Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

Rigorozní práce

Value–at–risk forecasting with the ARMA-GARCH family of models during the recent financial crisis

Author: **Mgr. Ivo Jánský**

Supervisor: **PhDr. Milan Rippel**

Academic Year: **2011/2012**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, September 11, 2011

Signature

Acknowledgments

The author is grateful especially to PhDr. Milan Rippel for a successful guidance while writing the thesis and for valuable comments.

Abstract

The thesis evaluates several hundred one–day–ahead VaR forecasting models in the time period between the years 2004 and 2009 on data from six world stock indices — DJI, GSPC, IXIC, FTSE, GDAXI and N225. The models model mean using the AR and MA processes with up to two lags and variance with one of GARCH, EGARCH or TARCH processes with up to two lags. The models are estimated on the data from the in–sample period and their forecasting accuracy is evaluated on the out–of–sample data, which are more volatile. The main aim of the thesis is to test whether a model estimated on data with lower volatility can be used in periods with higher volatility. The evaluation is based on the conditional coverage test and is performed on each stock index separately. Unlike other works in this field of study, the thesis does not assume the log–returns to be normally distributed and does not explicitly select a particular conditional volatility process. Moreover, the thesis takes advantage of a less known conditional coverage framework for the measurement of forecasting accuracy.

JEL Classification	C51, C52, C53, C58, G01, G24
Keywords	VaR, risk analysis, financial crisis, conditional volatility, conditional coverage, stock index, garch, egarch, tarch, moving average process, autoregressive process

Author's e-mail	Ivo.Jansky@seznam.cz
Supervisor's e-mail	milanrippel@seznam.cz

Abstrakt

Tato práce vyhodnocuje několik set modelů pro jednodenní předpověď VaR v období mezi roky 2004 až 2009 na datech ze šesti světových akciových indexů — DJI, GSPC, IXIC, FTSE, GDAXI a N225. Modely jsou založené na AR a MA procesech s maximálně dvěma předešlými pozorováními a zároveň modelují podmíněnou volatilitu pomocí jednoho z GARCH, EGARCH a TARCH procesů rovněž s maximálně dvěma předešlými pozorováními. Parametry modelů jsou odhadnuty na datech z prvního období a jejich odhadovací přesnost je otestována na datech z druhého období, které vykazuje podstatně větší volatilitu. Hlavním cílem práce je otestovat, zda modely s parametry odhadnutými v období menší volatility mohou být použity i v období s větší volatilitou. Vyhodnocení je založeno na conditional coverage testu a je provedeno pro každý index zvlášť. Na rozdíl od jiných prací zabývajících se tímto tématem, tato práce nepředpokládá normální rozdělení logaritmovaných výnosů a neomezuje se na jeden předem vybraný proces pro modelování podmíněné volatility. Tato práce navíc využívá méně známý aparát, tzv. conditional coverage, pro vyhodnocení přesnosti odhadu modelů, který oproti standardním metodám nabízí několik výhod.

Klasifikace JEL	C51, C52, C53, C58, G01, G24
Klíčová slova	VaR, analýza rizika, finanční krize, podmíněná volatilita, conditional coverage, odhad modelů, akciový index, garch, egarch, tarch, moving average proces, autoregresivní proces

E-mail autora	Ivo.Jansky@seznam.cz
E-mail vedoucího práce	milanrippel@seznam.cz

Contents

List of Tables

List of Figures

Acronyms

AIC	Akaike Information Criterion
AR	Autoregressive
ARCH	Autoregressive Conditional Heteroskedasticity
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
BIC	Bayesian Information Criterion
DJI	Dow Jones Industrial Average
DTARCH	Double Threshold Autoregressive Conditional Heteroskedasticity
DTB	Deutsche Börse
EGARCH	Exponential Generalized Autoregressive Conditional Heteroskedasticity
EUREX	Eurex
EURONEXT	Euronext
ES	Expected shortfall
FTSE	FTSE 100
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
GBM	Geometric Brownian Motion
GDAXI	German DAX Index
GED	Generalized Error Distribution
GSPC	SPDR S&P 500
HMSE	Heteroskedasticity Adjusted Mean Squared Error
HMAE	Heteroskedasticity Adjusted Mean Absolute Error
HS	Historical Simulation
i.i.d.	independent and identically distributed

ISE	International Securities Exchange
IXIC	NASDAQ Composite
KPSS	Kwiatkowski–Phillips–Schmidt–Shin
LIFFE	NYSE Liffe Futures/Options
LL	Log–Likelihood
LR	Likelihood Ratio
MA	Moving Average
MC	Monte Carlo
MLE	Maximum Likelihood Estimator
MSE	Mean Squared Error
MAE	Mean Absolute Error
N225	NIKKEY 225
NE	NYSE Euronext
NYSE	New York Stock Exchange
QL	Quantile Loss
STGARCH	Tree–Structured Generalized Autoregressive Conditional Heteroskedasticity
TARCH	Threshold Autoregressive Conditional Heteroskedasticity
VaR	Value–at–Risk

Master Thesis Proposal

Author	Mgr. Ivo Jánský
Supervisor	PhDr. Milan Rippel
Proposed topic	Value–at–risk forecasting with the ARMA-GARCH family of models during the recent financial crisis

Topic characteristics While trading with any kind of assets, the market risk must be carefully estimated. There are various methods for calculation the market risk, two of them being Value–at–Risk (VaR) and the expected shortfall.

