5.2 is devoted to the implementation of the ES bootstrap test and the testing of "Hypothesis Two". By dividing the different specifications into three groups, Section 5.1.2 and Section 5.2 exhibit a similar structure: (i.) the GARCH framework against the conditional EVT framework, (ii.) the relative performance of the Normal distribution, the Student-t distribution, and the JSU distribution, and (iii.) the leverage specification against the non-leverage specification. In order to provide a comprehensive answer on the research question and the two formulated hypotheses, a special focus is put on the explanations and potential implications of observed discrepancies and similarities between the three market classification groups.

5.1 Backtesting Value at Risk (99%)

The first part of this study is devoted to the determination of the 'optimal' Value at Risk model for each country index when forecasting at the 99 percent level of confidence. To achieve this goal, a large quantity of financial risk models is estimated for each specific market, and based on the criteria specified in Section 5.1.1, these models are utilized to arrive at the 'best' risk model in Section 5.1.2.

5.1.1 Stage One of the VaR (99%) Backtesting Procedure

This section carries out the first stage of the VaR backtesting procedure, which is dedicated towards the crucial process of distinguishing statistically sound VaR models from their inappropriate rivals. In order to accomplish such a separation, this stage includes the implementation of two decisive tests, namely the conditional coverage test of Christoffersen (1998) and the duration test of Christoffersen & Pelletier (2004). The first test is primarily a check on whether the frequency of violations is in line with the selected confidence level, which is in this study set equal to the BCBS standard of 99 percent. A violation rate that is greater than the expected rate would imply that the risk model induces insufficient capital allocation. On the other hand, an lower violation rate would imply that the risk model signals the need of a capital allocation that is greater than necessary. This highlights the importance of incorporating an interval for the amount of

TABLE 5.1 [A] - Stage 1 of the Value at Risk Backtesting Procedure – DEVELOPED markets

This table summarizes the Monte Carlo *p*-values of the Conditional Coverage (CC) test (Christoffersen, 1998) and the normal *p*-values of the Duration (DUR) test (Christoffersen & Pelletier, 2004) for each market within the *Developed* market classification. The bold font and the green cells indicate that the respective models have passed both tests at the 5 percent level of significance for the specific market. The abbreviations n, s, j, and c in the first column stand for Normal Distribution, Student-*t* Distribution, Johnson SU Distribution, and Conditional Extreme Value Theory. In the column of 'CC', the figure $^{-}$ indicates underestimation of risk (i.e. too many violations) and the figure $^{+}$ indicates overestimation of risk (i.e. too few violations).

	USA		Japan		Hong Kong		France		UK		Germany		Canada		Australia		Netherlands		Italy	
	CC	DUR	CC	DUR	CC	DUR	CC	DUR	CC	DUR	CC	DUR	CC	DUR	CC	DUR	CC	DUR	CC	DUR
n-GARCH	0.002^{-}	0.387	0.001^{-}	0.138	0.034^{-}	0.777	0.005^{-}	0.135	0.002^{-}	0.037	0.028^{-}	0.280	0.001^{-}	0.982	0.081	0.245	0.246	0.442	0.895	0.408
n-IGARCH	0.002^{-}	0.272	0.001^{-}	0.337	0.019^{-}	0.960	0.017^{-}	0.034	0.001^{-}	0.090	0.038^{-}	0.133	0.002^{-}	0.988	0.093	0.441	0.564	0.710	0.816	0.211
n-EGARCH	0.040^{-}	0.804	0.002^{-}	0.419	0.078	0.463	0.059	0.378	0.060	0.727	0.069	0.216	0.016^{-}	0.144	0.222	0.672	0.740	0.612	0.663	0.459
n-GJRGARCH	0.003^{-}	0.522	0.006^{-}	0.981	0.058	0.547	0.034^{-}	0.110	0.019^{-}	0.919	0.052	0.171	0.027^{-}	0.119	0.095	0.287	0.261	0.050	0.911	0.479
n-TGARCH	0.131	0.629	0.006^{-}	0.609	0.076	0.949	0.110	0.031	0.055	0.952	0.055	0.404	0.027^{-}	0.905	0.166	0.703	0.730	0.506	0.652	0.459
n-NAGARCH	0.073	0.365	0.001^{-}	0.784	0.114	0.660	0.086	0.357	0.074	0.972	0.187	0.341	0.023^{-}	0.715	0.149	0.505	0.577	0.523	0.667	0.459
s-GARCH	0.031^{-}	0.576	0.002^{-}	0.081	0.747	0.399	0.334	0.050	0.209	0.176	0.584	0.149	0.056	0.459	0.121	0.599	0.247	0.442	0.910	0.408
s-IGARCH	0.040^{-}	0.881	0.004^{-}	0.091	0.923	0.287	0.579	0.177	0.218	0.226	0.742	0.144	0.065	0.448	0.179	0.616	0.589	0.710	0.824	0.211
s-EGARCH	0.831	0.459	0.042^{-}	0.943	0.917	0.090	0.489	0.090	0.916	0.102	0.727	0.033	0.148	0.402	0.929	0.164	0.740	0.612	0.639	0.459
s-GJRGARCH	0.919	0.048	0.109	0.983	0.745	0.399	0.455	0.001	0.163	0.950	0.317	0.057	0.187	0.102	0.902	0.310	0.240	0.050	0.927	0.479
s-TGARCH	0.642	0.417	0.033^{-}	0.792	0.909	0.031	0.581	0.084	0.606	0.830	0.742	0.033	0.453	0.156	0.574	0.194	0.762	0.506	0.665	0.459
s-NAGARCH	0.833	0.417	0.064	0.580	0.583	0.501	0.741	0.312	0.597	0.419	0.906	0.312	0.153	0.579	0.918	0.404	0.584	0.523	0.680	0.459
j-GARCH	0.040^{-}	0.210	0.040^{-}	0.019	0.820	0.721	0.647	0.143	0.214	0.597	0.891	0.085	0.655	0.911	0.412	0.686	0.257	0.442	0.899	0.408
j-IGARCH	0.187	0.608	0.009^{-}	0.172	0.481	0.253	0.384	0.464	0.225	0.597	0.918	0.085	0.631	0.911	0.420	0.721	0.579	0.710	0.838	0.211
j-EGARCH	0.136	0.433	0.173	0.255	0.646	0.053	0.907	0.200	0.080	0.337	0.645	0.072	0.514	0.147	0.128	0.958	0.736	0.612	0.665	0.459
j-GJRGARCH	0.275	0.318	0.341	0.359	0.802	0.120	0.757	0.017	0.394	0.070	0.832	0.040	0.516	0.147	0.134	0.525	0.235	0.050	0.928	0.479
j-TGARCH	0.014^{+}	0.594	0.220	0.458	0.899	0.031	0.908	0.176	0.051	0.424	0.667	0.072	0.490	0.147	0.138	0.958	0.723	0.506	0.659	0.459
j-NAGARCH	0.026^{+}	0.541	0.346	0.252	0.921	0.204	0.651	0.401	0.084	0.310	0.522	0.143	0.640	0.184	0.384	0.572	0.588	0.523	0.637	0.459
c-n-GARCH	0.064	0.977	0.079	0.498	0.026^{+}	0.240	0.138	0.420	0.031^{+}	0.063	0.385	0.008	0.295	0.527	0.017^{+}	0.164	0.042^{+}	0.415	0.159	0.182
c-n-IGARCH	0.061	0.907	0.385	0.488	0.136	0.241	0.151	0.420	0.069	0.145	0.379	0.012	0.138	0.973	0.011^{+}	0.164	0.146	0.479	0.120	0.182
c-n-EGARCH	0.026^{+}	0.460	0.065	0.490	0.145	0.814	0.132	0.364	0.003^{+}	0.665	0.292	0.647	0.274	0.056	0.023^{+}	0.933	0.271	0.475	0.370	0.470
c-n-GJRGARCH	0.008^{+}	0.630	0.093	0.392	0.295	0.156	0.382	0.316	0.002^{+}	0.114	0.096	0.171	0.070	0.370	0.012^{+}	0.195	0.264	0.475	0.137	0.005
c-n-TGARCH	0.026^{+}	0.672	0.029^{+}	0.639	0.120	0.814	0.402	0.649	0.006^{+}	0.665	0.141	0.353	0.286	0.088	0.012^{+}	0.933	0.039^{+}	0.388	0.396	0.470
c-n-NAGARCH	0.015^{+}	0.911	0.063	0.976	0.072	0.699	0.075	0.188	0.007^{+}	0.665	0.140	0.208	0.284	0.717	0.017^{+}	0.195	0.031^{+}	0.388	0.376	0.470
c-s-GARCH	0.078	0.977	0.293	0.747	0.069	0.550	0.147	0.420	0.020^{+}	0.063	0.393	0.012	0.075	0.780	0.018^{+}	0.164	0.040^{+}	0.415	0.074	0.228
c-s-IGARCH	0.082	0.907	0.125	0.764	0.126	0.241	0.138	0.420	0.077	0.145	0.526	0.031	0.136	0.973	0.017^{+}	0.164	0.132	0.479	0.132	0.182
c-s-EGARCH	0.029^{+}	0.460	0.044^{+}	0.773	0.150	0.814	0.127	0.364	0.005^{+}	0.665	0.290	0.647	0.275	0.056	0.012^{+}	0.933	0.149	0.889	0.412	0.470
c-s-GJRGARCH	0.004^{+}	0.630	0.043^{+}	0.159	0.419	0.183	0.530	0.370	0.014^{+}	0.772	0.086	0.171	0.066	0.697	0.011^{+}	0.195	0.088	0.877	0.135	0.005
c-s-TGARCH	0.013^{+}	0.594	0.024^{+}	0.639	0.145	0.814	0.272	0.562	0.004^{+}	0.665	0.147	0.353	0.090	0.166	0.006^{+}	0.933	0.048^{+}	0.388	0.361	0.470
c-s-NAGARCH	0.022^{+}	0.541	0.068	0.976	0.143	0.814	0.048^{+}	0.045	0.004^{+}	0.665	0.131	0.208	0.149	0.276	0.011^{+}	0.195	0.048^{+}	0.388	0.400	0.470
c-j-GARCH	0.051	0.876	0.067	0.217	0.075	0.550	0.141	0.420	0.038^{+}	0.048	0.401	0.012	0.086	0.780	0.015^{+}	0.164	0.043^{+}	0.415	0.125	0.182
c-j-IGARCH	0.034^{+}	0.876	0.263	0.625	0.136	0.241	0.131	0.420	0.045^{+}	0.031	0.490	0.031	0.139	0.973	0.020^{+}	0.164	0.139	0.479	0.092	0.228
c-j-EGARCH	0.033^{+}	0.460	0.060	0.490	0.138	0.814	0.084	0.466	0.005^{+}	0.665	0.289	0.647	0.095	0.227	0.025^{+}	0.933	0.122	0.889	0.400	0.470
c-j-GJRGARCH	0.002^{+}	0.630	0.041^{+}	0.159	0.393	0.183	0.488	0.370	0.003^{+}	0.665	0.064	0.171	0.041^{+}	0.922	0.016^{+}	0.195	0.136	0.292	0.127	0.005
c-j-TGARCH	0.015^{+}	0.594	0.035^{+}	0.713	0.133	0.814	0.255	0.562	0.002^{+}	0.665	0.134	0.353	0.069	0.202	0.017^{+}	0.933	0.071	0.877	0.380	0.470
c-j-NAGARCH	0.021^{+}	0.541	0.057	0.976	0.130	0.814	0.079	0.188	0.005^{+}	0.665	0.124	0.208	0.043^{+}	0.362	0.011^{+}	0.195	0.040^{+}	0.388	0.383	0.470

TABLE 5.1 [B] - Stage 1 of the Value at Risk Backtesting Procedure – EMERGING markets

This table summarizes the Monte Carlo *p*-values of the Conditional Coverage (CC) test (Christoffersen, 1998) and the normal *p*-values of the Duration (DUR) test (Christoffersen & Pelletier, 2004) for each market within the *Emerging* market classification. The bold font and the green cells indicate that the respective models have passed both tests at the 5 percent level of significance for the specific market. The abbreviations n, s, j, and c in the first column stand for Normal Distribution, Student-*t* Distribution, Johnson SU Distribution, and Conditional Extreme Value Theory. In the column of 'CC', the figure $^{-}$ indicates underestimation of risk (i.e. too many violations) and the figure $^{+}$ indicates overestimation of risk (i.e. too few violations).

	China		India		South-Korea		Brazil		Mexico		Malaysia		Indonesia		Poland		Colombia		Peru	
	CC	DUR	CC	DUR	CC	DUR	CC	DUR	CC	DUR	CC	DUR	CC	DUR	CC	DUR	CC	DUR	CC	DUR
n-GARCH	0.006^{-}	0.309	0.112	0.539	0.220	0.895	0.047^{-}	0.473	0.023^{-}	0.836	0.251	0.875	0.002^{-}	0.294	0.588	0.496	0.324	0.013	0.388	0.796
n-IGARCH	0.005^{-}	0.498	0.101	0.191	0.104	0.660	0.193	0.567	0.030^{-}	0.727	0.341^{+}	0.257	0.002^{-}	0.324	0.927^{+}	0.106	0.399	0.030	0.068^{+}	0.171
n-EGARCH	0.023^{-}	0.852	0.247	0.825	0.431	0.730	0.034^{-}	0.890	0.329	0.699	0.591	0.506	0.001^{-}	0.217	0.464	0.303	0.673	0.173	0.495	0.499
n-GJRGARCH	0.013^{-}	0.658	0.755	0.197	0.449	0.450	0.037^{-}	0.981	0.115	0.019	0.119	0.801	0.002^{-}	0.106	0.406	0.223	0.830	0.772	0.261	0.050
n-TGARCH	0.032^{-}	0.981	0.456	0.609	0.561	0.819	0.027^{-}	0.696	0.255	0.150	0.473	0.604	0.002^{-}	0.098	0.515	0.303	0.273	0.913	0.134	0.085
n-NAGARCH	0.020^{-}	0.738	0.217	0.294	0.163	0.359	0.035^{-}	0.828	0.745	0.696	0.568	0.847	0.003^{-}	0.072	0.659	0.185	0.384	0.546	0.138	0.085
s-GARCH	0.723	0.143	0.814	0.855	0.131	0.721	0.447	0.979	0.730	0.947	0.915	0.764	0.338	0.720	0.583	0.496	0.308	0.013	0.407	0.796
s-IGARCH	0.366	0.649	0.830	0.855	0.645	0.036	0.922	0.844	0.244	0.808	0.903	0.686	0.753	0.572	0.905	0.106	0.366	0.030	0.066	0.171
s-EGARCH	0.753	0.966	0.813	0.372	0.814	0.597	0.598	0.708	0.515	0.064	0.280	0.851	0.908	0.251	0.515	0.303	0.666	0.173	0.482	0.499
s-GJRGARCH	0.735	0.549	0.505	0.387	0.393	0.760	0.125	0.692	0.798	0.361	0.372	0.712	0.128	0.321	0.366	0.223	0.852	0.772	0.272	0.050
s-TGARCH	0.911	0.729	0.795	0.372	0.654	0.905	0.726	0.279	0.414	0.120	0.272	0.851	0.670	0.207	0.504	0.303	0.282	0.913	0.122	0.085
s-NAGARCH	0.804	0.696	0.259	0.574	0.495	0.699	0.373	0.154	0.385	0.337	0.640	0.569	0.528	0.435	0.653	0.185	0.402	0.546	0.140	0.085
j-GARCH	0.588	0.259	0.518	0.578	0.122	0.910	0.668	0.866	0.404	0.613	0.279	0.800	0.275	0.747	0.579	0.496	0.325	0.013	0.386	0.796
j-IGARCH	0.909	0.532	0.525	0.968	0.130	0.910	0.136	0.900	0.826	0.727	0.087	0.544	0.051	0.477	0.901	0.106	0.385	0.030	0.095	0.171
j-EGARCH	0.761	0.966	0.821	0.372	0.263	0.430	0.162	0.394	0.085	0.429	0.079	0.523	0.290	0.317	0.531	0.303	0.651	0.173	0.529	0.499
j-GJRGARCH	0.757	0.549	0.281	0.902	0.277	0.867	0.848	0.545	0.132	0.645	0.142	0.964	0.304	0.860	0.368	0.223	0.816	0.772	0.272	0.050
j-TGARCH	0.916	0.729	0.652	0.294	0.069	0.876	0.078	0.478	0.084	0.429	0.051	0.043	0.396	0.264	0.511	0.303	0.277	0.913	0.129	0.085
j-NAGARCH	0.803	0.696	0.285	0.574	0.082	0.876	0.277	0.234	0.079	0.429	0.148	0.980	0.075	0.876	0.659	0.185	0.389	0.546	0.150	0.085
c-n-GARCH	0.401	0.284	0.020^{+}	0.279	0.016^{+}	0.958	0.011^{+}	0.559	0.012^{+}	0.923	0.036^{+}	0.322	0.005^{+}	0.538	0.037^{+}	0.396	0.040^{+}	0.879	0.035^{+}	0.772
c-n-IGARCH	0.288	0.801	0.142	0.544	0.015^{+}	0.958	0.004^{+}	0.183	0.028^{+}	0.169	0.019^{+}	0.812	0.016^{+}	0.474	0.048^{+}	0.396	0.045^{+}	0.879	0.019^{+}	0.816
c-n-EGARCH	0.395	0.889	0.048^{+}	0.114	0.002^{+}	0.420	0.001^{+}	0.724	0.026^{+}	0.944	0.027^{+}	0.812	0.048^{+}	0.570	0.039^{+}	0.117	0.051	0.639	0.022^{+}	0.088
c-n-GJRGARCH	0.678	0.239	0.024^{+}	0.134	0.025^{+}	0.992	0.002^{+}	0.724	0.014^{+}	0.923	0.024^{+}	0.049	0.003^{+}	0.538	0.033^{+}	0.117	0.023^{+}	0.772	0.014^{+}	0.088
c-n-TGARCH	0.407	0.889	0.022^{+}	0.134	0.002^{+}	0.227	0.002^{+}	0.724	0.008^{+}	0.531	0.014^{+}	0.167	0.044^{+}	0.570	0.044^{+}	0.117	0.019^{+}	0.772	0.004^{+}	0.502
c-n-NAGARCH	0.531	0.803	0.022^{+}	0.134	0.002^{+}	0.227	0.001^{+}	0.724	0.001^{+}	0.531	0.012^{+}	0.167	0.028^{+}	0.267	0.087	0.363	0.047^{+}	0.982	0.026^{+}	0.816
c-s-GARCH	0.291	0.801	0.020^{+}	0.279	0.018^{+}	0.958	0.017^{+}	0.559	0.012^{+}	0.923	0.027^{+}	0.812	0.005^{+}	0.538	0.052	0.396	0.044^{+}	0.879	0.019^{+}	0.816
c-s-IGARCH	0.283	0.801	0.088	0.732	0.016^{+}	0.958	0.002^{+}	0.183	0.020^{+}	0.169	0.028^{+}	0.812	0.019^{+}	0.474	0.037^{+}	0.396	0.047^{+}	0.879	0.040^{+}	0.816
c-s-EGARCH	0.661	0.913	0.032^{+}	0.134	0.005^{+}	0.402	0.002^{+}	0.724	0.015^{+}	0.944	0.005^{+}	0.047	0.027^{+}	0.884	0.044^{+}	0.117	0.037^{+}	0.639	0.013^{+}	0.088
c-s-GJRGARCH	0.658	0.239	0.016^{+}	0.134	0.015^{+}	0.992	0.001^{+}	0.457	0.006^{+}	0.923	0.017^{+}	0.167	0.003^{+}	0.538	0.044^{+}	0.117	0.019^{+}	0.772	0.017^{+}	0.088
c-s-TGARCH	0.507	0.918	0.021^{+}	0.134	0.003^{+}	0.227	0.004^{+}	0.724	0.006^{+}	0.531	0.008^{+}	0.167	0.044^{+}	0.570	0.043^{+}	0.117	0.056	0.639	0.003^{+}	0.502
c-s-NAGARCH	0.297	0.237	0.032^{+}	0.134	0.007^{+}	0.420	0.002^{+}	0.457	0.005^{+}	0.531	0.012^{+}	0.167	0.037^{+}	0.570	0.074	0.363	0.044^{+}	0.982	0.022^{+}	0.816
c-j-GARCH	0.144	0.421	0.011^{+}	0.279	0.017^{+}	0.958	0.015^{+}	0.559	0.016^{+}	0.923	0.024^{+}	0.812	0.002^{+}	0.538	0.051	0.396	0.039^{+}	0.879	0.054	0.772
c-j-IGARCH	0.290	0.801	0.083	0.732	0.014^{+}	0.958	0.005^{+}	0.183	0.013^{+}	0.923	0.033^{+}	0.812	0.018^{+}	0.474	0.037^{+}	0.396	0.030^{+}	0.879	0.030^{+}	0.816
c-j-EGARCH	0.671	0.673	0.028^{+}	0.134	0.004^{+}	0.420	0.002^{+}	0.724	0.011^{+}	0.944	0.004^{+}	0.047	0.006^{+}	0.540	0.058	0.117	0.043^{+}	0.639	0.013^{+}	0.088
c-j-GJRGARCH	0.673	0.239	0.024^{+}	0.134	0.020^{+}	0.992	0.001^{+}	0.457	0.022^{+}	0.923	0.025^{+}	0.812	0.004^{+}	0.538	0.049^{+}	0.117	0.028^{+}	0.772	0.011^{+}	0.088
c-j-TGARCH	0.824	0.925	0.021^{+}	0.134	0.001^{+}	0.227	0.003^{+}	0.724	0.003^{+}	0.531	0.004^{+}	0.047	0.016^{+}	0.540	0.043^{+}	0.117	0.034^{+}	0.772	0.005^{+}	0.502
c-j-NAGARCH	0.281	0.647	0.033^{+}	0.134	0.006^{+}	0.848	0.001^{+}	0.457	0.012^{+}	0.944	0.003^{+}	0.047	0.036^{+}	0.884	0.079	0.363	0.046^{+}	0.982	0.021^{+}	0.816

TABLE 5.1 [C] – Stage 1 of the Value at Risk Backtesting Procedure – FRONTIER markets

This table summarizes the Monte Carlo *p*-values of the Conditional Coverage (CC) test (Christoffersen, 1998) and the normal *p*-values of the Duration (DUR) test (Christoffersen & Pelletier, 2004) for each market within the *Frontier* market classification. The bold font and the green cells indicate that the respective models have passed both tests at the 5 percent level of significance for the specific market. The abbreviations n, s, j, and c in the first column stand for Normal Distribution, Student-*t* Distribution, Johnson SU Distribution, and Conditional Extreme Value Theory. In the column of 'CC', the figure $^-$ indicates underestimation of risk (i.e. too many violations) and the figure $^+$ indicates overestimation of risk (i.e. too few violations).

	Nigeria		Kazakhstan		Croatia		Romania		Kenya		Tunisia		Lebanon		Mauritius		Slovenia		Estonia	
	CC	DUR	CC	DUR	CC	DUR	CC	DUR	CC	DUR	CC	DUR	CC	DUR	CC	DUR	CC	DUR	CC	DUR
n-GARCH	0.035$^-$	0.031	0.001$^-$	0.433	0.219	0.473	0.057	0.637	0.001$^-$	0.662	0.359	0.728	0.340	0.591	0.021$^-$	0.789	0.249	0.471	0.243	0.472
n-IGARCH	0.046$^-$	0.042	0.016$^-$	0.368	0.153	0.628	0.051	0.637	0.002$^-$	0.973	0.163	0.985	0.230	0.449	0.042$^-$	0.889	0.011$^-$	0.468	0.254	0.472
n-EGARCH	0.026$^-$	0.047	0.001$^-$	0.284	0.902	0.762	0.084	0.724	0.001$^-$	0.088	0.126	0.648	0.266	0.502	0.029$^-$	0.829	0.164	0.043	0.192	0.275
n-GJRGARCH	0.039$^-$	0.150	0.001$^-$	0.388	0.246	0.788	0.065	0.621	0.003$^-$	0.796	0.154	0.663	0.342	0.591	0.034$^-$	0.889	0.107	0.102	0.249	0.472
n-TGARCH	0.038$^-$	0.030	0.001$^-$	0.532	0.666	0.369	0.096	0.509	0.001$^-$	0.801	0.123	0.648	0.165	0.453	0.039$^-$	0.611	0.115	0.073	0.163	0.506
n-NAGARCH	0.029$^-$	0.050	0.001$^-$	0.606	0.201	0.617	0.164	0.364	0.001$^-$	0.614	0.338	0.588	0.474	0.564	0.020$^-$	0.691	0.081	0.094	0.243	0.472
s-GARCH	0.021$^-$	0.006	0.574	0.902	0.819	0.422	0.902	0.283	0.457	0.248	0.666	0.240	0.070	0.660	0.285	0.473	0.912	0.101	0.460	0.641
s-IGARCH	0.106	0.023	0.406	0.653	0.924	0.245	0.560	0.373	0.733	0.327	0.269	0.729	0.129	0.491	0.519	0.299	0.745	0.329	0.327	0.418
s-EGARCH	0.025$^-$	0.002	0.721	0.850	0.669	0.369	0.470	0.219	0.463	0.737	0.368	0.463	0.142	0.276	0.491	0.294	0.926	0.207	0.467	0.488
s-GJRGARCH	0.017$^-$	0.006	0.748	0.997	0.529	0.055	0.828	0.564	0.449	0.248	0.474	0.303	0.127	0.268	0.909	0.293	0.918	0.207	0.45	0.641
s-TGARCH	0.016$^-$	0.001	0.810	0.901	0.509	0.586	0.560	0.373	0.684	0.342	0.391	0.463	0.071	0.377	0.495	0.816	0.821	0.288	0.462	0.488
s-NAGARCH	0.188	0.032	0.340	0.881	0.485	0.055	0.824	0.564	0.591	0.572	0.498	0.303	0.076	0.205	0.662	0.381	0.733	0.067	0.345	0.418
j-GARCH	0.028$^-$	0.026	0.654	0.740	0.395	0.051	0.504	0.722	0.743	0.323	0.799	0.193	0.535	0.173	0.286	0.473	0.814	0.263	0.228	0.472
j-IGARCH	0.135	0.072	0.395	0.653	0.672	0.369	0.821	0.564	0.653	0.977	0.279	0.729	0.287	0.387	0.515	0.299	0.894	0.177	0.262	0.472
j-EGARCH	0.042$^-$	0.026	0.396	0.653	0.846	0.884	0.755	0.350	0.736	0.393	0.516	0.873	0.652	0.221	0.484	0.294	0.777	0.270	0.169	0.275
j-GJRGARCH	0.032$^-$	0.026	0.635	0.942	0.365	0.051	0.654	0.980	0.918	0.278	0.518	0.436	0.486	0.207	0.657	0.297	0.914	0.207	0.241	0.472
j-TGARCH	0.039$^-$	0.003	0.519	0.804	0.650	0.369	0.684	0.342	0.735	0.393	0.368	0.463	0.679	0.221	0.510	0.816	0.893	0.207	0.235	0.472
j-NAGARCH	0.240	0.105	0.521	0.804	0.393	0.051	0.499	0.722	0.732	0.323	0.388	0.463	0.653	0.221	0.675	0.381	0.588	0.146	0.234	0.472
c-n-GARCH	0.046$^+$	0.014	0.147	0.549	0.384	0.051	0.035$^+$	0.402	0.006$^+$	0.030	0.081	0.985	0.148	0.268	0.012$^+$	0.812	0.026$^+$	0.455	0.143	0.211
c-n-IGARCH	0.004$^+$	0.008	0.146	0.402	0.391	0.051	0.136	0.303	0.013$^+$	0.030	0.085	0.985	0.273	0.387	0.043$^+$	0.094	0.049$^+$	0.330	0.140	0.211
c-n-EGARCH	0.018$^+$	0.015	0.138	0.549	0.065	0.401	0.073	0.211	0.047$^+$	0.485	0.070	0.255	0.283	0.763	0.037$^+$	0.257	0.253	0.316	0.400	0.030
c-n-GJRGARCH	0.033$^+$	0.014	0.160	0.549	0.265	0.093	0.023$^+$	0.848	0.011$^+$	0.030	0.081	0.985	0.136	0.613	0.024$^+$	0.812	0.056	0.550	0.154	0.211
c-n-TGARCH	0.024$^+$	0.007	0.140	0.549	0.075	0.401	0.034$^+$	0.402	0.011$^+$	0.033	0.088	0.255	0.079	0.205	0.022$^+$	0.593	0.148	0.469	0.376	0.030
c-n-NAGARCH	0.053	0.014	0.130	0.549	0.284	0.053	0.027$^+$	0.848	0.008$^+$	0.025	0.075	0.985	0.084	0.452	0.016$^+$	0.812	0.157	0.177	0.254	0.277
c-s-GARCH	0.044$^+$	0.014	0.088	0.867	0.398	0.051	0.029$^+$	0.848	0.145	0.627	0.135	0.908	0.145	0.491	0.007$^+$	0.240	0.023$^+$	0.455	0.077	0.053
c-s-IGARCH	0.005$^+$	0.008	0.082	0.867	0.403	0.051	0.153	0.303	0.526	0.710	0.078	0.985	0.269	0.387	0.002$^+$	0.468	0.037$^+$	0.330	0.306	0.269
c-s-EGARCH	0.027$^+$	0.009	0.150	0.549	0.140	0.447	0.077	0.211	0.280	0.416	0.064	0.255	0.281	0.204	0.012$^+$	0.531	0.035$^+$	0.550	0.426	0.030
c-s-GJRGARCH	0.039$^+$	0.014	0.086	0.867	0.267	0.231	0.029$^+$	0.848	0.125	0.627	0.303	0.537	0.281	0.204	0.001$^+$	0.555	0.033$^+$	0.550	0.303	0.021
c-s-TGARCH	0.011$^+$	0.009	0.064	0.624	0.129	0.447	0.133	0.910	0.018$^+$	0.002	0.075	0.255	0.304	0.717	0.005$^+$	0.288	0.157	0.343	0.263	0.004
c-s-NAGARCH	0.039$^+$	0.008	0.057	0.970	0.259	0.231	0.017$^+$	0.848	0.409	0.655	0.159	0.908	0.080	0.205	0.004$^+$	0.131	0.031$^+$	0.922	0.124	0.211
c-j-GARCH	0.005$^+$	0.003	0.070	0.867	0.148	0.447	0.047$^+$	0.663	0.274	0.570	0.075	0.798	0.133	0.491	0.015$^+$	0.031	0.024$^+$	0.455	0.143	0.204
c-j-IGARCH	0.005$^+$	0.008	0.052	0.970	0.129	0.447	0.056	0.830	0.406	0.536	0.127	0.908	0.292	0.387	0.006$^+$	0.468	0.044$^+$	0.330	0.141	0.204
c-j-EGARCH	0.023$^+$	0.009	0.145	0.549	0.149	0.447	0.041$^+$	0.663	0.274	0.416	0.150	0.319	0.142	0.268	0.015$^+$	0.531	0.046$^+$	0.550	0.365	0.077
c-j-GJRGARCH	0.004$^+$	0.003	0.042$^+$	0.970	0.133	0.447	0.025$^+$	0.848	0.259	0.570	0.131	0.319	0.093	0.660	0.005$^+$	0.011	0.044$^+$	0.550	0.151	0.015
c-j-TGARCH	0.024$^+$	0.009	0.095	0.624	0.139	0.447	0.054	0.663	0.123	0.604	0.141	0.319	0.142	0.268	0.008$^+$	0.288	0.081	0.528	0.404	0.003
c-j-NAGARCH	0.001$^+$	0.001	0.051	0.970	0.298	0.093	0.024$^+$	0.848	0.369	0.383	0.071	0.204	0.137	0.268	0.001$^+$	0.131	0.078	0.353	0.080	0.064

violations, as both underestimation and overestimation of risk need to be prevented. In theory, however, a viable VaR model does not only produce the correct amount of violations, but also produces violations that are evenly spread over time, as large losses that happen in succession are more likely to result in disasters such as solvency issues and bankruptcy. In order to capture this prominent danger for the risk taking vendor, the duration test is implemented.

Table 5.1 [A-C] documents the p-values belonging to the conditional coverage test and the duration test for the 36 implemented $VaR_{99\%}$ models when applied to the 30 different market indices. As a reference guide, the green cells indicate that the respective models pass both tests at the 5 percent level of significance. Furthermore, the minus (−) and the plus (+) signs indicate that the conditional coverage test is rejected for that specific model due to either underestimation or overestimation of the market risk.

Result 1 – There is a clear discrepancy in terms of adequate risk estimation between the relatively simplistic GARCH framework and the more sophisticated cEVT framework when forecasting Value at Risk at the 99 percent level of confidence.

The results from the conditional coverage test elucidate a clear pattern for the forecasting behavior of the standard GARCH framework. The large quantity of minus signs in the upper part of Table 5.1 [A-C] reveals a tendency of the GARCH framework to underestimate the true level of risk, i.e. to forecast series of VaR that include too many violations. This pattern of underestimation seems to be most prominent for GARCH models with a Normal distribution specification, as 170 out of the 180 estimated models from this type are rejected by the conditional coverage test for having too many violations, compared to only 10 rejections and 7 rejections for the Student-t distribution and the JSU distribution, respectively. This systematic underperformance is in line with the descriptive statistics from Section 4.1, where significant kurtosis and skewness coefficients already signaled a consistent deviation from the Normal distribution. This inability to account for leptokurtic and skewed return distributions causes the distribution to fail prominently for the vast majority of markets.

On the other hand, a pattern of risk overestimation seems to hold for models that originate from the more sophisticated framework of cEVT. The lower part of Table 5.1 [A-C] displays an exclusive usage of plus signs, where no less than 270 out of the 540 estimated cEVT models are rejected for including too few violations. This bias of overestimation should be mainly attributed to the interaction between the mathematical design of cEVT and the chosen level of confidence of 99 percent. The design of cEVT focusses explicitly on modeling the tail regions of the distribution, which enables a prediction of large and rare losses, possibly even larger than the ones that occurred within the range of available observations. However, if a situation appears where the specific asset does not produce a large amount of relatively extreme outliers that correspond with the chosen level of confidence, the cEVT model will systematically overestimate the financial risk. Therefore, in the context of the thirty considered market indices, one could argue that the chosen confidence level of 99 percent is not far enough in the tails of the return distributions in order for the cEVT framework to be statistically adequate. Nonetheless, the BCBS requires the usage of a 99 percent confidence level in their regulation framework, which makes it the most relevant one for both practitioners and academics.

Result 2 – The implemented models of risk are on a large scale capable of capturing the danger of violation clustering when forecasting Value at Risk at the 99 percent level of confidence.