VaR is one of the most common tools for the measurement of a risk of a loss. Typically, VaR is employed by banks where it is used in the management of a credit risk and by portfolio managers, since it helps them quantify the worst loss that can occur with some given probability. There are several approaches for the calculation of VaR. such as the historical approach, variance/covariance approach, etc., each of them depending on some parameters.

The expected shortfall is an alternative method of calculating the market risk and it is particularly useful when the risk of the most extreme cases must be considered. Some authors even consider it as an excellent replacement of the VaR approach.

The goal of the thesis is to analyze these approaches separately for each asset type as well as a sample portfolio based on the market data and to reveal how much they were influenced by the recent financial crisis. The effects are to be discussed separately for the markets in the USA, the UK and Europe (DAX).

Hypotheses The following hypotheses are to be tested in the master thesis.

1. The appropriate distribution function of market values for the all asset types is the normal distribution function.
2. Based on data prior to the financial crisis, the tested risk measurement approaches for the asset types and the portfolio give the same results or at least the results do not significantly differ.
3. Based on data including the period of the financial crisis, the tested risk measurement approaches for the asset types and the portfolio give the same results or at least the results do not significantly differ.
4. The financial crisis had a significant impact on the values of the tested risk measurement approaches for the asset types and the portfolio in the selected territories.

Methodology In order to test the various risk measurement approaches, the dataset consisting of data from the New York Stock Exchange (NYSE), NYSE Liffe Futures/Options (LIFFE) and Deutsche Börse (DTB) will be used. Since all of them provide data on a daily basis, the dataset should be large enough to validate the above stated hypotheses. The employed risk measurement approaches will be the expected shortfall and the VaR calculated using *the historical approach*, *the parametric approach* and *the Monte-Carlo simulation using the Brownian motion.*

In order to obtain maximal accuracy, the true distribution function of the dataset will be first discovered using standard econometric tools. If the discovered distribution function differs from the normal distribution function, it will be included in the later on calculations.

Given the confidence intervals of 90%, 95% and 99% the risk measurement values will be calculated and then compared. Since the impact of the financial crisis is to be determined, the risk measurement will be calculated once more, this time including the data during the period of the financial crisis. The values will be compared again. The most appropriate method will be selected employing a method called back-testing.

Finally, extending the basic dataset prior to the financial crisis by data following the Brownian motion or a similar method, the hypothesis that the financial crisis significantly altered the risk measurement values will be tested.

The results should suggest the most accurate approach to be employed in the case of a future financial crisis.

Outline

1. Introduction
2. Theoretical analysis of asset types
3. Risk measurement methods
4. Empirical analysis of the Value at risk method
5. Empirical analysis of the Expected shortfall method
6. Conclusion

Core bibliography

1. Jorion, P. "Value at Risk : The New Benchmark for Managing Financial Risk." 2 edition. [New York] : *McGraw-Hill*, 2000. 544 p. ISBN 978-0071355025.
2. Saunders, A. "Credit risk measurement : new approaches to value at risk and other paradigms." *New York : John Wiley & Sons*, 2002. 319 p. ISBN 047121910X.
3. Holton, G. A. "Value at Risk : theory and practice." *Amsterdam : Academic Press*, 2003. 405 p. ISBN 0123540100.
4. Schmid, B. "Credit risk pricing models : theory and practice." *Berlin : Springer*, 2004. 383 p. ISBN 354040466X.
5. Alexander, C. "Market risk analysis : Volume III, Pricing, hedging and trading financial instruments." *Chichester : John Wiley & Sons*, 2008. 386 p. ISBN 978-0470997895.

Author

Supervisor

Chapter 1

Introduction

The modern financial world knows many types of assets and the vast majority of them can be traded on various exchanges all over the world. The exchanges — *markets* — react very quickly on any new information that enters the price making process and therefore the prices of the traded assets are constantly evolving. The market participants thus need to asses the influence of such information entering the market on the price of the asset that they own — *market risk*. There are several possible ways how to calculate or estimate market risk, one of them being Value–at–Risk (VaR). VaR is a very popular method and it is used by banks mainly for assessment of credit risk and by portfolio managers in order to evaluate the level of loss that can occur with some probability. There exist several approaches for the calculation of VaR such as the historical approach, variance/covariance approach, Monte Carlo simulation, methods based on the GARCH family of conditional volatility processes, and many others.

The objective of the thesis is to analyze VaR forecasting methods based on several conditional mean and conditional variance modeling processes in the context of the recent financial crisis and several most traded stock indices. Moreover, the methods are to be evaluated without any apriori assumptions on the particular parameters of the conditional mean and conditional variance processes as well as the shape of the distribution of the logarithmic returns. The context of the financial crisis means that the thesis will focus on the evaluation of forecasting accuracy of the models in times of increased volatility, commonly observed during the financial crisis. Accuracy in this sense is represented by ability to provide results that closely follow the actual market development.

Such definition of accuracy is especially important in times of market turbulences as it helps the market participants to better asses market risk.

The thesis is structured into four main parts. Chapter 2 provides the reader with a theoretical introduction to the problematic of forecasting VaR, the theory behind conditional mean processes, conditional variance processes, and back–testing methods. After the necessary introduction follows Chapter 3, which applies the selected VaR methods on the data from several market indices and evaluates their accuracy. Finally, Chapter 4 summarizes the obtained results.

Chapter 2

Theoretical background

To measure risk, from the financial point of view, means to try to forecast what are the possible losses that are caused by adverse movements of market prices. Market prices are influenced by many variables such as interest rates and foreign exchange rates. There are several methods, how risk can be measured and translated into real financial figures that are required by the risk managers in companies and banks in order to take appropriate investment decisions. Probably the most popular method is the so–called Value–at–Risk (VaR), which is the main risk measure that this thesis is concerned with. Several VaR forecasting methods are introduced later on in the text together with methods that are used to evaluate their accuracy — *back–testing methods*. The definition of risk, however, has to be tackled first.

2.1 Risk

The main topic of the thesis is the measurement of financial risk on the market and one of the first terms that is to be explained is *risk*. Even though there is no exact definition of what risk actually is, one can find several explanations of this word, such as the following one: "*Risk can be defined as the volatility of unexpected outcomes, which can represent the value of assets, equity, or earnings*" (Jorion 2007, pp. 3). This definition is, however, very broad and for the purpose of this thesis it should be narrowed. Since the thesis is concerned with financial markets, the main category of risk is *financial risk*.

Financial risk comprises all the typical sources of changes in the underlying variables that affect the price of an asset. The most known sources are, for example, interest rate movements that influence a vast majority of assets, and foreign exchange rate changes, which are of particular interest while trading with currency futures. In other words, financial risk represents the possibility of a loss stemming from participation on financial markets.

The importance of financial risks evaluation is easily illustrated on the development of the financial markets during the past 40 years. In the period between the year 1970 and the present day, there have been many significant events that caused the markets to drop in value. Several examples include oil–price shocks, Japanese stock–bubble, the Asian turmoil, the Russian default, the terrorist attacks in 2001, and of course, the most recent financial crisis. Each of these events was unexpected therefore it was not possible to avoid it, when it actually occurred.

2.1.1 Risk components

The sources of risk are also diverse. Most of risk sources are connected with human activities that include running the economy such as inflation, interest rates, innovations, political decisions, wars, and so on. Another source of risk comes from the nature itself. There have always been bad weather, floods, earthquakes, bad crops, etc. Therefore, it is easy to see that the world we live in is full of risk and financial markets are no exception. Since the term *risk* is quite broad, the following groups of risks are typically considered — market risk, liquidity risk, credit risk, and operational risk.

Market risk

Market risk is the main financial risk category that the thesis is concerned with. It represents the losses that result from adverse price movements of assets on the market. As it has already been said, market risk is caused by the movements of interest rates, exchange rates, etc. To get a better grasp at such virtual measure, the market risk value can be looked at from several points of view. Many authors, such as for example Alexander (2009), distinguishes the following:

- absolute risk — the absolute difference between the two prices over time,
- relative risk — the difference of the price of an asset and some arbitrary benchmark index,
- directional risk — represents risk related to the movements of financial variables such as interest rates, foreign exchange rates, etc.,

- non–directional risk — other risk variables not included in directional risk, usually exposures to hedged positions,
- basis risk — represents unexpected "*movements of the relative prices of an asset in a hedged position*",
- volatility risk — stems from changes of actual volatility.

Market risk is usually being closely monitored and it can be significantly reduced by setting *stop–loss limits*, employing the VaR method, or even by supervising market activities by an independent supervisor.

Liquidity risk

The second type of risk, included in the financial risks category, is liquidity risk. This is a specific type of risk as it is not caused by movements of financial variables. The main sources of liquidity risk are the market itself – *the depth of the market* — and its participants. As well as in the previous discussion of market risk, Jorion (2007) distinguishes two types of liquidity risk:

- market and product liquidity risk,
- cash–flow risk.

Market and product liquidity risk, sometimes called *asset–liquidity risk*, is the case when a market transaction cannot be successfully closed due to some constraints on the size or the price of the transaction. Transaction is considered as either too big or too small for the market, when there is no other entity that has a desire to engage on the other side of the transaction, which usually happens when the transaction size differs from a typical trading lot. This, however, does not seem to be a problem with major currencies or with treasury bonds, as their markets are very deep. Exotic derivatives, on the other hand, face higher liquidity risk.

The second subgroup, cash–flow risk, which is also known as *funding–liquidity risk*, represents actual inability to provide funds. Cash–flow risk is particularly obvious in leveraged portfolios. Typically, this situation leads to increased losses due to the fact that traders need to make marking–to–market payments and if they fail to do so, they are forced to sell the portfolio, which in turn translates into even higher losses.

Asset–liquidity risk can be effectively minimized by trading standardized amounts of assets on sufficiently liquid markets, however, it might not always be possible to avoid. Cash–flow risk is a bit easier to manage. The easiest way is to engage in proper cash–flow management in combination with portfolio hedging, diversification and setting limits on maximum possible cash–flow differences.

Credit risk

The third major source of risk is credit risk. Credit risk represents losses that stem from inability or unwillingness of a counter–party to meet its liabilities. From the costs point of view, credit risk can be further on divided into two components. The most obvious component is the nominal value that is lost due to credit risk and the second component is the so called *recovery rate*. The recovery rate states, how much from one lost dollar gets returned back to the lender. Naturally, the recovery rate is in most situations smaller than one, however, in rare situations it can even be slightly higher than one thanks to additional fees related to the workout process. As Holton (2003) points out, credit risk can be caused by the process of marking–to–market of debt, which is often the result of changes in market prices of debt caused by a changed credit rating of the debt.