The results from the duration test of Table 5.1 [A-C] show that the null hypothesis of no dependency between VaR violations gets hardly ever rejected at the 5 percent level of significance. This observation is in sharp contrast to the large number of rejections that occurred for the conditional coverage test. The low number of rejections suggests that the estimated risk models are at large capable of estimating series of VaR that have a statistically sound spread of violations. If one looks at the amount of rejections for each market, it becomes apparent that the number tends to vary consistently between zero and eight. This should minimize the risk of opportunity costs for the risk taking vendor. The only exception could be found within the context of Frontier markets, namely for Nigeria, were 33 out of the 36 estimated models did not pass the test. The vast majority of risk models are having serious trouble in forecasting series of VaR with randomly distributed violations for the Nigerian index. This should constitute an urgent alarm to any risk manager who considers to neglect the potential danger of violation clustering.

TABLE 5.2 [A] – Stage 2 of the Value at Risk Backtesting Procedure – DEVELOPED MARKETS

This table summarizes the Asymmetric Loss Function values from the backtest of González-Rivera et al. (2004) and the corresponding rankings for each market within the *Developed* market classification. It includes only 'LOSS' values for the $VaR_{99\%}$ models that have passed the first stage of the Value at Risk backtesting procedure. Low values speak in favor of a model, while high values do the opposite. The green cells indicate that the respective models have the number one ranking for the specific market. The abbreviations n, s, j, and c in the first column stand for Normal Distribution, Student-*t* Distribution, Johnson SU Distribution, and Conditional Extreme Value Theory.

	USA		Japan		Hong Kong		France		UK		Germany		Canada		Australia		Netherlands		Italy	
	LOSS	RANK	LOSS	RANK	LOSS	RANK	LOSS	RANK	LOSS	RANK	LOSS	RANK	LOSS	RANK	LOSS	RANK	LOSS	RANK	LOSS	RANK
n-GARCH															60.49	18	64.67	7	94.62	7
n-IGARCH															60.20	17	64.96	13	95.41	16
n-EGARCH					52.39	8	63.22	10	59.58	8	72.19	3			58.02	5	64.70	10	95.14	10
n-GJRGARCH					53.21	13					73.73	13			58.37	8	65.14	16	95.28	13
n-TGARCH	45.32	5			52.38	7			58.80	2	71.62	1			58.28	7	63.76	1	94.55	4
n-NAGARCH	45.13	4			52.94	11	61.35	2	59.87	10	71.83	2			58.10	6	64.14	4	94.11	1
s-GARCH					53.69	23	62.57	5	60.71	14	72.48	5	53.08	22	59.57	13	64.67	8	94.62	8
s-IGARCH					53.86	25	62.90	8	60.79	15	72.96	7	52.76	21	59.57	14	64.96	14	95.41	17
s-EGARCH	44.81	3			51.53	3	62.88	7	58.86	3			50.08	2	57.84	3	64.70	11	95.14	11
s-GJRGARCH			65.84	16	52.10	5			59.44	6	73.15	10	50.54	4	57.81	1	65.14	17	95.28	14
s-TGARCH	44.58	2					62.39	4	57.87	1			49.76	1	57.97	4	63.76	2	94.55	5
s-NAGARCH	43.61	1	66.47	18	51.33	1	61.28	1	59.24	5	72.32	4	50.77	5	57.83	2	64.14	5	94.11	2
j-GARCH					53.81	24	62.66	6	60.46	12	72.98	8	52.29	16	59.73	15	64.67	9	94.62	9
j-IGARCH	49.03	11			54.06	26	63.53	11	60.70	13	73.43	12	52.34	17	59.97	16	64.96	15	95.41	18
j-EGARCH	47.14	7	65.83	15	51.77	4	64.04	12	60.08	11	73.08	9	50.29	3	58.61	9	64.70	12	95.14	12
j-GJRGARCH	46.49	6	65.26	13	52.22	6			59.52	7			50.84	6	58.73	12	65.14	18	95.28	15
j-TGARCH			66.18	17			63.03	9	58.96	4	72.60	6	50.87	7	58.63	11	63.76	3	94.55	6
j-NAGARCH			65.27	14	51.46	2	62.33	3	59.6	9	73.34	11	51.29	9	58.61	10	64.14	6	94.11	3
c-n-GARCH	48.52	8	64.00	6			65.36	17					54.19	28					98.08	22
c-n-IGARCH	49.28	12	63.45	3	55.99	31	66.46	24	63.18	17			54.01	26			67.66	24	98.74	28
c-n-EGARCH			64.09	7	53.56	19	66.55	26			76.01	16	51.50	10			67.54	20	99.06	31
c-n-GJRGARCH			64.63	11	53.43	17	65.90	20			77.25	21	52.18	14			67.70	25		
c-n-TGARCH					53.23	14	64.86	13			75.54	14	52.57	20					98.47	25
c-n-NAGARCH			64.38	9	52.85	9	64.91	14			77.03	20	52.53	19					97.93	19
c-s-GARCH	48.82	9	63.97	5	55.83	28	65.25	15					53.91	23					97.96	20
c-s-IGARCH	49.43	13	63.26	2	55.93	30	66.24	23	62.91	16			53.99	24			67.80	27	98.7	27
c-s-EGARCH					53.57	20	67.61	28			76.39	18	51.66	11			67.65	23	99.09	32
c-s-GJRGARCH					53.49	18	66.46	25			77.61	24	51.95	13			67.83	28		
c-s-TGARCH					53.23	15	65.38	18			75.93	15	52.19	15					98.56	26
c-s-NAGARCH			64.19	8	52.92	10					77.32	22	51.81	12					98.00	21
c-j-GARCH	48.91	10	63.94	4	55.82	27	65.28	16					54.01	25					98.09	23
c-j-IGARCH			62.96	1	55.88	29	66.12	21					54.17	27			67.63	22	98.76	29
c-j-EGARCH			64.87	12	53.66	21	68.53	29			76.85	19	51.20	8			67.62	21	99.38	33
c-j-GJRGARCH					53.68	22	66.87	27			77.82	25					67.74	26		
c-j-TGARCH					53.37	16	65.74	19			76.09	17	52.48	18			67.00	19	98.85	30
c-j-NAGARCH			64.49	10	53.12	12	66.18	22			77.51	23							98.21	24

TABLE 5.2 [B] – Stage 2 of the Value at Risk Backtesting Procedure – EMERGING MARKETS

This table summarizes the Asymmetric Loss Function values from the backtest of González-Rivera et al. (2004) and the corresponding rankings for each market within the *Emerging* market classification. It includes only 'LOSS' values for the $VaR_{99\%}$ models that have passed the first stage of the Value at Risk backtesting procedure. Low values speak in favor of a model, while high values do the opposite. The green cells indicate that the respective models have the number one ranking for the specific market. The abbreviations n, s, j, and c in the first column stand for Normal Distribution, Student-*t* Distribution, Johnson SU Distribution, and Conditional Extreme Value Theory.

	China		India		South-Korea		Brazil		Mexico		Malaysia		Indonesia		Poland		Colombia		Peru	
	LOSS	RANK	LOSS	RANK	LOSS	RANK	LOSS	RANK	LOSS	RANK	LOSS	RANK	LOSS	RANK	LOSS	RANK	LOSS	RANK	LOSS	RANK
n-GARCH			64.42	14	62.22	7					44.26	13			87.00	16			81.93	13
n-IGARCH			65.01	16	62.18	6	81.74	10			44.71	15			86.07	10			83.31	16
n-EGARCH			62.59	2	62.05	4			63.50	5	43.07	9			84.62	4	65.19	4	81.23	1
n-GJRGARCH			62.73	4	61.88	3					43.09	10			85.88	7	65.99	10	81.84	10
n-TGARCH			62.38	1	61.81	2			63.27	3	42.73	5			84.16	1	65.07	1	81.32	4
n-NAGARCH			62.63	3	61.42	1			63.13	1	42.87	7			86.43	13	65.46	7	81.56	7
s-GARCH	63.05	24	64.29	13	62.14	5	81.48	9	67.30	12	44.13	12	77.93	7	87.00	17			81.93	14
s-IGARCH	62.64	22	64.76	15			81.83	11	68.02	14	44.74	16	77.77	6	86.07	11			83.31	17
s-EGARCH	60.12	9	62.85	6	62.44	10	79.03	3	63.61	6	41.97	2	77.35	4	84.62	5	65.19	5	81.23	2
s-GJRGARCH	59.83	5	62.81	5	62.51	11	78.71	2	63.82	7	42.24	3	76.88	2	85.88	8	65.99	11	81.84	11
s-TGARCH	59.89	6	62.91	7	62.42	9	79.13	4	63.19	2	41.90	1	78.26	8	84.16	2	65.07	2	81.32	5
s-NAGARCH	58.83	1	62.98	8	62.38	8	77.98	1	63.39	4	42.48	4	76.08	1	86.43	14	65.46	8	81.56	8
j-GARCH	63.02	23	65.20	17	63.35	12	82.08	12	68.02	13	44.64	14	78.79	10	87.00	18			81.93	15
j-IGARCH	62.52	21	65.79	18	64.03	13	82.87	13	68.54	15	45.36	17	80.00	12	86.07	12			83.31	18
j-EGARCH	60.32	11	63.42	10	65.00	15	80.22	7	64.93	11	42.74	6	78.27	9	84.62	6	65.19	6	81.23	3
j-GJRGARCH	59.94	7	63.42	9	64.45	14	79.32	5	64.83	10	43.13	11	77.67	5	85.88	9	65.99	12	81.84	12
j-TGARCH	60.01	8	63.47	11	65.11	17	80.42	8	64.57	8			79.22	11	84.16	3	65.07	3	81.32	6
j-NAGARCH	58.87	2	63.81	12	65.09	16	79.46	6	64.82	9	42.99	8	77.16	3	86.43	15	65.46	9	81.56	9
c-n-GARCH	64.34	29																		
c-n-IGARCH	64.09	27	68.78	21																
c-n-EGARCH	61.46	20															70.37	13		
c-n-GJRGARCH	60.75	15																		
c-n-TGARCH	61.35	19																		
c-n-NAGARCH	60.79	16													89.62	20				
c-s-GARCH	64.26	28													89.65	21				
c-s-IGARCH	63.97	25	68.65	20																
c-s-EGARCH	60.86	17																		
c-s-GJRGARCH	60.28	10																		
c-s-TGARCH	60.65	13															71.20	14		
c-s-NAGARCH	59.77	4													89.68	22				
c-j-GARCH	64.36	30													89.78	24			88.60	19
c-j-IGARCH	64.04	26	68.62	19																
c-j-EGARCH	60.94	18													88.46	19				
c-j-GJRGARCH	60.40	12																		
c-j-TGARCH	60.65	14																		
c-j-NAGARCH	59.73	3													89.76	23				

TABLE 5.2 [C] – Stage 2 of the Value at Risk Backtesting Procedure – FRONTIER MARKETS

This table summarizes the Asymmetric Loss Function values from the backtest of González-Rivera et al. (2004) and the corresponding rankings for each market within the *Frontier* market classification. It includes only 'LOSS' values for the $VaR_{99\%}$ models that have passed the first stage of the Value at Risk backtesting procedure. Low values speak in favor of a model, while high values do the opposite. The green cells indicate that the respective models have the number one ranking for the specific market. The abbreviations n, s, j, and c in the first column stand for Normal Distribution, Student-*t* Distribution, Johnson SU Distribution, and Conditional Extreme Value Theory.

	Nigeria		Kazakhstan		Croatia		Romania		Kenya		Tunisia		Lebanon		Mauritius		Slovenia		Estonia	
	LOSS	RANK	LOSS	RANK	LOSS	RANK	LOSS	RANK	LOSS	RANK	LOSS	RANK	LOSS	RANK	LOSS	RANK	LOSS	RANK	LOSS	RANK
n-GARCH					54.58	13	77.97	12			42.34	5	43.43	2			57.16	3	68.83	28
n-IGARCH					54.91	16	78.81	17			41.83	1	43.94	4					68.47	23
n-EGARCH					52.62	1	78.04	15			42.28	4	44.36	8					68.22	21
n-GJRGARCH					54.24	8	76.53	3			42.42	6	43.10	1			58.43	16	68.77	27
n-TGARCH					52.91	2	76.51	2			42.57	10	44.61	10			58.49	17	68.09	19
n-NAGARCH					54.31	9	75.96	1			41.89	2	43.59	3			58.53	18	68.89	29
s-GARCH			108.83	9	54.68	14	77.99	13	53.41	4	42.86	16	46.20	13	37.40	7	56.90	2	67.19	12
s-IGARCH			109.42	12	55.27	17	78.88	18	53.07	3	42.53	9	46.65	22	37.98	11	57.92	9	67.05	11
s-EGARCH			107.91	7	53.72	4	77.83	11	53.59	6	42.59	11	48.78	35	36.81	1	57.87	8	66.03	5
s-GJRGARCH			107.78	4	54.45	11	77.21	8	52.89	1	43.16	18	46.44	19	37.17	5	57.51	6	66.70	8
s-TGARCH			107.48	3	53.71	3	78.03	14	53.61	7	42.74	12	49.48	36	36.89	3	57.47	5	65.89	2
s-NAGARCH			106.49	2	53.89	5	76.67	5	52.93	2	42.83	15	46.81	26	37.46	9	57.99	11	66.73	9
j-GARCH			108.86	10	54.51	12	77.80	10	54.06	9	42.77	13	44.23	7	37.40	8	56.79	1	67.50	14
j-IGARCH	90.18	1	109.73	13	55.28	18	78.69	16	54.90	12	42.50	7	44.22	6	37.98	12	58.28	13	67.31	13
j-EGARCH			108.31	8	54.23	7	77.58	9	54.47	11	42.26	3	46.02	12	36.81	2	57.99	10	66.02	4
j-GJRGARCH			107.82	6	54.77	15	76.94	7	53.66	8	42.95	17	44.03	5	37.31	6	57.25	4	66.87	10
j-TGARCH			107.82	5	54.08	6	76.91	6	54.20	10	42.50	8	46.42	18	36.89	4	57.81	7	65.69	1
j-NAGARCH	92.06	2	106.18	1	54.32	10	76.61	4	53.48	5	42.79	14	44.39	9	37.46	10	58.29	14	66.66	6
c-n-GARCH			112.47	18	58.13	25					44.06	21	46.28	15					68.59	24
c-n-IGARCH			115.78	26	58.50	32	84.65	25			43.40	19	46.49	20					68.42	22
c-n-EGARCH			110.62	14	58.08	22	83.00	22			44.20	23	47.05	28			59.97	23		
c-n-GJRGARCH			110.88	15	57.66	20					44.11	22	45.82	11			58.39	15	68.61	26
c-n-TGARCH			110.98	16	58.10	23					44.41	25	47.67	32			59.17	22		
c-n-NAGARCH			109.10	11	67.22	36					43.52	20	47.12	29			58.24	12	68.60	25
c-s-GARCH			114.71	25	58.37	31			59.75	17	44.63	27	46.34	16					68.16	20
c-s-IGARCH			117.40	29	58.64	34	83.75	24	60.53	21	44.22	24	46.80	25					67.90	18
c-s-EGARCH			112.92	20	57.95	21	81.49	20	60.41	20	44.63	28	47.61	31						
c-s-GJRGARCH			113.20	21	58.28	28			59.06	13	45.06	32	46.73	23						
c-s-TGARCH			113.31	22	58.12	24	80.81	19			44.94	31	48.25	34			58.90	20		
c-s-NAGARCH			111.60	17	57.45	19			59.35	14	44.65	29	46.95	27					67.54	15
c-j-GARCH			115.78	27	58.62	33			60.02	18	45.18	33	46.35	17					67.81	17
c-j-IGARCH			117.33	28	58.32	29	83.33	23	61.07	22	44.50	26	46.76	24					67.68	16
c-j-EGARCH			113.85	23	58.17	26			61.10	23	44.82	30	47.38	30					66.00	3
c-j-GJRGARCH					58.93	35			59.70	16	45.58	36	46.25	14						
c-j-TGARCH			114.17	24	58.33	30	81.52	21	60.09	19	45.29	34	48.04	33			59.13	21		
c-j-NAGARCH			112.91	19	58.19	27			59.43	15	45.30	35	46.49	21			58.85	19	66.68	7

TABLE 5.3 – Number One Ranked Models of Value at Risk

This table provides an overview of the respective 99 percent Value at Risk models that enjoy the highest ranking within each market. It shows the models that are classified as optimal by the Asymmetric Loss function of González-Rivera et al. (2004), conditional upon passing the Conditional Coverage test (Christoffersen, 1998) and the Duration test (Christoffersen & Pelletier, 2004) at the 5 percent level of significance. In the column 'Leverage', the figure ✓ indicates whether the highest ranked model belongs to the class of leverage GARCH models. A similar structure is chosen for the column 'EVT', where the figure ✓ denotes whether the highest ranked model comes from the framework of Conditional Extreme Value Theory.