Two more types of credit risk also belong to the same category. On the contrary to company–specific risk as introduced in the previous paragraph, there is also country–specific risk, *sovereign risk*, which is mainly caused by political bodies that attempt to control the foreign exchange market. The result of such control is inability to meet ones liabilities and therefore increased credit risk.

Last risk belonging to the credit risk group is settlement risk. Settlement risk occurs, as its name suggests, at the time of settling the transaction. Especially when different currencies are being traded and the physical settlement occurs at two different times, settlement risk increases and the trader's exposure equals the full value of the transaction. Again there is a possibility to manage credit risk. The minimization of credit risk is usually performed by setting various limitations such as marking to market.

Operational risk

Finally, the fourth major source of risk is represented by operational risk. This group comprises of several parts that do not fit in any of the previous categories, mainly due to the fact that they arise from internal processes in companies and banks. Typically, one can clearly identify the following sources of operational risk:

- process risk,
- people risk,
- model risk,
- legal risk.

Process risk, as the name suggests, is present in every process connected with the participation on financial markets. A typical example can be found in the banking sector. The process of making a transaction on a financial market, from the perspective of a bank, is a process instrumented by several actors. First, there is a trader who enters the market and creates a deal. Information about the deal is then sent to the back–office where it is entered into banking information systems. It is clear that during the process there can arise a problem due to false information provided and/or received and due to incorrect information entered into banking systems.

People risk is closely related to process risk, as people are parts of processes, but the definition is different. People risk actually represents a possibility that some person intentionally provides false information or, in other words, frauds the other party. This type of risk is often connected with market risk, simply because traders might be tempted to falsely identify their position after incurring a loss of some significant amount. An example could be the fall of the Barings bank in England in 1995.

Third risk – *model risk* — is again partially correlated with process risk, but it is considered as a specific type of operational risk. Model risk is represented by inaccurately valuated models that result in wrong decisions about market position and later on incur a loss. Practically any used model can be inaccurately valuated. The option pricing model is one example. This risk can be largely minimized by "*independent evaluation using market prices, when available, or objective out–of–sample evaluations*" (Jorion 2007, pp. 26).

The fourth mentioned source of operational risk is legal risk. Legal risk stems from the legal environment, in which the entity lives and in which it operates. In the modern largely globalized and integrated world, legal environment is very complex as it can include legal environments of many parties that participate on a particular transaction. Therefore risk of doing something that is against the law or regulation can easily occur. Results of legal risk are "*fines, penalties, or punitive damages resulting from supervisory actions, as well as private settlements*" (Jorion 2007, pp. 26).

Operational risk can be largely minimized by engaging in several activities such as minimizing the number of systems involved, preparing effective internal controlling, and, probably the most important step, requiring a clear separation of responsibilities. Operational risk and its minimization has been widely discussed in the recent years, especially in the banking industry. The BASEL II regulatory framework, for example, clearly identifies operational risk as one of the risks that banks face and it requires them to be prepared for possible losses that might occur due to operational risk.

2.1.2 Value at risk

VaR is a widely used method for the assessment of financial risk. The history of the method dates back to the beginning of the 1990s, when the method was first introduced in response to the recent financial turmoils. At the time VaR was published it was primarily intended to be used as measure of market risk. However, in the past two decades VaR has evolved into a universal risk management tool, which is now used to calculate both credit and operational risk. This is, however, not the complete enumeration. Many authors, such as Alexander (2009), provide a list of VaR applications as they evolved over time. It is interesting to observe how a passive tool became an active daily–used method in almost any financial institution in the world.

- *Passive* — At this time, VaR was used mainly for information reporting. Financial operations were assessed by this method and results were presented to the management and shareholders.

- *Defensive* — Later on, as the potential of VaR was better understood, the method began to represent virtual boundaries for traders, while trading on the markets. Position limits calculated by VaR were introduced.

- *Active* — Finally, VaR became an active tool in the risk management industry. Its primary use now is to manage risk and it serves the purposed of allocating capital inside a company or a bank. Another example from this category is the employment of risk adjusted performance measures that serve the purpose of better assessing risks connected with an investment.

The evolution of VaR clearly shows how a simple statistical method can become one of the mostly used tools in the financial industry. The clear benefits are mainly for shareholders in general, because the management can see the risk impact of some desired investment on the portfolio as a whole, before even carrying out the investment. That is the reason, why VaR is being widely used all over the world. Institutions that use VaR range from financial institutions to non-financial institutions, as it is summarized in the following list:

- *Financial institutions* — Financial institutions and especially banks are very sensitive to any negative changes in their portfolio, because they usually manage very large portfolios consisting of many types of assets. That is the reason why there is a strong tendency among these institutions to create centralized rules for risk management.
- *Regulators* — Probably the largest growth of the VaR methodology employment can be seen by regulatory offices around the world. As examples can very well server the Basel Committee on Banking Supervision or securities commissions in the USA and in the European Union. BASEL II requires banks to hold a certain amount of reserves for adverse situations and it also employs VaR in order to calculate credit and operational risk.
- *Non–financial institutions* — It is very common that ordinary companies hold assets in the form of foreign currencies, because they operate on several foreign markets. These companies are also subjects to financial risk that can be evaluated by the VaR method.
- *Asset managers* — Finally, asset managers heavily use VaR, because it gives them the possibility to see a complex picture of risk that they are undertaking, while trading with various assets. As it is described in Holton (2003), there are even possibilities to show asset managers the breakdown of total risk by markets, asset types and so on.

Holton (2003) arguments that transparent risk reporting procedures could have prevented financial losses that occurred in the past. The application of the VaR method has the advantage that it provides information about possible losses that could occur as a result of adverse movements of market variables. Another benefit of employing VaR lies in the fact that it changes the way people think about risk. This results in a clear risk management governance in institutions and forces the institutions to consider risk more carefully than before.

Since the risk of a loss actually represents a dichotomy — there is exposure to risk factors and then there are distributions of risky factors — there exist several methods for VaR calculation. First, models that work with the exposure can be divided into two groups — *local valuation* and *full valuation* methods. The main difference between these two groups is in the frequency with which they valuate the portfolio. The local valuation performs the valuation only once and it is represented by the *delta–normal* method and the *variance–covariance* approach. The second group, the full valuation, performs valuation of the portfolio over a range of possible scenarios. The models that focus on the distribution of risky factors are represented by both *parametric* approaches — such as the normal distribution — and *non–parametric* approaches that work with historical data.

2.2 Conditional mean

While working with time series data, it is quite common that data is timely dependent. In other words, the value of the observed variable at time t might be dependent upon some information that has been observed earlier. Quite common are the lagged values of the observed variable and the lagged values of the error term. Thus the two concepts of conditional mean covered by the thesis are the Autoregressive (AR) process and the Moving Average (MA) process.

2.2.1 Autoregressive process

An AR process is described by the Equation 2.1 below. The order of the AR process is given by the value of the lag parameter p. From the equation it is clear that the value of the random variable y_t at time t depends on the previously realized values of y_t. The other components in the equation are constant mean μ and white noise ϵ_t. The importance of the previously realized value of the random variable y_{t-i} is captured by the parameter ρ_i, where i is the i–th lag of the random variable. For simplicity, the constant μ is usually omitted.

$$y_t = \mu + \sum_{i=1}^{p} \rho_i y_{t-i} + \epsilon_t \tag{2.1}$$

A typical constraint on a time series random variable is that the random variable is stationary. The stationarity condition for the AR process is usually described by restrictions on roots of the polynomial $z^p - \sum_{i=1}^{p} \rho_i z^{p-i}$. In order for the AR process to be wide–sense stationary, roots of the polynomial must lie within a unit circle. In other words, for each z_i it must hold that $|z_i| < 1$. Since the task of checking for the stationarity is quite common, there exist several statistical tests, which test for stationarity and can be applied on the time series.

One of the well known tests is the Dickey–Fuller test introduced by Dickey & Fuller (1979), which tests the null–hypothesis that a unit root is present in a time series. This test is available in most statistical software. A derivation of the Dickey–Fuller test for larger datasets is the Augmented Dickey–Fuller test introduced in Said & Dickey (1984). Based on the Dickey–Fuller test is also the Phillips–Perron test. According to Phillips & Perron (1988), the test is able to deal with autocorrelation and heteroskedasticity in the time series in a more robust way. A complement to the unit–root tests described so far is the Kwiatkowski–Phillips–Schmidt–Shin (KPSS) test introduced by Kwiatkowski *et al.* (1992). The null–hypothesis of the KPSS test is a stationarity of the time series and it is typically used together with the three above mentioned tests.

2.2.2 Moving–average process

Another process that is commonly used with univariate time series is the MA process. The mathematical representation of the MA process is provided in the Equation 2.2 below. The order of the MA process is described by the parameter q and it represents the number of lagged errors, on which the current value of the random variable y_t depends. The idea behind the MA process lies in the inclusion of past unexpected shocks or errors ϵ_{t-i} as variables that help to explain the value of the random variable at specific time. Errors are assumed to be white noise and the importance of the lagged error ϵ_{t-i} is captured by the θ_i parameter, the mean is represented by the variable μ and the error term for the current period t is represented by ϵ_t.

$$y_t = \mu + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t \tag{2.2}$$

2.2.3 Autoregressive moving–average process

The two described processes for modeling the mean of a time series are often combined in order to create an Autoregressive Moving Average (ARMA) process. An ARMA process is able to capture both the autoregressive part of the time series and lagged shocks introduced by the error term. Therefore the ARMA process has two parameters, p and q, which specify orders of the AR process and the MA process respectively. The equation of the ARMA process is provided below in Equation 2.3.

$$y_t = \mu + \sum_{i=1}^{p} \rho_i y_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t \tag{2.3}$$

An ARMA process is a special case of the Autoregressive Integrated Moving Average (ARIMA) process. In addition to the p and q parameters, the ARIMA process has another parameter d that specifies the degree by which the random variable is differenced. Differencing a random variable basically means calculating the differences between two adjacent realizations of the random variable. Even thought, the time series random variable is typically differenced once to become stationary, there might be cases when the first difference is not enough or the model requires a higher order differencing. In case no differencing is necessary, the parameter d is equal to zero. The ARMA process is thus an ARIMA process with the d parameter equal to zero.

Processes that model the mean of the time series variable are not the only dynamic part in the procedure of identifying an appropriate model for a time series. Many authors, such as Ghahramani & Thavaneswaran (2008) argue that it is quite common with time series random variables that the squared error term is not homoskedastic. In order to improve the predicting capability of the one–day–ahead VaR, conditional variance processes are often used. The conditional process used in the thesis are described in the following section.