DEVELOPED MARKETS				
Country	**Distribution**	**Model**	**Leverage**	**EVT**
USA	Student-*t*	NAGARCH	✓	
Japan	JSU	IGARCH		✓
Hong Kong	Student-*t*	NAGARCH	✓	
France	Student-*t*	NAGARCH	✓	
United Kingdom	Student-*t*	TGARCH	✓	
Germany	Normal	TGARCH	✓	
Canada	Student-*t*	TGARCH	✓	
Australia	Student-*t*	GJRGARCH	✓	
Netherlands	Normal	TGARCH	✓	
Italy	Normal	NAGARCH	✓	

EMERGING MARKETS				
Country	**Distribution**	**Model**	**Leverage**	**EVT**
China	Student-*t*	NAGARCH	✓	
India	Normal	TGARCH	✓	
South-Korea	Normal	NAGARCH	✓	
Brazil	Student-*t*	NAGARCH	✓	
Mexico	Normal	NAGARCH	✓	
Malaysia	Student-*t*	TGARCH	✓	
Indonesia	Student-*t*	NAGARCH	✓	
Poland	Normal	TGARCH	✓	
Colombia	Normal	TGARCH	✓	
Peru	Normal	GJRGARCH	✓	

FRONTIER MARKETS				
Country	**Distribution**	**Model**	**Leverage**	**EVT**
Nigeria	JSU	IGARCH		
Kazakhstan	JSU	NAGARCH	✓	
Croatia	Normal	EGARCH	✓	
Romania	Normal	NAGARCH	✓	
Kenya	Student-*t*	GJRGARCH	✓	
Tunisia	Normal	IGARCH		
Lebanon	Normal	GJRGARCH	✓	
Mauritius	Student-*t*	EGARCH	✓	
Slovenia	JSU	GARCH		
Estonia	JSU	TGARCH	✓	

5.1.2 Stage Two of the VaR (99%) Backtesting Procedure

The first stage of the VaR backtesting procedure tested the statistical adequacy of each specified VaR model: if the null hypothesis of both the conditional coverage test and the duration test could not be rejected, the VaR model was characterized as a statistically adequate model for volatility forecasting. However, these two tests were not capable of providing any direct information on whether an 'adequate' risk model is more accurate than another 'adequate' one. This lack of power could be attributed to the fact that a higher p-value does not directly indicate statistical superiority of that specific model among its competitors. Nonetheless, the ranking of competing models is deemed to be of high importance. It would allow risk practitioners and academics to select the optimal risk model for each market. Therefore, this section carries out the asymmetric loss function of González-Rivera et al. (2004) on each model that passed the first stage. The function measures the accuracy of the VaR forecasts by looking at the distance between the observed returns and the forecast series of VaR. By design, the asymmetric design penalizes underestimation of risk more heavily than overestimation of risk. The final step is to compare the risk models based on their values in the loss function: low values peak in favor of a risk model, while high values to the exact opposite.

Table 5.2 [A-C] presents the summary results for the asymmetric loss function when applied to the respective models that are found to be statistically acceptable in the first stage of the model selection procedure. As a reference guide, the green cells indicate that the respective risk models enjoy the highest ranking for that specific country index. In order to provide a more convenient overview of all the number one ranked VaR models, Table 5.3 was created. This table adds an indicator column for both the leverage term and the conditional EVT specification on the ground of easing the detection of potential patterns in model performance between the three market classification groups.

HYPOTHESIS ONE	
$H0_1$	The underlying market classification of a country index is unrelated to the relative forecasting performance of $VaR_{99\%}$ models.
HA_1	The underlying market classification of a country index is related to the relative forecasting performance of $VaR_{99\%}$ models.

The main purpose of this section is to test whether the abovementioned null hypothesis needs to be rejected in favor of the alternative hypothesis. In order to enable a comparative analysis, the examination of the results will first be divided into three subparts, from which at the end a conclusion will be extracted.

I. <u>The simplistic GARCH framework against the sophisticated cEVT framework.</u>

The ranking results from Table 5.3 illustrate that the standard GARCH framework is the favored approach for no less than 29 out of the 30 considered markets. This strong pattern of preference could be attributed to the adverse interaction between the design of the cEVT framework and the design of the chosen 'two-stage' backtesting procedure. The approach of cEVT focusses explicitly on modeling the tail regions of the return distribution, which enables the prediction of large and rare losses, possibly even larger than the ones that have ever occurred within the range of available observations. As a result, the design of the cEVT framework will forecast systematically higher values for VaR than the design of the standard GARCH framework. If a situation arises where the return distribution of a financial asset does not exhibit a large amount of relatively extreme outliers, this pattern of overestimation could have two serious consequences regarding the ranking results of the asymmetric loss function:

(i.) The first issue would be that models originating from the cEVT framework will be more likely to forecast a series of VaR that includes too few violations to pass the conditional coverage test and the corresponding first stage of the backtesting procedure. In other words, these models will not make it to the ranking procedure from the second stage of the backtesting procedure.

(ii.) The second issue would be that models originating from the cEVT framework that do pass the first stage will face a higher 'overestimation' penalty term on normal trading days and a lower 'underestimation' penalty term on turbulent trading days. Furthermore, a scenario could emerge where the overestimation bias is so dominant that the relatively large size of the 'overestimation' penalty term completely negates the minimum size of the 'underestimation' penalty term. As a result, these models then face a higher value in the asymmetric loss function and a lower ranking result than models that originate from the standard GARCH framework.

Based on the conditional coverage testing results of Table 5.1 [A-C] and the ranking results of Table 5.2 [A-C], one could argue that both scenarios played a crucial role for the vast majority of the considered markets. Overall, the findings indicate that the relatively simplistic GARCH framework outperforms the more sophisticated CEVT framework for all three market classification groups when forecasting $VaR_{99\%}$. This consistent outperformance could be interpreted as evidence that the market classification groups exhibit similar return dynamics. However, the field of empirical literature has documented the exact opposite. Therefore, this study looked deeper into the interaction between the design of the backtesting procedure and the unique characteristics of each market classification group.

Table 5.4 provides the general return characteristics of the three market classification groups on both normal trading days and turbulent trading days. The next logical step would be to transform these characteristics into an 'overestimation' penalty term and 'underestimation' penalty term for both the GARCH framework and the cEVT framework, and to see whether these penalty terms correspond with the optimal ranking results from Table 5.3.

TABLE 5.4 – Unique Characteristics of the Three Market Classification Groups

This table provides an overview of the general empirical characteristics of the three market classification groups regarding the return behavior on both normal trading days and turbulent trading days, respectively. The return behavior on normal trading days is related to the level of volatility, whereas the return behavior on turbulent trading days is related to the number and the magnitude of extreme outliers.

	Normal Trading Days	Turbulent Trading Days
Developed	Normal level of volatility.	A few extreme outliers.
Emerging	Lower level of volatility.	More extreme outliers.
Frontier	Lowest level of volatility.	Even more extreme outliers.

As illustrated before, the cEVT framework will by design generate series of VaR that are systematically higher than the ones forecasted by the GARCH framework, which means in practice that the 'overestimation' penalty term for cEVT will always be higher and the 'underestimation' penalty term for cEVT will always be lower. However, it is the relative size of the two penalty terms between the frameworks that will be decisive on which one will enjoy the highest ranking.

Developed Markets: The literature shows that country indices belonging to this market classification tend to exhibit only a limited amount of extreme outliers in their return distributions. As a result, one would expect that the more subtle framework of GARCH will be sufficient on turbulent trading days. At the same time, the cEVT framework will most likely produce a much higher 'overestimation' penalty term on normal trading days than its competitor. Looking at the low ranking results for cEVT models within the context of Developed markets, one could indeed argue that the higher 'overestimation' penalty term negatively dominated the value of the loss function.

Emerging Markets: The field of literature has shown that countries belonging to this market classification tend to exhibit low volatility in general, but they simultaneously exhibit frequent financial shocks that give rise to substantial extreme volatility. This characteristic of extreme volatility could be mainly attributed to the fact that Emerging markets are facing a higher illiquidity than Developed markets, as numerous empirical studies have confirmed the existence of a positive relation between illiquidity and return volatility. This extreme behavior in volatility increases the need of implementing cEVT in order to cover for potential large losses and to minimize the 'underestimation' penalty term. However, treating volatility as a uniform measure with a homogeneous relation with liquidity will overlook the important structure of the overall volatility. The overall volatility is merely the integration of two types of volatility: (i.) volatility patterns generated by a discontinuous jump process, which arises from infrequent, large, isolated 'surprise' price changes, and (ii.) diffusive volatility that arises from smooth and more 'expected' small price changes. In the context of Emerging markets, this means that the illiquidity will in general be associated with large financial shocks on turbulent trading days, but where low price changes on normal trading days can still exist (Božović & Totić, 2015). As a result of the lower volatility on normal trading days, the cEVT framework will produce a higher 'overestimation' penalty term in the environment of Emerging markets than for Developed markets. Overall, the low ranking results suggest that this increase in the 'overestimation' penalty term cancels out the decrease in the 'underestimation' penalty term within the context of Emerging markets.

Frontier Markets: According to the literature, Frontier markets are best described as markets that are found to be even more complex than Emerging markets. The first consequence of this description would be that the issues that plague many existing Emerging markets are amplified in the environment of Frontier markets. For example, the market classification framework from the MSCI denotes that Frontier markets have a substantial decrease in liquidity relative to the other two market groups. All these different characteristics are giving rise to the frequency of the outliers and the magnitude of the volatility on turbulent trading days. In terms of the loss function, this means that there is a higher urgency for the implementation of cEVT in order to minimize the 'underestimation' penalty term. However, the low level of volatility on normal trading days is also amplified in the environment of Frontier markets. In terms of the loss function, this results in the fact that the cEVT models will face an 'overestimation' penalty term that is even larger than the one observed for Emerging markets. Overall, the low ranking results for the cEVT models from Table 5.3 in the context of Frontier markets suggest that the decrease in the 'underestimation' penalty term is outweighed by the increase in the 'overestimation' penalty term.

To summarize the above, one could argue that the consistent outperformance of the GARCH framework is not caused by the existence of similar return dynamics of the three market classification groups, but is instead caused by the interaction between the unique characteristics of the three market classification groups and the design of the asymmetric loss function.

An important deduction from the ranking results of Table 5.3 and the underlying market characteristics of Table 5.4 would be that the chosen confidence level of 99 percent is not far enough in the tails of the return distribution in order for the framework of cEVT to be accurate. One would expect that in the context of a higher confidence level, the cEVT approach will offset its overestimation bias on normal trading days with a crucial minimization of extreme losses, especially in the environment of the extreme Frontier markets. Simultaneously, one would expect that in the context of a higher confidence level, the GARCH framework will not pass the filtering stage of the backtesting procedure due to its tendency to underestimate risk. Nonetheless, the BCBS requires the usage of a 99 percent confidence level in their regulation framework, which makes it the most relevant percentage for both practitioners and academics.

II. The relative performance of the three error distributions.

The ranking results from Table 5.2 [A-C] and Table 5.3 illustrate that the Normal distribution is favored the most (13 times), followed by the Student-t distribution (12 times) and the JSU distribution (5 times). The finding that the Normal distribution dominates the ranking in general could be surprising if one recalls its inability to account for excess kurtosis and skewness.

The design of the Normal distribution tends to underestimate the risk of financial assets by ignoring crucial stylized facts exhibited by financial return distributions (see Section 3.4). Conversely, the Student-t distribution incorporates a parameter to account for the stylized fact of fat tails, and the JSU distribution adds on top of that a parameter to account for the stylized fact of return asymmetry. This means that if a situation arises where a return distribution shows significant kurtosis and/or skewness, the models with a Normal distribution specification will forecast levels of VaR that are systematically lower than compared to the ones forecasted by the other two distributions. As a reminder, the descriptive statistics from Section 4.1 denoted a clear deviation from normality for all considered markets. Based on these statistics, one would expect that the Normal distribution would fail prominently when forecasting VaR at the 99 percent level of confidence. Actually, the conditional coverage testing results from Section 5.1.1 are in support of this statement, as models with a Normal distribution were rejected much more frequently for having too many violations. Nonetheless, the ranking results from Table 5.3 show outperformance for the Normal distribution in 13 markets. This remarkable pattern should be attributed to the interaction between the design of the asymmetric loss function and the composition of the three considered distributions.

To explain this interaction, one should first understand that if a risk model with a Normal distribution produces a number of VaR violations that is within the desired interval of the confidence level, and if these VaR violations are not clustered over time, this specific model will still pass the first stage of the VaR backtesting procedure, despite the highlighted inabilities. In fact, Table 5.2 [A-C] shows that no less than 22 out of the 30 considered markets do exhibit return dynamics that are adequately captured by at least one model with a Normal distribution specification. After passing the first stage of the backtesting procedure, the model will be evaluated by the asymmetric loss function. As normally distributed models will most likely forecast lower levels of VaR than its two rivals, it will achieve a lower 'overestimation' penalty term on normal trading days and a higher 'underestimation' penalty term on turbulent trading days. Overall, the ranking results suggest that for 13 countries the minimization of the opportunity costs played a larger role than the minimization of the potential losses.

However, it should be noted that the Normal distribution is not by definition granted with a highest ranking after passing the first stage of the backtesting procedure. The incorporated loss function is asymmetric by design and when the VaR violations on turbulent trading days are too large, these models will be punished more severely by the 'underestimation' penalty term. At the same time, the Normal distribution models will most likely produce a higher number of violations, which means that the 'underestimation' penalty term will again be larger. The markets from Malaysia and Estonia are two good examples where all VaR models with a Normal distribution specification passed the first stage, but did not receive the highest ranking in the second stage.

Finally, in order to provide sufficient information for the testing of the first hypothesis, the forecasting analysis of the three considered distributions needs to be tailored towards the three market classification groups. In the context of both Developed and Emerging markets, the choice for the optimal distribution seems to vacillate between the Normal distribution and the Student-t distribution, with Japan being the only exception. This preference for one of these two distributions could be interpreted as evidence that these markets do not exhibit many relatively extreme returns for the 99 percent level of confidence. At the same time, the results have shown that the JSU distribution is favored solely within the context of Japan and four Frontier markets. As a reminder, the JSU distribution incorporates both a tail parameter and a skewness parameter, which induces higher forecasting series of $VaR_{99\%}$ than for the two competing distributions. Taken the results altogether, one could conclude that the Normal distribution and the Student-t distribution are no safe bet when one wants to capture the extreme outliers that are more common for the group of Frontier markets.

The ranking results are in line with the strong believe that risk managers should have a proper understanding on the return dynamics of each market and each market classification group. The results have shown that one should not blindly implement the Normal distribution in his market risk quantification framework, as this could lead to severe underestimation of risk for 17 out of the 30 considered markets, especially in the environment of Frontier markets. However, naively ignoring the Normal distribution would cause severe opportunity costs that could have easily been avoided for 13 out of the 30 considered markets.

III. The leverage specification against the non-leverage specification.

The last point of interest would be whether the inclusion of an asymmetric component in the framework of GARCH will add equal value for the three different market classification groups. The ranking results from Table 5.3 indicate that at a large scale, the leverage models are favored above their rivals, more specifically, for 26 out of the 30 considered markets. Based on this consistent outperformance, one could argue that the volatility dynamics corresponding to the 99 percent confidence level are in general more efficiently captured when a 'leverage-effect' term is added to the equation, i.e. returns and volatility seem to be negatively correlated in general. In terms of the asymmetric loss function, the leverage models produce series of VaR that are relatively close to the actual losses on turbulent trading days, which minimizes the 'underestimation' penalty term, and relatively close to the actual returns on normal trading days, which minimizes the 'overestimation' penalty term. In other words, the leverage models seem to be superior when it comes to finding a balance between the minimization of potential losses and the minimization of opportunity costs.

The countries that are in need of a special treatment regarding leverage are Japan, Nigeria, Slovenia, and Tunisia. It should be noted that the last three countries all belong to the Frontier market classification group. These ranking results could be interpreted as evidence that these four market do not exhibit an overall asymmetric correlation between returns and volatility, or it could be interpreted as evidence that these markets incurred some sudden and extreme losses that could not be fully captured by the leverage effect. The second explanation asks for a deeper justification. In the literature, there is the believe that during times of increased volatility and high turbulence some markets treat both positive and negative shocks in the far ends of the tails in a similar way. Meaning, a high positive shock will induce the same difference in volatility than a high negative shock will. As a result, the models with a leverage specification will tend to underestimate the downside reaction of the markets if such extreme and sudden behavior takes place in practice. Based on this explanation, one could argue that risk managers should take into deep consideration the chance of sudden and extreme losses at the 99 percent confidence level when judging the added value of the leverage component in the context of Japan and Frontier markets.

IV. The analysis of "Hypothesis One".

The issue that remains is whether the null hypothesis of this section [$H0_1$] needs to be rejected in favor of the alternative hypothesis [HA_1]. First of all, looking at the ranking results of the GARCH framework against the cEVT framework, one finds strong evidence that the former framework is the better choice across all three market classification groups and that the null hypothesis should not be rejected. However, this consistent outperformance should not be interpreted as evidence that these three groups of markets exhibit similar return dynamics. Instead, the pattern of outperformance should be attributed to the interaction between their unique characteristics and the design of the asymmetric loss function. One would even expect that the implementation of a higher confidence level would increase the need for the cEVT framework in order to cover for the extreme losses in the far ends of the tail, especially in the context of the more extreme Frontier markets. Nonetheless, this study is tailored towards the 99 percent level of confidence from the standard regulation framework of the BCBS, which means that the null hypothesis should not be rejected.

However, the ranking results belonging to the three distributions seem to deliver another conclusion. The optimal choice for Developed and Emerging markets seems to vacillate between the Normal distribution and the Student-t distribution. Conversely, the ranking results for Frontier markets illustrate that no less than four markets favor the JSU distribution. This discrepancy in preference could originate from the fact that the former two market classification groups tend to exhibit not many relatively extreme outliers, whereas the latter one is often more extreme by nature. An important implication of this conflict of preference would be that risk practitioners acknowledge that the Normal distribution and/or the Student-t distribution are not by definition a safe bet when one wants to capture the risk dynamics in Frontier markets. Instead, they should at least consider the implementation of the JSU distribution in order to minimize the potential damage of extreme losses. Overall, the ranking results suggest strongly that the null hypothesis should be rejected.

Lastly, the ranking results show that almost all markets favor the inclusion of the leverage effect in order to account for an asymmetric relation between return and volatility. However, again, the exception to the rule could be found in the context of Frontier markets. A possible explanation for this crucial discrepancy would be that these countries exhibit more extreme shocks, both positive and negative, and that they tend to treat extreme shocks equally during times of increased volatility and high turbulence. As a result, these ranking results provide evidence that the null hypothesis should be rejected.