2.3 Conditional variance

The calculation methods used for one–day–ahead VaR forecasting range from very simple ones to more advanced models. The first group is represented by the Historical Simulation (HS) method, which is the easiest method to calculate. Then there is the Monte Carlo (MC) method that is a part of the parametric approach family of models. Some authors such as Huang (2010) employ an advanced modification of the MC method called the optimized MC method, which is supposed to provide more accurate results. Since the MC method is based on a simulation of different outcomes, there is usually a substantial increase in the complexity of the calculation in comparison to the HS method — especially the duration of the calculation. On the other side of the VaR forecasting spectrum lie models based on conditional volatility processes such as the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) process with all its special cases such as the Exponential Generalized Autoregressive Conditional Heteroskedasticity (EGARCH) and Threshold Autoregressive Conditional Heteroskedasticity (TARCH). These three mentioned processes will be the main conditional volatility processes used in the thesis.

2.3.1 GARCH

Econometric models dealing with time series usually work under the assumption of homoskedasticity, in other words, constant volatility. However, in the case of financial time series, volatility is rarely a constant, as many authors such as Akgiray (1989) proved. Therefore, volatility can vary at different points in time. One of the models that is able to deal with conditional heteroskedasticity is the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model introduced by Bollerslev (1986), which is a generalization of the Autoregressive Conditional Heteroskedasticity (ARCH) model that was originally developed by Engle (1982).

The ARCH family of models correctly distinguishes between unconditional variance and conditional variance and allows conditional variance to change over time, leaving unconditional variance constant. The ARCH model allows for long lags in conditional variance and the GARCH model extends it in the way that it allows for both long lags in conditional variance and a more flexible lag structure.

Let's start with the definition of the GARCH model, which is described by the equations 2.4 and 2.6. The Equation 2.4 is a general equation for the value of the error term denoted as ε_t. Under the assumption that the distribution of the random variable z_t is normally distributed with mean zero, the Equation 2.4 can be replaced by the Equation 2.5. More details about possible distributions of the random variable z_t are provided later on in this section.

$$\varepsilon_t = z_t\sqrt{h_t} \tag{2.4}$$

$$\varepsilon_t|\mathcal{F}_{t-1} \sim N(0, h_t) \tag{2.5}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q}\alpha_i\varepsilon_{t-i}^2 + \sum_{i=1}^{p}\beta_i h_{t-i} \tag{2.6}$$

The Equation 2.5 says that the distribution of the random variable ε_t is normal with mean zero and variance h_t conditional of $\mathcal{F}_{t-1}$, which is an information set of all information up to time $t-1$. The Equation 2.6 then models conditional variance h_t based on both past realizations of h_t and the square of the past realization of the random variable ε_t. By changing the parameters p and q, the GARCH model can be transformed into other models, some of which are well known.

To show how the GARCH model relates to the ARCH model, let's consider a special case when $p = 0$. The sum of lags of h_t then drops from the equation of conditional variance h_t and the equation is identical to the equation of conditional variance in the ARCH model. Therefore, ARCH model is only a special case of the GARCH model. Another special case takes place when $p = q = 0$. In this case, the GARCH model simply becomes white noise ε_t. It is therefore obvious that parameters p and q are crucial to the model, along with estimates of parameters α and β. There are, however, several restrictions on values of parameters α and β.

$$\begin{aligned} p &\geq 0, & q &> 0 & & \\ \alpha_0 &> 0, & \alpha_i &\geq 0, & i &= 1, \cdots, q \\ \beta_i &\geq 0, & & & i &= 1, \cdots, p \end{aligned} \tag{2.7}$$

The interpretation of above written conditions is straightforward. The value of α_0 must be greater than zero, so that conditional variance is also greater than zero. Values of α_i and β_i must as well be greater than zero, in order to achieve overall positive conditional variance. However, α_i and β_i are also allowed to be equal to zero, which is typically the case when parameters are not significant at some i. The last condition, as written in the Equation 2.8, assures that the random process described by ε_t is covariance stationary. The Equation 2.9

simply represents the value of unconditional variance of the GARCH model.

$$\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1 \tag{2.8}$$

$$h_t = \frac{\alpha_0}{\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i} \tag{2.9}$$

A useful summary of above stated and described conditions and equations is represented by the Theorem 2.1 introduced below. The proof of the theorem can be found in Bollerslev (1986).

Theorem 2.1. *The GARCH(p,q) process as defined in 2.5 and 2.6 is wide–sense stationary with* $E(\varepsilon_t) = 0$, $var(\varepsilon_t) = \frac{\alpha_0}{\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i}$ *and* $cov(\varepsilon_t, \varepsilon_s) = 0$ *for* $t \neq s$ *if and only if* $\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i < 1$.

Proof. See Bollerslev (1986, Appendix) □

Distributions of z_t

In his original work, Engle (1982) assumed that the distribution of the random variable z_t in the Equation 2.4 is normal. Therefore, this equation could be simply replaced by the Expression 2.5. Normal distribution, however, is not the only possible distribution of the random variable. Bollerslev (1986), for example, suggested to use the standardized Student–t distribution with $\upsilon > 2$ degrees of freedom. The density function of the Student–t distribution is in the form described by the Equation 2.10.

$$D(z_t, \upsilon) = \frac{\Gamma\big((\upsilon+1)/2\big)}{\Gamma(\upsilon/2)\sqrt{\pi(\upsilon-2)}} \left(1 + \frac{z_t^2}{\upsilon-2}\right)^{-\frac{\upsilon+1}{2}}, \tag{2.10}$$

where $\Gamma(\upsilon) = \int_0^\infty e^{-x} x^{\upsilon-1} dx$ is the gamma function and υ represents the so–called tail–thickness parameter of the distribution.