Taking these results altogether, one could make the strong argument that the first null hypothesis should be **rejected**. The final rankings show that the relative forecasting accuracy of market risk quantification frameworks for $VaR_{99\%}$ differs substantially between country indices with heterogeneous market classifications. In particular, the results show that Frontier markets and their unique characteristics in the tails of their return distributions ask for a deviating approach in terms of both leverage inclusion and choice of distribution. This distinction in behavior stresses the importance for risk practitioners to acquire a deeper understanding of the unique return dynamics of each market classification group before setting up a market risk quantification policy. Overall, it speaks highly against the implementation of a 'one-size fits all' approach when forecasting Value at Risk at the 99 percent level of confidence.

TABLE 5.5 [A] – Expected Shortfall Ranking – DEVELOPED MARKETS

This table summarizes the values of p for the Expected Shortfall Bootstrap Test of McNeil & Frey (2000) and the corresponding rankings for each market within the *Developed* market classification. High values speak in favor of a model, while low values do the opposite. The green cells indicate that the respective $ES_{97.5\%}$ models have the number one ranking for the specific market. The abbreviations n, s, j, and c in the first column stand for Normal Distribution, Student-*t* Distribution, Johnson SU Distribution, and Conditional Extreme Value Theory. A p-value of 0.000 should be interpreted as something less than 0.0005, not zero.

	USA		Japan		Hong Kong		France		UK		Germany		Canada		Australia		Netherlands		Italy	
	p	RANK	*p*	RANK	*p*	RANK	*p*	RANK	*p*	RANK	*p*	RANK	*p*	RANK	*p*	RANK	*p*	RANK	*p*	RANK
n-GARCH	0.000	36	0.000	36	0.005	35	0.013	36	0.023	36	0.024	35	0.001	35	0.010	35	0.000	34	0.003	29
n-IGARCH	0.000	35	0.001	35	0.002	36	0.014	34	0.026	35	0.025	34	0.001	34	0.009	36	0.000	33	0.003	30
n-EGARCH	0.018	33	0.001	33	0.014	32	0.014	35	0.060	31	0.028	31	0.006	32	0.105	32	0.001	27	0.007	26
n-GJRGARCH	0.007	34	0.001	32	0.010	33	0.052	31	0.040	34	0.027	33	0.008	31	0.164	29	0.000	28	0.006	27
n-TGARCH	0.029	32	0.001	31	0.018	31	0.030	33	0.059	32	0.008	36	0.004	33	0.074	33	0.002	26	0.005	28
n-NAGARCH	0.042	31	0.001	34	0.007	34	0.048	32	0.058	33	0.028	32	0.000	36	0.051	34	0.002	25	0.019	25
s-GARCH	0.277	15	0.003	30	0.286	16	0.349	6	0.277	9	0.209	8	0.068	26	0.139	31	0.000	35	0.001	36
s-IGARCH	0.341	12	0.010	27	0.390	10	0.273	10	0.267	11	0.125	14	0.061	27	0.173	28	0.000	32	0.001	35
s-EGARCH	0.375	11	0.005	29	0.324	12	0.158	21	0.197	17	0.135	12	0.108	23	0.224	25	0.000	31	0.001	32
s-GJRGARCH	0.593	7	0.012	25	0.199	28	0.230	14	0.273	10	0.117	19	0.122	22	0.511	5	0.000	36	0.001	33
s-TGARCH	0.341	13	0.011	26	0.321	13	0.227	15	0.229	14	0.138	10	0.197	18	0.180	27	0.000	29	0.001	34
s-NAGARCH	0.274	16	0.006	28	0.154	30	0.191	18	0.190	18	0.273	7	0.019	30	0.147	30	0.000	30	0.002	31
j-GARCH	0.760	6	0.020	24	0.634	4	0.585	2	0.689	2	0.524	3	0.831	6	0.606	2	0.554	4	0.586	2
j-IGARCH	0.907	5	0.103	19	0.765	1	0.749	1	0.702	1	0.622	1	0.897	3	0.746	1	0.628	1	0.618	1
j-EGARCH	0.935	4	0.049	22	0.728	2	0.254	13	0.542	5	0.408	5	0.905	1	0.482	7	0.493	6	0.369	4
j-GJRGARCH	0.967	2	0.050	21	0.585	5	0.339	8	0.595	3	0.476	4	0.840	5	0.353	17	0.517	5	0.359	5
j-TGARCH	0.944	3	0.042	23	0.675	3	0.529	3	0.517	6	0.530	2	0.900	2	0.535	3	0.577	3	0.234	6
j-NAGARCH	0.982	1	0.053	20	0.514	8	0.271	11	0.584	4	0.356	6	0.897	4	0.411	13	0.612	2	0.495	3
c-n-GARCH	0.053	28	0.701	5	0.223	23	0.128	27	0.183	24	0.145	9	0.074	25	0.440	10	0.424	10	0.164	9
c-n-IGARCH	0.047	29	0.864	3	0.218	25	0.225	16	0.165	29	0.137	11	0.048	29	0.415	12	0.438	9	0.158	11
c-n-EGARCH	0.304	14	0.421	17	0.179	29	0.394	4	0.186	22	0.113	21	0.250	15	0.461	8	0.159	24	0.110	16
c-n-GJRGARCH	0.171	21	0.724	4	0.207	27	0.140	26	0.345	7	0.120	16	0.274	13	0.267	22	0.256	16	0.090	21
c-n-TGARCH	0.225	17	0.475	15	0.220	24	0.102	29	0.215	15	0.100	23	0.321	11	0.319	19	0.214	19	0.121	15
c-n-NAGARCH	0.137	23	0.567	10	0.501	9	0.209	17	0.185	23	0.064	29	0.060	28	0.226	24	0.233	17	0.083	23
c-s-GARCH	0.070	27	0.634	8	0.294	15	0.142	25	0.186	21	0.119	17	0.105	24	0.445	9	0.412	12	0.161	10
c-s-IGARCH	0.045	30	0.921	2	0.276	18	0.291	9	0.138	30	0.098	24	0.206	17	0.418	11	0.466	7	0.175	8
c-s-EGARCH	0.194	19	0.430	16	0.213	26	0.344	7	0.199	16	0.118	18	0.275	12	0.383	15	0.171	22	0.096	18
c-s-GJRGARCH	0.192	20	0.625	9	0.232	20	0.089	30	0.239	12	0.089	27	0.373	10	0.270	21	0.264	14	0.091	20
c-s-TGARCH	0.403	10	0.507	14	0.332	11	0.144	24	0.239	13	0.104	22	0.444	7	0.328	18	0.225	18	0.121	14
c-s-NAGARCH	0.123	24	0.550	11	0.543	6	0.152	23	0.170	28	0.049	30	0.145	19	0.222	26	0.181	21	0.075	24
c-j-GARCH	0.094	25	0.683	6	0.228	21	0.159	20	0.189	20	0.115	20	0.142	20	0.483	6	0.418	11	0.215	7
c-j-IGARCH	0.072	26	0.936	1	0.278	17	0.350	5	0.174	27	0.097	25	0.130	21	0.531	4	0.454	8	0.153	12
c-j-EGARCH	0.445	9	0.415	18	0.306	14	0.263	12	0.182	25	0.124	15	0.394	9	0.369	16	0.169	23	0.101	17
c-j-GJRGARCH	0.210	18	0.667	7	0.225	22	0.121	28	0.297	8	0.090	26	0.227	16	0.288	20	0.257	15	0.096	19
c-j-TGARCH	0.492	8	0.544	12	0.259	19	0.155	22	0.177	26	0.133	13	0.410	8	0.407	14	0.265	13	0.127	13
c-j-NAGARCH	0.164	22	0.543	13	0.542	7	0.173	19	0.190	19	0.076	28	0.251	14	0.256	23	0.213	20	0.087	22

TABLE 5.5 [B] – Expected Shortfall Ranking – EMERGING MARKETS

This table summarizes the values of p for the Expected Shortfall Bootstrap Test of McNeil & Frey (2000) and the corresponding rankings for each market within the *Emerging* market classification. High values speak in favor of a model, while low values do the opposite. The green cells indicate that the respective $ES_{97.5\%}$ models have the number one ranking for the specific market. The abbreviations n, s, j, and c in the first column stand for Normal Distribution, Student-*t* Distribution, Johnson SU Distribution, and Conditional Extreme Value Theory. A p-value of 0.000 should be interpreted as something less than 0.0005, not zero.

	China		India		South-Korea		Brazil		Mexico		Malaysia		Indonesia		Poland		Colombia		Peru	
	p	RANK	*p*	RANK	*p*	RANK	*p*	RANK	*p*	RANK	*p*	RANK	*p*	RANK	*p*	RANK	*p*	RANK	*p*	RANK
n-GARCH	0.001	36	0.066	31	0.201	29	0.035	36	0.009	35	0.011	33	0.002	33	0.000	35	0.000	35	0.012	36
n-IGARCH	0.009	33	0.065	32	0.070	36	0.070	34	0.005	36	0.031	31	0.005	31	0.000	36	0.000	36	0.020	31
n-EGARCH	0.007	35	0.029	35	0.152	33	0.093	31	0.027	33	0.009	35	0.001	36	0.001	27	0.000	33	0.012	35
n-GJRGARCH	0.009	34	0.055	33	0.175	31	0.049	35	0.012	34	0.005	36	0.005	32	0.001	26	0.000	34	0.014	33
n-TGARCH	0.012	32	0.020	36	0.150	34	0.074	32	0.037	30	0.010	34	0.002	34	0.001	29	0.001	32	0.013	34
n-NAGARCH	0.033	31	0.048	34	0.109	35	0.073	33	0.037	31	0.024	32	0.001	35	0.001	25	0.003	31	0.019	32
s-GARCH	0.572	10	0.359	14	0.762	7	0.492	30	0.078	29	0.313	26	0.157	30	0.000	34	0.008	29	0.030	28
s-IGARCH	0.495	14	0.343	15	0.934	4	0.690	27	0.036	32	0.330	24	0.418	17	0.000	33	0.013	27	0.061	25
s-EGARCH	0.773	7	0.544	8	0.537	18	0.838	16	0.191	20	0.439	21	0.277	25	0.001	30	0.008	28	0.024	30
s-GJRGARCH	0.466	15	0.567	7	0.628	14	0.572	29	0.130	27	0.292	27	0.198	28	0.001	28	0.006	30	0.032	27
s-TGARCH	0.950	4	0.489	9	0.589	16	0.752	24	0.209	17	0.560	18	0.369	22	0.000	32	0.017	26	0.026	29
s-NAGARCH	0.930	5	0.325	18	0.579	17	0.643	28	0.163	22	0.467	20	0.401	19	0.001	31	0.040	25	0.041	26
j-GARCH	0.682	8	0.577	6	0.997	2	0.915	12	0.226	14	0.701	11	0.780	6	0.187	22	0.935	4	0.381	7
j-IGARCH	0.635	9	0.710	4	0.998	1	0.991	1	0.246	12	0.708	9	0.891	3	0.269	13	0.965	2	0.539	1
j-EGARCH	0.962	3	0.786	2	0.872	6	0.954	7	0.476	3	0.903	1	0.958	1	0.197	19	0.925	5	0.345	13
j-GJRGARCH	0.876	6	0.712	3	0.618	15	0.989	2	0.456	6	0.702	10	0.764	7	0.313	10	0.888	6	0.372	8
j-TGARCH	0.983	2	0.798	1	0.920	5	0.951	8	0.418	7	0.870	3	0.910	2	0.169	23	0.948	3	0.336	15
j-NAGARCH	0.991	1	0.628	5	0.944	3	0.987	3	0.629	1	0.862	4	0.887	4	0.216	16	0.981	1	0.451	2
c-n-GARCH	0.250	19	0.195	27	0.637	13	0.802	19	0.271	10	0.150	30	0.399	20	0.654	6	0.349	23	0.279	24
c-n-IGARCH	0.217	23	0.157	30	0.728	10	0.759	23	0.509	2	0.404	22	0.300	24	0.751	2	0.695	9	0.295	22
c-n-EGARCH	0.223	22	0.213	24	0.217	27	0.811	18	0.132	26	0.540	19	0.276	26	0.280	12	0.534	16	0.330	16
c-n-GJRGARCH	0.194	24	0.312	19	0.346	24	0.939	10	0.193	18	0.655	15	0.367	23	0.433	9	0.349	22	0.289	23
c-n-TGARCH	0.514	13	0.230	23	0.421	19	0.760	22	0.217	16	0.656	14	0.413	18	0.192	21	0.666	10	0.366	9
c-n-NAGARCH	0.053	30	0.325	17	0.162	32	0.933	11	0.147	25	0.696	12	0.184	29	0.192	20	0.186	24	0.392	6
c-s-GARCH	0.298	17	0.208	25	0.701	12	0.784	21	0.271	11	0.324	25	0.520	14	0.700	4	0.468	18	0.304	19
c-s-IGARCH	0.245	20	0.160	29	0.730	9	0.743	25	0.457	5	0.353	23	0.393	21	0.748	3	0.615	11	0.299	20
c-s-EGARCH	0.180	26	0.236	22	0.271	25	0.812	17	0.149	24	0.690	13	0.480	15	0.293	11	0.550	14	0.353	11
c-s-GJRGARCH	0.259	18	0.446	11	0.352	22	0.949	9	0.224	15	0.620	16	0.617	9	0.460	8	0.376	21	0.340	14
c-s-TGARCH	0.523	12	0.268	20	0.359	20	0.841	15	0.235	13	0.597	17	0.265	27	0.202	18	0.559	13	0.392	5
c-s-NAGARCH	0.087	29	0.373	13	0.208	28	0.955	6	0.163	23	0.828	6	0.606	11	0.202	17	0.534	17	0.413	3
c-j-GARCH	0.307	16	0.197	26	0.710	11	0.787	20	0.319	8	0.262	29	0.547	13	0.659	5	0.425	19	0.297	21
c-j-IGARCH	0.227	21	0.163	28	0.734	8	0.737	26	0.469	4	0.289	28	0.566	12	0.819	1	0.715	8	0.314	17
c-j-EGARCH	0.181	25	0.254	21	0.196	30	0.878	13	0.126	28	0.732	7	0.641	8	0.248	14	0.542	15	0.351	12
c-j-GJRGARCH	0.162	27	0.474	10	0.353	21	0.957	5	0.274	9	0.729	8	0.616	10	0.498	7	0.600	12	0.304	18
c-j-TGARCH	0.544	11	0.391	12	0.348	23	0.869	14	0.166	21	0.875	2	0.438	16	0.217	15	0.741	7	0.392	4
c-j-NAGARCH	0.097	28	0.338	16	0.229	26	0.970	4	0.191	19	0.849	5	0.805	5	0.116	24	0.411	20	0.361	10

TABLE 5.5 [C] – Expected Shortfall Ranking – FRONTIER MARKETS

This table summarizes the values of *p* for the Expected Shortfall Bootstrap Test of McNeil & Frey (2000) and the corresponding rankings for each market within the *Frontier* market classification. High values speak in favor of a model, while low values do the opposite. The green cells indicate that the respective $ES_{97.5\%}$ models have the number one ranking for the specific market. The abbreviations n, s, j, and c in the first column stand for Normal Distribution, Student-*t* Distribution, Johnson SU Distribution, and Conditional Extreme Value Theory. A *p*-value of 0.000 should be interpreted as something less than 0.0005, not zero.

	Nigeria		Kazakhstan		Croatia		Romania		Kenya		Tunisia		Lebanon		Mauritius		Slovenia		Estonia	
	p	RANK	*p*	RANK	*p*	RANK	*p*	RANK	*p*	RANK	*p*	RANK	*p*	RANK	*p*	RANK	*p*	RANK	*p*	RANK
n-GARCH	0.050	35	0.001	34	0.004	36	0.005	32	0.001	34	0.012	31	0.004	31	0.009	36	0.049	31	0.000	32
n-IGARCH	0.065	19	0.003	31	0.004	34	0.003	35	0.002	31	0.010	32	0.001	32	0.014	33	0.010	33	0.000	36
n-EGARCH	0.051	34	0.001	33	0.006	32	0.003	34	0.000	35	0.006	34	0.001	34	0.036	31	0.005	36	0.000	34
n-GJRGARCH	0.056	27	0.000	36	0.004	35	0.005	31	0.001	32	0.005	35	0.001	33	0.013	34	0.008	34	0.000	31
n-TGARCH	0.048	36	0.000	35	0.006	33	0.004	33	0.003	30	0.007	33	0.000	36	0.034	32	0.007	35	0.000	35
n-NAGARCH	0.054	31	0.001	32	0.006	31	0.011	30	0.001	33	0.003	36	0.000	35	0.010	35	0.013	32	0.000	33
s-GARCH	0.088	9	0.146	30	0.032	27	0.147	24	0.321	26	0.219	29	0.891	14	0.162	29	0.680	16	0.037	28
s-IGARCH	0.117	2	0.188	29	0.021	30	0.136	27	0.571	22	0.301	27	0.974	2	0.268	26	0.548	21	0.051	25
s-EGARCH	0.087	10	0.239	26	0.024	29	0.049	28	0.167	29	0.476	23	0.968	3	0.084	30	0.309	29	0.024	30
s-GJRGARCH	0.086	11	0.196	28	0.044	23	0.145	25	0.319	27	0.174	30	0.858	17	0.318	25	0.475	24	0.049	26
s-TGARCH	0.083	12	0.404	24	0.025	28	0.002	36	0.000	36	0.508	22	0.955	6	0.250	27	0.442	26	0.062	24
s-NAGARCH	0.090	8	0.200	27	0.035	26	0.024	29	0.285	28	0.259	28	0.777	21	0.200	28	0.402	27	0.048	27
j-GARCH	0.105	5	0.540	18	0.125	19	0.283	19	0.711	18	0.398	26	0.955	7	0.666	4	0.884	9	0.035	29
j-IGARCH	0.137	1	0.718	3	0.061	22	0.343	14	0.962	5	0.608	21	0.986	1	0.800	1	0.936	3	0.070	23
j-EGARCH	0.113	3	0.757	2	0.043	25	0.401	9	0.679	19	0.730	19	0.900	13	0.549	8	0.636	19	0.083	21
j-GJRGARCH	0.102	6	0.683	6	0.089	20	0.426	5	0.834	14	0.467	24	0.769	22	0.586	5	0.793	14	0.071	22
j-TGARCH	0.099	7	0.703	4	0.044	24	0.385	12	0.722	17	0.706	20	0.945	10	0.799	2	0.739	15	0.110	19
j-NAGARCH	0.111	4	0.677	7	0.088	21	0.256	20	0.805	16	0.445	25	0.861	16	0.701	3	0.651	18	0.085	20
c-n-GARCH	0.053	32	0.497	22	0.329	9	0.391	11	0.558	24	0.976	4	0.756	23	0.556	6	0.919	5	0.425	13
c-n-IGARCH	0.068	17	0.534	20	0.343	6	0.340	15	0.564	23	0.949	12	0.726	25	0.392	21	0.895	7	0.376	16
c-n-EGARCH	0.071	15	0.440	23	0.264	17	0.319	16	0.992	1	0.967	7	0.403	30	0.506	11	0.393	28	0.300	18
c-n-GJRGARCH	0.056	28	0.828	1	0.330	8	0.139	26	0.648	20	0.953	10	0.742	24	0.551	7	0.580	20	0.418	14
c-n-TGARCH	0.079	13	0.652	9	0.307	11	0.304	17	0.635	21	0.955	9	0.516	28	0.442	14	0.269	30	0.359	17
c-n-NAGARCH	0.056	29	0.561	15	0.258	18	0.218	23	0.434	25	0.990	2	0.682	26	0.402	20	0.856	12	0.401	15
c-s-GARCH	0.056	30	0.662	8	0.339	7	0.419	6	0.936	10	0.943	13	0.961	5	0.382	22	0.873	10	0.520	10
c-s-IGARCH	0.062	21	0.610	12	0.518	2	0.408	7	0.971	3	0.966	8	0.966	4	0.378	23	0.952	2	0.570	5
c-s-EGARCH	0.073	14	0.546	17	0.269	15	0.286	18	0.932	12	0.984	3	0.835	19	0.466	13	0.655	17	0.469	11
c-s-GJRGARCH	0.051	33	0.631	10	0.357	4	0.235	22	0.933	11	0.907	15	0.947	8	0.537	9	0.867	11	0.548	8
c-s-TGARCH	0.066	18	0.585	13	0.356	5	0.638	1	0.821	15	0.970	6	0.543	27	0.421	18	0.497	22	0.468	12
c-s-NAGARCH	0.060	24	0.617	11	0.387	3	0.405	8	0.939	9	0.950	11	0.926	12	0.441	15	0.921	4	0.568	6
c-j-GARCH	0.060	25	0.687	5	0.282	14	0.516	4	0.943	8	0.940	14	0.927	11	0.429	17	0.890	8	0.543	9
c-j-IGARCH	0.060	23	0.557	16	0.548	1	0.604	3	0.982	2	0.991	1	0.849	18	0.319	24	0.902	6	0.598	4
c-j-EGARCH	0.061	22	0.342	25	0.265	16	0.238	21	0.910	13	0.973	5	0.814	20	0.436	16	0.480	23	0.553	7
c-j-GJRGARCH	0.057	26	0.537	19	0.300	13	0.344	13	0.949	6	0.874	18	0.945	9	0.480	12	0.801	13	0.683	2
c-j-TGARCH	0.068	16	0.523	21	0.302	12	0.393	10	0.945	7	0.901	16	0.477	29	0.527	10	0.468	25	0.643	3
c-j-NAGARCH	0.062	20	0.572	14	0.317	10	0.634	2	0.968	4	0.879	17	0.885	15	0.415	19	0.966	1	0.755	1

TABLE 5.6 – Number One Ranked Models of Expected Shortfall

This table provides an overview of the respective 97.5 percent Expected Shortfall models that enjoy the highest ranking within each market. It shows the models that are classified as optimal by the Expected Shortfall Bootstrap Test of McNeil & Frey (2000). In the column 'Leverage', the figure ✓ indicates whether the highest ranked model belongs to the class of leverage GARCH models. A similar structure is chosen for the column 'EVT', where the figure ✓ denotes whether the highest ranked model comes from the framework of Conditional Extreme Value Theory.

DEVELOPED MARKETS				
Country	**Distribution**	**Model**	**Leverage**	**EVT**
USA	JSU	NAGARCH	✓	
Japan	JSU	IGARCH		✓
Hong Kong	JSU	IGARCH		
France	JSU	IGARCH		
United Kingdom	JSU	IGARCH		
Germany	JSU	IGARCH		
Canada	JSU	EGARCH	✓	
Australia	JSU	IGARCH		
Netherlands	JSU	IGARCH		
Italy	JSU	IGARCH		

EMERGING MARKETS				
Country	**Distribution**	**Model**	**Leverage**	**EVT**
China	JSU	NAGARCH	✓	
India	JSU	TGARCH	✓	
South-Korea	JSU	IGARCH		
Brazil	JSU	IGARCH		
Mexico	JSU	NAGARCH	✓	
Malaysia	JSU	EGARCH	✓	
Indonesia	JSU	EGARCH	✓	
Poland	JSU	IGARCH		✓
Colombia	JSU	NAGARCH	✓	
Peru	JSU	IGARCH		

FRONTIER MARKETS				
Country	**Distribution**	**Model**	**Leverage**	**EVT**
Nigeria	JSU	IGARCH		
Kazakhstan	Normal	GJRGARCH	✓	✓
Croatia	JSU	IGARCH		✓
Romania	Student-*t*	TGARCH	✓	✓
Kenya	Normal	EGARCH	✓	✓
Tunisia	JSU	IGARCH		✓
Lebanon	JSU	IGARCH		
Mauritius	JSU	IGARCH		
Slovenia	JSU	NAGARCH	✓	✓
Estonia	JSU	NAGARCH	✓	✓

5.2 Backtesting Expected Shortfall (97.5%)

The final part of this study is devoted to the selection of the 'optimal' Expected Shortfall model for each individual market portfolio. To achieve this goal, a large quantity of financial risk models frameworks has been estimated for each market, and based on the criteria specified, these models were utilized to arrive at the 'best' risk models. An important aspect of Expected Shortfall would be that its field of backtesting is still in its infancy and that there is yet no clear consensus in literature on what should be the standard backtesting method. At the time being, only a few approaches have been deemed to be accurate or sufficient. An exception would be the Bootstrap ES test of McNeil & Frey (2000), which is a widely respected approach that measures the discrepancy between realized losses and Expected Shortfall estimates on days when Value at Risk violations take place. For correctly specified models, these discrepancies should form a sample from a distribution with a mean equal to zero.

Table 5.5 [A-C] documents the ranking results from the ES Bootstrap test for the 36 implemented models when applied to the thirty individual market indices. The backtest delivers a value between zero and one, where high values speak in favor of a model and low values do the exact opposite. These results are then ranked from 1 to 36, where models with the number one ranking are deemed to be the most accurate ones in terms of forecasting performance. As a reference guide, the green cells indicate that the respective risk models enjoy the highest ranking for that specific market. In order to provide the most convenient overview of all the number one ranked ES models, Table 5.6 was formed. Similar to the table that was created for VaR in Section 5.1.2, this table adds an indicator column for both the leverage term and the conditional EVT specification in order to ease the detection of similarities and discrepancies in model performance between the three market classification groups.

HYPOTHESIS TWO	
$H0_2$	The underlying market classification of a country index is unrelated to the relative forecasting performance of $ES_{97.5\%}$ models.
HA_2	The underlying market classification of a country index is related to the relative forecasting performance of $ES_{97.5\%}$ models.

The main purpose of this section is to test whether the abovementioned null hypothesis needs to be rejected in favor of the alternative hypothesis. In order to enable a comparative analysis, the examination of the results will first be divided into three main subjects, from which at the end a conclusion will be extracted.

I. <u>The simplistic GARCH framework against the sophisticated cEVT framework.</u>

The first area of interest would be the performance of the standard GARCH framework against the more sophisticated framework of cEVT when forecasting Expected Shortfall at the 97.5 percent level of confidence. The ranking results from Table 5.5 [A-C] and Table 5.6 depict that for Developed markets only Japan favors the cEVT framework, that for Emerging markets only Poland favors the cEVT framework, and that for Frontier markets no less than 7 out of the 10 considered markets favor the cEVT framework. These findings are in sharp contrast to the ranking results of Value at Risk, where only Japan favored the cEVT framework. This difference in forecasting performance across the three market classification groups could be attributed to a combination of four important elements, namely (i.) the design of Expected Shortfall, (ii.) the design of cEVT, (iii.) the design of the Expected Shortfall Bootstrap backtest, and (iv.) the unique characteristics exhibited by the Frontier markets.

(i.) First of all, there exists a sharp contrast between the design of Value at Risk and the design of Expected Shortfall. The former risk measure seeks to answer the question “How bad can things get?”, whereas the latter looks at “If things go bad and VaR is exceeded, how much can we expect to lose?”. In terms of mathematical interpretation, this means that VaR only looks at the quantile that corresponds with the level of confidence, while ES also looks behind the quantile. In other words, the ES measure needs to capture the entire tail, so also the most extreme outliers, whereas the VaR measure will ignore these extreme outliers. An important implication of this behavior would be that although the confidence level of ES (97.5%) in this study is set lower than the confidence level of VaR (99%), one should not make the mistake that VaR looks at higher losses than ES.

(ii.) Secondly, one should recognize the substantial disadvantage that is associated with the GARCH framework. The approach has a naïve focus on the whole return distribution, while ES is in essence only related to the far ends of the probability distribution. So instead of forcing a single distribution on the entire return series, the cEVT framework offers an innovative alternative by explicitly modeling the tail regions of the distributions. This allows the framework to predict extreme and rare losses, possibly even more extreme than the ones observed within the range of available return observations.

(iii.) Thirdly, there exists a discrepancy between the implemented backtesting procedures of VaR and ES, respectively. The former one takes into account the entire ‘out-of-sample’ period, from which it calculates a daily ‘underestimation’ penalty term and a daily ‘overestimation’ penalty term. Conversely, the backtesting procedure of ES is only interested in days when the forecasted value of VaR is violated, i.e. the most turbulent trading days. This design is in accordance with the concept of ES, which focuses on the calculation of how large losses can become on average in situations where the value of VaR is violated. As a result, the overestimation on normal trading days should not play a significant role in the determination of the most accurate ES model. This decision is supported by the McNeil & Frey (2000), who remark that underestimation of ES is the more likely direction of failure and the more dangerous way in which ES can be wrong.

(iv.) Lastly, the unique characteristics exhibited by the three different market classification groups are of high importance in terms of return dynamics. A significant part of literature has been devoted to the difference between Developed and Emerging markets, and in general they have documented that the latter are subject to different cultural, institutional, econom, and political circumstances. A crucial distinguishing feature of Emerging markets in the context of market risk quantification would be that the low level of liquidity present in these markets causes them to exhibit low volatility in general, but where frequent systematic financial shocks are giving rise to substantial extreme volatility (Božović & Totić, 2015). At the same time, the dynamics encountered in Frontier markets are presumed to be even more complex than the ones observed for Emerging markets. These markets have shown to exhibit more violent patterns of extreme and sudden negative returns in times of turbulence. Issues that plague some Emerging markets, such as political instability, weak regulations, and illiquidity, are all amplified in many of the existing Frontier markets. By definition, the Frontier markets are a class described best by small, illiquid, less accessible, and less known Emerging markets. Due to their characteristic of extreme illiquidity, the equity prices tend to respond more strongly to selling and trading pressures during flight-to-safety episodes where investors try to reduce their risk exposure quickly (Chan-Lau, 2011). As a consequence, the extreme events that concur the Frontier markets are expressed more severely in terms of extreme and sudden negative returns.

Taking these four concepts altogether, one could argue that the relative outperformance of the cEVT framework within the context of Frontier market originates from the fact that these markets exhibit return distributions that are more likely to include very large and sudden exceedances, and that the numerous regime switches in these markets cause the parameters of the return distribution to change more frequently. In other words, these markets have some outliers that are hard to be derived directly from the available data. As a consequence, the standard GARCH framework will be a likely victim of systematical underestimation of financial risk when expressed in ES. Conversely, the cEVT allows for a prediction of large and rare losses that are not even within the range of available observations from the past. As underestimation of ES is the more likely direction of failure on turbulent trading days and the more dangerous way in which ES can be wrong, the cEVT framework has a higher chance of being preferred in the context of Frontier markets and their unique characteristics.

At the same time, the ranking results of Table 5.5 [A-C] and Table 5.6 depict that only one Emerging market prefers the framework of cEVT. This suggests that the sudden and extreme returns that concur Emerging markets are not as pronounced as those that are observed for Frontier markets. This difference is fully in line with the findings of the existing literature on this subject, as they denote that the issues that plague many Emerging markets are amplified in the environment of Frontier markets. Overall, one could conclude that the more sophisticated framework of cEVT seems to have the most added value in the context of Frontier markets, which highlights the importance for risk practitioners to be very aware of the unique risk dynamics before investing into the relatively new investment opportunity.

II. The relative performance of the three error distributions.

The next analysis that needs to be conducted is related to the three distributions: the Normal distribution, the Student-t distribution, and the JSU distribution. The ranking results of Table 5.5 [A-C] and Table 5.6 reveal that models with JSU innovations enjoy the highest position for almost all markets, more specifically, for 27 out of the 30 markets. This gives spark to the prominent question why the JSU is so dominant in the context of ES, and simultaneously so weak for VaR. As a reminder, Section 5.1.2 showed that for VaR only 5 out of the 30 considered markets favored the JSU distribution.

First of all, the consistent underperformance of the JSU distribution within the context of VaR should be attributed to the design of the VaR backtesting procedure and the chosen level of confidence. The JSU distribution includes both a skewness and a fat tail parameter, which implies that it will produce forecasts of VaR that are systematically higher than its two competitors when the return distribution exhibits both stylized facts. In a situation where the Normal distribution or the Student-t distribution forecasts a series of VaR that falls in the acceptable interval of violations, the chance is substantial that the minimization of the 'underestimation' penalty term of these two distributions outweighs the minimization of the 'overestimation' penalty term of the JSU distribution. However, one would expect that a confidence level set higher than 99 percent would increase the importance of extreme outliers and the suitability of JSU.

Secondly, the outperformance of the JSU distribution within the context of ES could be attributed to the backtesting framework of ES, which acknowledges that underestimation of risk is the more likely and the more dangerous way in which ES can be wrong. The concept of ES focusses on how large the losses can get on average in situations where VaR is violated, and the most important issue would then be to minimize the underestimation of the largest losses that occur in the far ends of the return distribution. As the JSU distribution will by design be a less likely victim of underestimation if the empirical return distribution exhibits significant excess kurtosis and skewness, this specification will produce levels of VaR that are closer to the far extremes.

An important observation would be that the preference for the JSU distribution is violated in only three countries, who all belong to the Frontier classification. A possible explanation could be that these three markets exhibit more frequent regime switches, with deviating periods of positive and negative skewness. This could result in severe risk underestimation at the beginning of a new period if extreme returns occur. Another explanation could be that these countries have a relatively large gap between the smallest and biggest outliers, i.e. the normal trading days exhibit only small outliers and the turbulent trading days exhibit only very extreme outliers. These dynamics will make the average underestimation of risk larger for series of ES that are forecasted by the JSU distribution, as the other two distributions will lower the average underestimation of risk by including a few smaller violations. This mechanism prevents indirectly the scenario that the ES is set so high that all outliers are automatically captured.

III. The leverage models against the non-leverage models.

The last point of interest would be whether or not the leverage specification adds equal value for all three market classification groups when forecasting Expected Shortfall at the 97.5 percent level of confidence. Overall, the ranking results of Table 5.6 illustrate that the non-leverage IGARCH model is the favored model in 16 out of the 30 considered markets. These high rankings of the IGARCH model could be interpreted as evidence that these 16 markets do not exhibit an asymmetric correlation between return and volatility. However, this would contradict with the results from the previous section, where 26 out of the 30 considered markets preferred a VaR model that incorporated a leverage specification. If a country exhibits a general leverage effect, this should be visible for the entire return distribution and all confidence levels. Therefore, a better explanation for this pattern could be found in the reasoning that, during times of increased volatility and high turbulence, some markets treat extreme positive and extreme negative shocks in a similar fashion. In other words, the highest gains and highest losses that occur in these markets induce similar shocks in volatility. This behavior influences the estimates of ES more prominently than it does for VaR, as the former risk measure is more tailored towards the extreme outliers that lie far in the tail. As a result, the negative shocks that happen in the far end of the tail are larger than one would expect based on the leverage effect, i.e. based on the previous return structure.

The most peculiar result in terms of forecasting performance could be observed in the context of the three individual market classification groups. To reveal this distinctive element, one should look at the combination of the leverage specification and the cEVT framework. In order to ease the detection of the pattern, the following two groups are designed:

- **Group 1** – This group consists of the markets that prefer a leverage model <u>without</u> a conditional Extreme Value Theory specification.
- **Group 2** – This group consists of the markets that prefer a leverage model <u>with</u> a conditional Extreme Value Theory specification.