The motivation for the use of the Student–t distribution lies in thick tails of financial time series.[1] The mean of the standardized Student–t distribution is equal to zero and degrees of freedom, controlled by the parameter υ, influence the tails–thickness. As the value of υ goes to infinity, the shape of the distribution resembles the shape of the normal distribution.

Another popular distribution is the Generalized Error Distribution (GED),

[1] The presence of thick tails in financial time series has been reported by many authors such as Bollerslev *et al.* (1994); Ghysels *et al.* (1996).

which has been proposed by Nelson (1991). He suggested that the GED distribution should be used in the EGARCH model. The density function of the GED with zero mean is described by the Equation 2.11.

$$D(z_t, \upsilon) = \frac{\upsilon \exp(-0.5|z_t/\lambda|^{\upsilon})}{2^{(1+1/\upsilon)}\Gamma(\upsilon^{-1})\lambda}, \tag{2.11}$$

where $-\infty < z < \infty$, $0 < \upsilon \leq \infty$, $\Gamma(\cdot)$ is the gamma function and

$$\lambda \equiv \left[2^{(-2/\upsilon)}\Gamma(1/\upsilon)/\Gamma(3/\upsilon)\right]^{1/2}. \tag{2.12}$$

The parameter υ is the tail–thickness parameter and its value effectively changes the shape of the distribution function. For example, when $\upsilon = 2$ then the random variable z is distributed as $z \sim N(0, 1)$. Thicker tails can be achieved by setting $\upsilon < 2$ and thinner tails vice versa by setting $\upsilon > 2$. In case $\upsilon = \infty$, z_t has a uniform distribution on the interval $\langle -\sqrt{3}, \sqrt{3}\rangle$. Even though, the three briefly introduced distributions are not the only distributions that can be used in the GARCH family of models, they suffice for the scope of the thesis. More information about other distributions and their use in the GARCH family of models can be found in works of Guermat & Harris (2002) and Lambert & Laurent (2001).

Parameter estimation

One possible procedure, how to obtain parameters of the GARCH model, is to use the Maximum Likelihood Estimator (MLE), which provides an asymptotically efficient estimator.[2] Under the assumption that innovations z_t that enter the MLE are independent and identically distributed (i.i.d.) and the density function of innovations is $D(z_t, \upsilon)$, the log–likelihood function of $\{y_t(\theta)\}$ is described by the Equation 2.13.

$$\ell(\{y_t\}, \theta) = \sum_{t=1}^{T} \left[\ln\Big[D\big(z_t(\theta), \upsilon\big)\Big] - \frac{1}{2}\ln\big(h_t(\theta)\big)\right], \tag{2.13}$$

where T is the number of observations and θ is the vector of parameters that are to be estimated. The likelihood estimator that maximizes the Equation 2.13 is denoted as $\hat{\theta}$. Angelidis *et al.* (2004) in their work provide equations for the three discussed distribution functions. Therefore, the derivation

[2] See Wooldridge (2008) for introduction to the MLE estimation.

of these equation will not be done step–by–step in this thesis, instead, equations from the referred source will be used.

Log–likelihood for z_t that is normally distributed:

$$\ell(\{y_t\}, \theta) = -\frac{1}{2}\left[T \ln(2\pi) + \sum_{t=1}^{T} z_t^2 + \sum_{t=1}^{T} \ln(h_t)\right]. \tag{2.14}$$

Log–likelihood for z_t with Student–t distribution:

$$\begin{aligned}\ell(\{y_t\}, \theta) &= T\left[\ln\Gamma\left(\frac{\upsilon+1}{2}\right) - \ln\Gamma\left(\frac{\upsilon}{2}\right) - \frac{1}{2}\ln\left[\pi(\upsilon-2)\right]\right] - \\ &\quad \frac{1}{2}\sum_{t=1}^{T}\left[\ln(h_t) + (1+\upsilon)\ln\left(1 + \frac{z_t^2}{\upsilon-2}\right)\right]. \end{aligned}\tag{2.15}$$

Log–likelihood for z_t with GED distribution:

$$\begin{aligned}\ell(\{y_t\}, \theta) &= \sum_{t=1}^{T}\left[\ln\left(\frac{\upsilon}{\lambda}\right) - \frac{1}{2}\left|\frac{z_t}{\lambda}\right|^{\upsilon} - \right. \\ &\quad \left. (1+\upsilon)^{-1}\ln(2) - \ln\Gamma\left(\frac{1}{\upsilon}\right) - \frac{1}{2}\ln(h_t)\right]. \end{aligned}\tag{2.16}$$

The vector of estimated parameters $\hat{\theta}$ of the particular distribution function of the random variable z_t is then obtained by employing a standard maximization method, in which Θ is the parameter space, as captured in the Equation 2.17.