The results from Table 5.7 illustrate that the first group consists of markets that belong to either the Developed or the Emerging market classification group. The combination of a leverage specification and a standard GARCH framework should be interpreted as evidence that these markets exhibit an asymmetric reaction between returns and volatility, where the size of the downward reaction can be derived directly from the available return data. In other words, these losses are not very extreme and sudden, which makes the sophisticated framework of cEVT undesirable. The second group, however, appears only within the context of Frontier markets. In the case of Frontier markets, five countries prefer a leverage specification, and all these countries combine this component with the sophisticated cEVT framework. This connection could be interpreted as evidence that these markets exhibit an asymmetric response between returns and volatility, and that these responses can be very large and sudden, which makes them too severe to be extracted by the standard GARCH framework from the available return data.

This sparks the prominent question why this particular combination of a leverage component and cEVT only occurs within the context of Frontier markets. A reasonable explanation would be the difference in the general liquidity between the three market classification groups. The Developed and Emerging markets have the benefit of a higher liquidity in general, which increases the chance that investors can express their feelings

TABLE 5.7 – Group 1 and Group 2

This table provides an overview of markets that correspond with the selection criteria of either Group 1 or Group 2. The first group consists of markets that prefer a leverage model without a conditional Extreme Value Theory specification. The second group consists of markets that prefer a leverage model with a conditional Extreme Value Theory specification. The 'n.a.' means that there is no country from that specific market classification group that satisfies the selection criteria.

	Group 1	Group 2
Developed	Canada and USA.	n.a.
Emerging	China, Colombia, India, Indonesia, Malaysia, and Mexico.	n.a.
Frontier	n.a.	Estonia, Kazakhstan, Kenya, Romania, and Slovenia.

accurately and easily. Conversely, the extreme illiquidity observed for Frontier markets could cause equity prices to respond more strongly to selling and trading pressures during flight-to-safety episodes where investors try to reduce their risk exposure. In other words, for Frontier markets, when there is extremely low liquidity, the investors magnify the downward movements, making it such that the outliers are becoming so extreme that only cEVT can accurately account for them.

IV. The analysis of the hypothesis.

The issue that remains is whether the null hypothesis of this section [$H0_2$] needs to be rejected in favor of the alternative hypothesis [HA_2]. First of all, the ranking results demonstrate that the framework of cEVT is only consistently preferred within the context of Frontier markets. The other two market classification groups favor the standard framework of GARCH instead. This important inequality in preference should be attributed to the fact that Frontier markets are more likely to exhibit return distribution with very large and sudden exceedances. In other words, Frontier markets exhibit outliers that are much harder to be derived directly from the data. The design of the cEVT tackles this problem by allowing for a prediction of large and rare losses that are not even within the range of available return observations. Taking into account that the concept of Expected Shortfall is mostly related to the far ends of the tails, the framework of cEVT has a substantially higher chance by nature of being preferred within the context of Frontier markets. On a side note, the fact that only one Emerging market favors the framework of cEVT suggests that the sudden and extreme returns exhibited by this market classification group are not as pronounced as those observed for Frontier markets. Overall, the ranking results are in support of the alternative hypothesis, which highlights the importance for risk practitioners to be very aware of the unique risk dynamics before investing in the relatively new environment of Frontier markets.

Another interesting result would be that the particular combination of a leverage component and the cEVT framework only occurs within the context of Frontier markets. This could be explained by the difference in liquidity between the three market classification groups. The countries that belong to Frontier market classification tend to face the disadvantage of extreme illiquidity, which cause the equity prices to respond more strongly to selling and trading pressures during flight-to-safety episodes where investors try to reduce their risk exposure quickly. In other words, when there is extremely low liquidity during turbulent times, the investors will magnify the downward movements, which makes it such that the outliers are becoming so extreme that only cEVT can accurately account for it. Furthermore, the relative rankings for the three considered error distributions also support the alternative hypothesis, as the preference for the JSU distribution is only violated within the context of three Frontier countries.

Taking these analyses altogether, the ranking results strongly suggest that the null hypothesis needs to be **rejected** in favor of the alternative hypothesis. The accuracy of market risk quantification frameworks when forecasting $ES_{97.5\%}$ seems to change substantially across the specified market classification groups. The most distinctive inequality in performance is observed in the environment of Frontier markets, where the unique characteristics of their return distributions generally demand a different approach regarding all layers of the market risk quantification framework, especially regarding the choice between a standard GARCH framework or a sophisticated cEVT framework.

(6) CONCLUSION

The global financial crisis of 2007-2009 and the subsequent European debt crisis have repeatedly underlined the importance for banks and regulators alike to re-examine their ways in which they perform financial risk management. The crises have reminded the world that distressed markets are capable of producing extreme losses that are far in excess of what many popular risk models can predict. This knowledge has even let the BCBS to announce a complete revision of their regulations in which they replace the standard risk measure of Value at Risk by Expected Shortfall. At the same time, the global character of the crisis of 2007-2009 raises the deep concern among investors whether traditional Emerging markets are still capable of delivering diversification benefits. As a result, investors seek for new investment opportunities that are by nature less affected by the increasing rate of globalization. Over the last decade, this urge of finding a new investment platform has created a huge increase of interest among investors in countries belonging to the Frontier market classification. While investors and academics tend to agree on their added value in international portfolios, they seem to be less certain about their risk profile. Empirical literature has documented that this group of markets exhibit several unique and extreme characteristics that influence the properties of return series significantly. This anomaly suggests that it would be a dangerous misnomer to treat international markets as being completely homogeneous when performing financial risk measurement.

Therefore, in order to capture important present-day challenges for the financial industry, this study examined whether a 'one-size fits all' approach works properly when forecasting market risk for country indices with heterogeneous market classifications (i.e. Developed, Emerging and Frontier), both under current and future BCBS regulations. That is, "Will the best risk model for country A also be the best risk model for country B?". In order to achieve maximum relevance and practicality, this study covered a large number of country indices belonging to all three market groups, with return data spanning a period of more than 11 years. At the same time, this study adopted sophisticated risk quantification frameworks, included innovative backtesting methods, and performed daily recalculations for each included parameter.

All in all, this study has provided strong evidence against implementing a 'one-size fits all' approach when measuring risk in different markets, both under current and future regulations of the BCSB. In particular, the ES backtesting results showed clear evidence that a homogeneous approach of risk management among country indices would bear serious consequences if one included countries with a Frontier market classification. It was argued that this difference originates from the fact that these countries tend to exhibit extraordinary return dynamics: they are more likely to show return distributions with very extreme and sudden outliers. In terms of model preference, the ES backtesting results revealed that the

Frontier markets are in need of a more sophisticated quantification framework such as conditional Extreme Value Theory. Conversely, the results showed that both Developed and Emerging markets are well described by the more simplistic GARCH framework. The similarities between Developed and Emerging markets suggest that the sudden and extreme returns that concur Emerging markets are not as pronounced as those observed for Frontier markets.

In practice, this means that if one would decide to apply the optimal risk models for Developed or Emerging markets (i.e. GARCH models) on Frontier markets, one would have a higher chance of severe risk underestimation and, potentially, bankruptcy. On the other hand, if one would apply the optimal risk models for Frontier markets (i.e. conditional EVT models) on either Developed or Emerging markets, one would have an increased chance of severe risk overestimation and high opportunity costs.

Therefore, the major implication of this study would be that risk practitioners need to acknowledge the Frontier markets' unique characteristics and that they should be very cautious before considering homogeneous financial risk measurement. More specifically, if one would decide to invest in country indices with a Frontier market classification, one should deeply consider the usage of techniques that are better capable of dealing with extreme and sudden fluctuations – e.g. the framework of conditional Extreme Value Theory – and they should discourage the employment of nothing but traditional methodologies such as GARCH that do only provide accurate risk measurement for Developed and Emerging markets. It should be pointed out that although the implementation of cEVT is more tedious than simply relying on the standard framework of GARCH, the results obtained are worth the extra effort and present the user with a better picture of the tail risk involved in Frontier markets.

A limitation of this study could be that it included only one confidence level per risk measure, namely 97.5% for Expected Shortfall and 99.0% for Value at Risk. These confidence levels were chosen in order to match with the regulation framework of the BCBS. However, regarding further research, it could be of high interest to break with these standard regulations of the BCBS and to implement different confidence levels for VaR and ES that are more tailored towards the unique levels of risk aversion of the user. For instance, it would be interesting to examine whether a higher (lower) confidence level increases (decreases) the need of cEVT when investing in Frontier markets. Another attractive direction for further research could be to look at countries that have undergone a market reclassification during the time sample (e.g. Argentina, which was transitioned from an Emerging market to a Frontier market in May 2009) and to test whether this transition has any significant effect on the forecasting performance of risk models.

(7) REFERENCES

Abdalla, S. Z. (2012). Modelling Exchange Rate Volatility using GARCH Models: Empirical Evidence from Arab Countries. *International Journal of Economics and Finance, 4 (3)*, 216-229.

Acerbi, C., & Tasche, D. (2002). On the Coherence of Expected Shortfall. *Journal of Banking & Finance, 26 (7)*, 1487-1503.

Aggarwal, R., Inclan, C., & Leal, R. (1999). Volatility in Emerging Stock Markets. *Journal of Financial and Quantitative Analysis, 34*, 33-55.

Akgiray, V. (1989). Conditional Heteroscedasticity in Time Series of Stock Returns: Evidence and Forecasts. *Journal of Business*, 55-80.

Alexander, C., Lazar, E., & Stanescu, S. (2011). Approximations to GARCH Aggregated Returns Distributions with Applications to VaR and ETL. *ICMA Centre Discussion Papers in Finance DP 2011-08*, 1-26.

Angelidis., T., Benos, A., & Degiannakis, S. (2004). The Use of GARCH Models in VaR Estimation. *Statistical Methodology, 1*, 105-128.

Artzner, P., Delbaen, F., Eber, J., & Heath, D. (1999). Coherent Measures of Risk. *Mathematical Finance, 9*, 203-228.

Balkema, A. A., & Heen, L. D. (1974). Residual Life Time at Great Age. *The Annals of Probability*, 792-804.

Basel Committee on Banking Supervision. (2011). *Messages from the Academic Literature on Risk Measurement for the Trading Book*. Bank for International Settlements.

Basel Committee on Banking Supervision. (2012). *Fundamental Review of the Trading Book*. Bank for International Settlements.

Basel Committee on Banking Supervision. (2016). *Minimum Capital Requirements for Market Risk*. Bank for International Settlements.

Bekaert, G., & Harvey, C. R. (2003). Emerging Markets Finance. *Journal of Empirical Finance, 10*, 3-55.

Bello, A., & Adedokun, L. (2011). Analysis of the Risk-Return Characteristics of the Quoted Firms in the Nigerian Stock Market. *International Journal of Business and Social Science, 17*, 158-165.

Bera, A. K., & Higgings, M. L. (1993). ARCH models: Properties, Estimation and Testing. *Journal of Economic Sruveys, 7 (4)*, 305-366.

Berger, D., Pukthuanthong, J., & Yang, J. J. (2011). International Diversification with Frontier Markets. *Journal of Financial Economics, 101*, 227-242.

Berkowitz, J., & O'Brien, J. (2002). How Accurate are Value-at-Risk Models at Commerical Banks? *The Journal of Finance, 57 (3)*, 1093-1111.

Bidarkota, P. V., & Dupoyet, B. V. (2004). The Impact of Fat Tails on Equilibrium Rates of Return and Term Premia. *Working Paper.*

Black, F. (1976). Studies of Stock Market Volatility Changes. *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 177-181.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Economtircs, 31 (3)*, 307-327.

Bollerslev, T. (1987). A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. *The Review of Economics and Statistics*, 542-547.

Bollerslev, T., Chou, R. Y., & Kroner, K. F. (1992). ARCH Modeling in Finance: A Review of the Theory and Empirical Evidence. *Journal of Econometrics, 52 (1-2)*, 5-59.

Box, G. E., & Jenkins, G. M. (1970). *Time Series Models for Forecasting and Control.* San Francisco.

Božović, M., & Totić, S. (2015). Tail Risk in Emerging Markets of Southeastern Europe. *Applied Economics*, 1-14.

Božović, M., & Totić, S. (2016). Tail Risk in Emerging Markets of Southeastern Europe. *Applied Economics, 48 (19)*, 1785–1798.

Brooks, C. (2008). *Introductory: Econometrics for Finance, 2nd Edition.* Cambridge: Cambridge University Press.

Cappiello, L., Engle, R. F., & Sheppard, K. (2006). Asymmetric Dynamics in the Correlations of Global Equity and Bond Returns. *Journal of Financial Econometrics, 4 (4)*, 537-572.

Chan-Lau, J. A. (2011). *Frontier Markets: Punching Below their Weight? A Risk Parity Perspective on Asset Allocation.* Center for Emerging Market Enterprises.

Chen, J., Kong, H., & Stein, J. (2000). Forecasting Crashes: Trading Volume, Past Returns and Conditional Skewness in Stock Prices. *Journal of Financial Economics, 61*, 345-381.

Christie, A. (1982). The Stochastic Behaviour Common Stock Variances: Value, Leverage, and Interest Rate Effects. *Journal of Financial Economics, 10*, 407-432.

Christoffersen, P. F. (1998). Evaluating Interval Forecasts. *International Economic Review*, 841-862.

Christoffersen, P. F. (2003). *Elements of Financial Risk Management.* San Diego: Academic Press.

Christoffersen, P. F., & Diebold, F. X. (2000). How Relevant is Volatility Foreacsting for Financial Risk Management? *Review of Economics and Statistics, 82 (1)*, 12-22.

Christoffersen, P. F., & Pelletier, D. (2004). Backtesting Value-at-Risk: A Duration-Based Approach. *Journal of Financial Econometrics, 2 (1)*, 84-108.

Coles, S., Bawa, J., Trenner, L., & Dorazio, P. (2001). *An Introduction to Statistical Modeling of Extreme Values.* Springer: London.

Cont, R. (2001). Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues. *Quantitative Finance, 1*, 223-236.

Damodaran, A. (1985). Economic Events Information Strucutre, and the Return-Generating Process. *Journal of Financial and Quantitative Analysis*, 425-434.

Danielsson, J. (2002). The Emperor has No Clothes: Limits to Risk Modelling. *Journal of Banking & Finance, 26*, 127301296.

Danielsson, J. (2011). *Financial Risk Forecasting: The Theory and Practice of Forecasting Market Risk with Implementation in* R *and Matlab.* Oxford: John Wiley & Sons, Ltd.

Danielsson, J., Haan, L. D., Peng, L., & Vries, C. D. (2001). Using a Bootstrap Method to Choose the Sample Fraction in Tail Index Estimation. *Journal of Multivariate Analysis, 76 (2)*, 226-248.

Demirgüç-Kunt, A., & Levine, R. (1996). *The Financial Systems and Public Enterprise Reform: Concepts and Cases.* London: Routledge.

Dimitrakopoulos, D. N., Kavussanos, M. G., & Spyrou, S. I. (2010). Value at Risk Models for Volatile Emerging Markets Equity Portfolios. *The Quarterly Review of Economics and Finance, 50*, 515–526.

Dowd, K. (2005). *Measuring Market Risk, Second Edition.* Oxford: John Wiley & Sons, Ltd.

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling Extremal Events: For Insurance and Finance.* Berlin: Springer-Verlag.

Embrechts, P., Resnick, S. I., & Samorodnitsky, G. (1999). Extreme Value Theory as a Risk Management Tool. *North American Actuarial Journal, 3 (2)*, 30-41.

Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica: Journal of the Econometric Society*, 987-1007.

Engle, R. F. (2001). GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics. *Journal of Economic Perspectives, 15 (4)*, 157-168.

Engle, R. F., & Bollerslev, T. (1985). Modelling the Persistence of Conditional Variances. *Econometric Reviews, 5 (1)*, 1-50.

Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business & Economic Statistics, 22 (4)*, 367-381.

Engle, R. F., & Ng, V. K. (1993). Measuring and Testing the Impact of News on Volatility. *The Journal of Finance, 48 (5)*, 1749-1778.

Engle, R. F., & Patton, A. J. (2001). What Good is a Volatility Model. *Quantitative Finance, 1*, 237-245.

Errunza, V. (1983). Emerging Markets: New Opportunity for Improving Global Portfolio Performance. *Financial Analysts Journal, 39*, 51-58.

Fama, E. F. (1965). The Behavior of Stock Market Prices. *Journal of Business, 38*, 34-105.

Fisher, R. A., & Tippett, L. H. (1928). Limiting Forms of the Frequency Distribution of the Largest and Smallest Member of a Sample. *Mathematical Proceedings of the Cambridge Philosophical Society, 24 (2)*, 180-190.

French, K., Schwert, W., & Stambaugh, R. (1987). Expected Stock Returns and Volatility. *Journal of Financial Economics, 19*, 3-29.

Frey, R., & Michaud, P. (1997). The Effect of GARCH-type Volatilities on Prices and Payoff-Distributions of Derivative Assets - A Simulation Study. *Working Paper*.

Gençay, R., & Selçuk, F. (2004). Extreme Value Theory and Value at Risk: Relative Performance in Emerging Markets. *International Journal of Forecasting, 20*, 287-303.

Giot, P. (2005). Relationships between Implied Volatility Indices and Stock Index Returns. *Journal of Portfolio Management, 31*, 92-100.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *The Journal of Finance, 48 (5)*, 1779-1801.

Gnedenko, B. (1943). Sur la Distribution Limite du Terme Maximum d'Une Serie Aleatoire. *Annals of Mathmatics*, 423-453.

Goeij, P. D., & Marquering, W. (2003). Modeling the Conditional Covariance between Stock and Bond Returns: A Multivariate GARCH Approach. *Working Paper*.

Gokcan, S. (2000). Forecasting Volatility of Emerging Stock Markets: Linear Versus Non-Linear GARCH Models. *Journal of Forecasting, 19 (6)*, 499-504.

González-Rivera, G., Lee, T., & Mishra, S. (2004). Forecasting Volatility: A Reality Check Based on Option Pricing, Utility Function, Value-at-Risk, and Predictive Likelihood. *International Journal of forecasting, 20 (4)*, 629-645.

Group, R. (1996). *RiskMetrics – Technical Document*. New York: J.P. Morgan/Reuters.

Guldimann, T. (2000). The Story of RiskMetrics. *Risk, 13 (1)*, 56-58.

Gustafsson, M., & Lundberg, C. (2009). An Empirical Evaluation of Value at Risk.

Hendricks, D. (1996). Evaluation of Value-at-Risk Models using Historical Data. *Economic Policy Review Federal Reserve Bank of New York, 2 (1)*, 39-67.

Hibbert, A., Daigler, R., & Dupoyet, R. (2008). A Behavioral Explanation for the Negative Asymmetric Return-Volatility Relation. *Journal of Banking and Finance, 32*, 2254-2266.

Holton, G. A. (2002). History of Value at Risk: 1922-1998. *Working Paper*.

Hull, J. C. (2015). *Options, Futures, and Other Derivatives, Ninth Edition.* New Jersey: Pearson Education, Inc.

Jarque, C. M., & Bera, A. K. (1980). Efficient Tests for Normality, Homoscedasticity and Serial Independence of Regression Residuals. *Economics Letters, 6 (3)*, 255-259.

Johansson, J., & Engblom, A. (2015). Models for Credit Risk in Static Portfolios. *Working Paper.*

Jorion, P. (2001). *Value at Risk: The New Benchmark for Managing Financial Risk.* New York: McGraw-Hill.

Kerkhof, J., & Melenberg, B. (2004). Backtesting for Risk-Based Regulatory Capital. *Journal of Banking & Finance, 28 (8)*, 1845-1865.

Kjelsson, B. (2013). Forecasting Expected Shortfall: An Extreme Value Approach.

Kupiec, P. H. (1995). Techniques for Verifying the Accuracy of Risk Measurement Models. *The Journal of Derivatives, 3 (2)*, 73-84.

Maghyereh, A. I., & Al-Zoubi, H. A. (2006). Value-at-Risk under Extreme Values: the Relative Performance in MENA Emerging Stock Markets. *International Journal of Managerial, 2 (2)*, 154-173.

Mandelbrot, B. (1963). The Variation of Certain Speculative Prices. *The Journal of Business, 36*, 394-419.

Marcellino, M., Stock, J. H., & Watson, M. W. (2006). A Comparison of Direct and Iterated Multistep AR Methods for Forecasting Macroeconomic Time Series. *Journal of Econometrics, 135 (1)*, 499-526.

Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance, 7 (1)*, 77-91.

McNeil, A. J. (1997). Estimating the Tails of Loss Severity Distributions using Extreme Value Theory. *Astin Bulletin, 27 (1)*, 117-137.

McNeil, A. J., & Frey, R. (2000). Estimation of Tail-Related Risk Measures for Heteroscedastic Financial Time Series: An Extreme Value Approach. *Journal of Empirical Finance, 7 (3)*, 271-300.

McNeil, A. J., Frey, R., & Embrechts, P. (2005). *Quantitative Risk Management.* Princeton University Press.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns; A New Approach. *Econometrica: Journal of the Econometric Society*, 347-370.

Nordbo, N., Roynstrand, T., & Strat, V. (2012). Evaluating Power of Value-at-Risk Backtests . *Master Thesis: Norwegian University of Science and Technology.*

Perignon, C., & Smith, D. (2010). The Level and Quality of Value-at-Risk Disclosure by Commericial Banks. *Journal of Banking & Finance, 34 (2)*, 362-377.

Pfaff, B. (2012). *Financial Risk Modelling and Portfolio Optimization with R.* New Jersey: John Wiley & Sons, Ltd.

Pickands, J. (1975). Statistical Inference using Extreme Order Statistics. *The Annals of Statistics*, 119-131.

Poon, S. H., & Granger, C. (2005). Practical Issues in Forecasting Volatility. *Financial Analysts Journal, 61 (1)*, 45-56.

Prayag, C., Toit, D. D., Kenmuir, K., Morrison, A., & Tembo, C. (2010). Do Frontier Market Equities have a Role to Play in a Diversified International Equity Portfolio. *African Finance Journal; Special Issue*, 75-97.

Rachev, S. T., Menn, C., & Fabozzi, F. J. (2005). *Fat-Tailed and Skewed Asset Return Distributions.* New Jersey: John Wiley & Sons, Ltd.

Reiss, R., & Thomas, M. (1997). *Statistical Analysis of Extreme Values with Applications to Insurance, Finance, Hydrology and Other Fields.* Basel: Birkhäuser Verlag.

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized Additive Models for Location, Scale and Shape (with Discussion). *Applied Statistics, 54 (3)*, 507-554.

Roy, A. D. (1952). Safety First and Holding of Assets. *Econometrica, 20 (3)*, 431-449.

Shams, M. F., Sheikhi, Z., & Sheikhi, M. (2012). Application of VaR Methodology to Risk Management in the Stock Market in Iran. *International Journal of Business and Management, 7 (1)*, 70-77.

Shapiro, S. S., Wilk, M. B., & Chen, H. J. (1968). A Comparitve Study of Various Tests for Normality . *Journal of The American Statistical Association, 63*, 1343-1372.

Sharpe, W. F. (1964). Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk. *Journal of Finance, 19 (3)*, 425-442.

Sheppard, K. (2013). *Financial Econometric Notes.* UK: University of Oxford.

Simkowitz, M. A., & Beedles, W. L. (1980). Asymmetric Stable Distributed Security Returns. *Journal of the American Statistical Association, 75*, 306-312.

Simon, D. (2003). The Nasdaq Volatility Index during and After the Bubble. *Journal of Derivatives, 11*, 9-24.

Simonato, J. G. (2012). GARCH Processes with Skewed and Leptokurtic Innovations: Revisiting the Johnson SU case. *Finance Research Letters, 9*, 213-219.

Speidell, L., & Krohne, A. (2007). The Case for Frontier Equity Markets. *Journal of Investing, 16*, 12-22.

Thupayagale, P. (2010). Evaluation of GARCH-based Models in Value-at-Risk Estimation: Evidence from Emerging Equity Markets. *Investment Analysts Journal, 39 (72)*, 13-29.

Tobin, J. (1958). Liquidity Preference as Behavior Towards Risk. *The Review of Economic Studies, 25*, 65-86.

Treynor, J. (1961). Towards a Theory of Market Value of Risky Assets. *Unpublished manuscript, 6*.

Tsay, R. S. (2010). *Analysis of Financial Time Series, 3rd Edition.* Chicago: John Wiley & Sons, Inc.

Vee, D. A., Gonpot, P. N., & Sookia, N. U. (2014). Forecasting Value-at-Risk for Frontier Stock Market Indexes using GARCH-type Models and Extreme Value Theory: Model Validation for Dynamic Models. *Journal of Risk Model Validation, 8 (4)*, 47–67.

Veronesi, P. (1999). Stock Market Overreaction to Bad News in Good Times: A Rational Expectations Equilibrium Model. *Review of Financial Studies, 12 (5)*, 975-1007.

Yamai, Y., & Yoshibita, T. (2005). Value at Risk versus Expected Shortfall: A Practical Perspective. *Journal of Banking & Finance, 29*, 997-1015.

Zakoian, J. M. (1994). Threshold Heteroskedastic Models. *Journal of Economic Dynamics and Control, 18 (5)*, 931-955.

Žiković, S. (2007). Testing Popular VaR Models in EU New Member and Candidate States. *Journal of Economics and Business, 25*, 325-346.

Žiković, S., & Aktan, B. (2009). Global Financial Crisis and VaR Performance in Emergings: a case of EU Candidate States – Turkey and Croatia. *Journal of Economics and Business, 27*, 149–170.

(8) APPENDIX

8.1 Appendix 1

TABLE A.1 – MSCI Market Classifications

This table provides an overview of the MSCI (Morgan Stanley Capital International) market indices that are designed to measure the performance of the larger and mid cap segments of the respective markets, along with the appropriate market classification. The market cap weighted indexes of MSCI are among the most respected and widely used benchmarks in the financial industry. The grey cells indicate that the respective markets have been included in this study. The selection of the chosen indices has been made carefully, as these markets (i) have the highest stock market capitalization within their market classification group, (ii) have return data availability since 01.12.2005 in the financial database DataStream, and (iii) have consistent market classifications according to MSCI over the whole sample period, e.g. Argentina is only a Frontier market since May 2009 and should therefore not be included.

DEVELOPED	EMERGING	FRONTIER
Australia	Brazil	Argentina
Austria	Chile	Bahrain
Belgium	China	Bangladesh
Canada	Colombia	Croatia
Denmark	Czech Republic	Estonia
Finland	Egypt	Jordan
France	Greece	Kazakhstan
Germany	Hungary	Kenya
Hong Kong	India	Kuwait
Ireland	Indonesia	Lebanon
Israel	Korea	Lithuania
Italy	Malaysia	Mauritius
Japan	Mexico	Morocco
Netherlands	Peru	Nigeria
New Zealand	Philippines	Oman
Norway	Poland	Pakistan
Portugal	Qatar	Romania
Singapore	Russia	Serbia
Spain	South Africa	Slovenia
Sweden	Taiwan	Sri Lanka
Switzerland	Thailand	Tunisia
United Kingdom	Turkey	Vietnam
USA[1]	United Arab Emirates	WAEMU[2]

1. United States of America
2. The West African Economic and Monetary Union (WAEMU) - Senegal, Ivory Coast and Burkina Faso

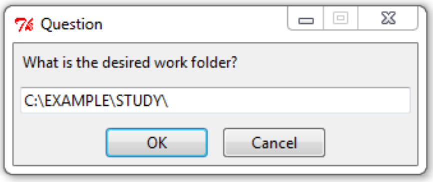
8.2 Appendix 2

SCRIPT ONE:

This script will download and install the required packages for R.

SCRIPT TWO:

First of all, this script will ask for the desired working directory.



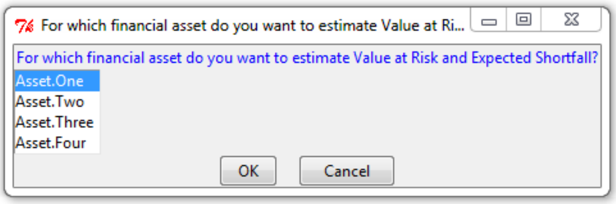
In order to perform the script, one has to supply the R software environment with only one Excel file (file type *.xlsx*). This file should contain the price data on the desired asset in the following format:

Date	*Price Asset 1*	…	*Price Asset N*
t			
…			
T			

Next, the script will automatically transform the price series to logarithmic return series, from which it will calculate a large quantity of descriptive statistics (i.e. mean, standard deviation, minimum, maximum, skewness, kurtosis, JB, LB, LM, ADF).

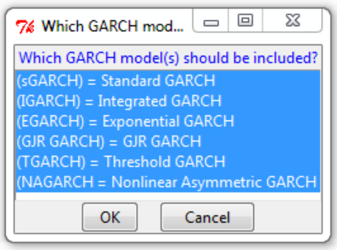
SCRIPT THREE:

Based on the Excel file provided by the user, this script will show a menu with all the available financial assets. Subsequently, the user can select the desired asset for which it wants to forecast VaR and ES.



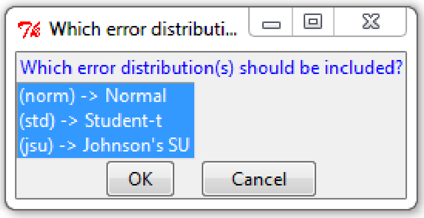
Question 1

This study implemented all of the six available risk models, but one could choose to drop one or more models from the selection if desired.



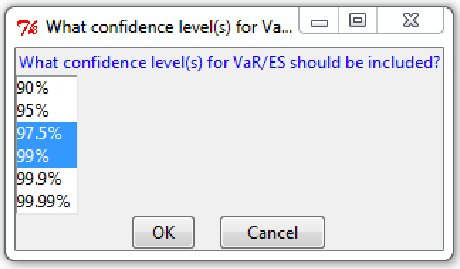
Question 2

This study implemented all of the three available distributions, but one could choose to drop one or more distributions from the selection if desired.



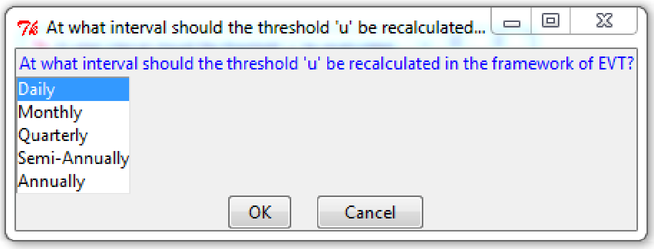
Question 3

This study implemented both the 97.5% and the 99% confidence interval. The first one is chosen to follow the future regulations of the BCBS in the framework of Expected Shortfall, while the second one is chosen to cope with the current regulations of the BCBS in the framework of Value at Risk.



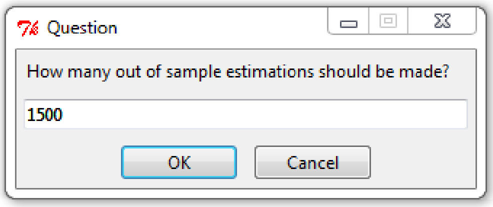
Question 4

This script incorporates the 'double bootstrap approach' of Danielsson et al. (2001), which is best described as a very computationally intensive procedure that automates the search for the appropriate threshold value. Due to the needed processing power, one could lower the estimation frequency. However, this study has chosen the daily frequency in order to make the study as realistic as possible.



Question 5

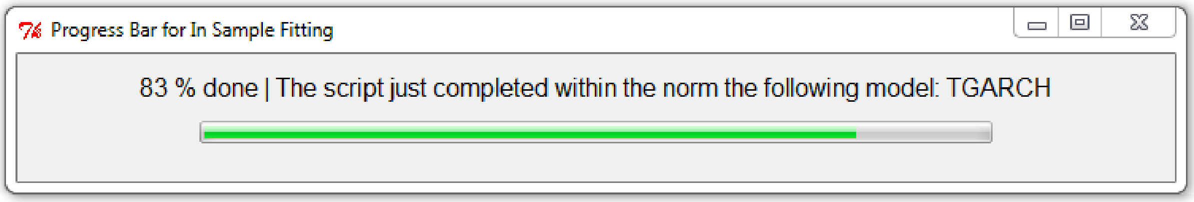
According to the current academic literature, the length of the 'out-of-sample' is subjective, but there are some guidelines (see Section 4.1). This study used 1500 observations. The remainder of the observations will then be automatically the 'in-sample', which was 1321 observations for this study.



All these answers need to be converted into variables. From this point on, the user does not need to do make any choices anymore.

SCRIPT FOUR:

The next thing in line is the 'In-Sample Fitting'-stage. In this study, we used 6 VaR models and 3 error distributions, leading to a total of 18 GARCH specifications. Each specification was estimated accordingly on the in sample data. A progress bar was added for convenience: the example on the following page shows that the TGARCH model was just estimated within the Normal Distribution. If the script crashes, one could see which model caused the inconvenience and could fix it accordingly. These type of progress bars are implemented for the whole script.



SCRIPT FIVE:

Next, the rolling estimates for Value at Risk and Expected Shortfall need to be made. This script does the calculations for the GARCH-family models, so not for the Conditional EVT. This is in the next step.

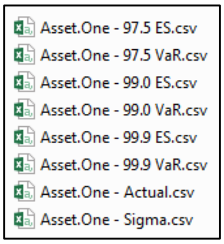
SCRIPT SIX:

The first part of the script extracts the conditional variance and the conditional mean from the fitted GARCH-models, along with the standardized residuals. The second part of the script is then devoted to the threshold calculation, with the 'double bootstrap approach' of Danielsson et al. (2001). The last part of the script then calculates the VaR and ES estimates by implementing the GPD.

SCRIPT SEVEN:

This part collects all the VaR and ES forecasts. In this study we used 6 types of GARCH models, 3 error distributions, and the Conditional EVT framework, which leads to a total of 36 models (6*3*2). This study uses the 99 percent quantile and the 97.5 quantile in order to follow the current and future BCBS regulations. The first one is for the Value at Risk framework and the second one for the Expected Shortfall framework. However, in order to backtest Expected Shortfall, one must also calculate the VaR for the same confidence level (the backtest looks at the violations of VaR). The last parameter is the length of the out-of-sample period, which is in the study set to 1500. This leads to a total of 1500 (out of sample) * 36 (models) * 3 (confidence levels) = 162.000.

The necessary objects are then all stored in .csv files. This looks as follows in the 'Output' folder.



If one wants to perform the research on multiple assets, one can repeat script three till seven for each asset. Regarding this study, this is done for the 30. This gives a total of 162.000 * 30 (markets) = 4.860.000 forecasts.

SCRIPT EIGHT:

This is the final script and it will perform the backtest analysis on all considered assets. First of all, it will implement the Conditional Coverage test of Christoffersen (1998) with bootstrap p-values and the Duration test of Christoffersen & Pelletier (2004) to filter out the statistically wrong VaR models. Then the González-Rivera et al. (2004) test will be conducted to show the ranking of the VaR models. Finally, the ES Bootstrap Test of McNeil & Frey (2000) will be conducted on the Expected Shortfall models to show the ranking. At the end, the script will provide the user with convenient tables with the p-values of the tests, the values of the loss functions, and the final rankings.