$$\hat{\theta} = \arg\max_{\theta\in\Theta} \ \ell(\{y_t\}, \theta) \tag{2.17}$$

Value–at–risk forecast

Once parameters of the random variable z_t with the desired distribution are known, the calculation of the VaR forecast for time $t+1$ conditional on the information up to time t, is straightforward. The general formula for one–day–ahead VaR is in 2.18, where $F(\alpha)$ is the α–quantile of the distribution of innovations and $\hat{h}_{t+1|t}^{1/2}$ is the conditional standard deviation of the innovation process at time $t+1$, based on the information set up to time t, obtained from the Equation 2.19.

$$VaR_{t+1|t} = F(\alpha)\hat{h}_{t+1|t}^{1/2}, \tag{2.18}$$

$$\hat{h}_{t+1|t} = \alpha_0^{(t)} + \sum_{i=1}^{q} \alpha_i^{(t)} \varepsilon_{t-i+1}^2 + \sum_{i=1}^{p} \beta_i^{(t)} h_{t-i+1} \tag{2.19}$$

Asymmetric models

The advantage of using the GARCH model lies in the fact that the model is able to deal with some characteristics of typical time series. These characteristics include, for example, thick tails and volatility clustering, as pointed out by Mandelbrot (1963) and Mandelbrot (1967). There are, however, some characteristics of financial time series that the GARCH model is not able to deal with. The main disadvantage of the GARCH model is that conditional variance depends on the squared value of ε_t, which in turn means that the model is sensitive only to the absolute magnitude of the variable but not to its sign. This represents a complication as Angelidis *et al.* (2004) point out, since a so called leverage effect introduced by Black (1976) might be present.

The leverage effect represents a negative correlation between asset returns and volatility of returns. In other words, volatility of returns has a tendency to rise as a result of negative news on the market and decrease as a result of positive news on the market, ie. when $\varepsilon_t < 0$ resp. $\varepsilon_t > 0$. Some authors such as Brooks & Persand (2003) and Rabemananjara & Zakoïan (1993) therefore suggest that models that take into consideration asymmetries in volatility of returns should be preferred. Failing to do so might, according to these authors, result in inaccurate forecasts of VaR. The appropriate solution to handle such asymmetry lies in using appropriate extensions of ARCH respectively GARCH models. Two most popular asymmetric models are the EGARCH model of Nelson (1991) and the TARCH model created by Zakoian (1994).

2.3.2 EGARCH

The introduction of the EGARCH model by Nelson (1991) reacted on the criticism of the popular GARCH model. According to Teräsvirta (2006), Nelson saw a problem especially in restrictions on parameters of the GARCH model, as represented by expressions 2.7 in the preceding section. The source of the criticism was induced by the requirement on positivity of conditional variance at all times. Second problem was the actual ignorance of possible asymmetries in volatility by the GARCH model.

$$\ln h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \big(\theta z_{t-i} + \gamma_i(|z_{t-i}| - E|z_{t-i}|)\big) + \sum_{i=1}^{p} \beta_i \ln h_{t-i} \tag{2.20}$$

The EGARCH model is represented by the Equation 2.20. At the first sight, it is clear that the parameter γ_i is the asymmetric effect parameter and its value is a direct determinant of the behavior of the model. In the case when $\gamma_i = 0$, both positive and negative shocks are treated the same way. In general it holds that the component θz_{t-i} determines the sign of the change and the component $\gamma_i(|z_{t-i}| - E|z_{t-i}|)$, on the other hand, the magnitude of the change.

Considering the fact that the Equation 2.20 is a logarithmic transformation, restrictions on values of α_i and β_i parameters, as described by expressions in 2.7, can be levied. Parameters α_i and β_i can therefore attain both positive and negative values. The logarithmic transformation also ensures that the actual value of conditional volatility is always non–negative. Stationarity conditions for the EGARCH model can be found in the original work of Nelson (1991).

Value–at–risk forecast

The parameter estimation of the random variable z_t can be obtained in the same way as in the case of the GARCH model. The calculation of the VaR forecast at time $t+1$, conditional on the information up to time t, is again based on the one–day–ahead VaR formula described by the Equation 2.18. The calculation of $\hat{h}_{t+1|t}$ differs from the GARCH case though. The conditional standard deviation of the innovation process in the EGARCH model at time $t+1$ based on the information set up to time t, is obtained from the Equation 2.21.

$$VaR_{t+1|t} = F(\alpha)\hat{h}_{t+1|t}^{1/2} \quad ,$$

where

$$\begin{aligned} \ln \hat{h}_{t+1|t} \quad &= \quad \alpha_0^{(t)} + \sum_{i=1}^{q} \alpha_i^{(t)} \big(\theta z_{t-i+1} + \gamma_i^{(t)}(|z_{t-i+1}| - E|z_{t-i+1}|)\big) + \\ & \sum_{i=1}^{p} \beta_i^{(t)} \ln h_{t-i+1}. \end{aligned} \tag{2.21}$$

Model discussion

In their work, He *et al.* (2002) suggest that for the popular EGARCH(1,1) model, the decay of autocorrelations of squared residuals is faster at the beginning and it slows down until it begins to decay at an approximately exponential rate. Therefore, they do not recommend to use the EGARCH(1,1) model for such time series where autocorrelations decay slowly. Moreover, Malmsten & Teräsvirta (2004) reported that EGARCH(1,1) model with the assumption of normally distributed errors is not appropriate for time series with slowly decaying autocorrelations and high kurtosis. These findings are in line with the recommendation of Nelson (1991) who suggested that errors should be modeled using the GED. Nelson (1991) in his work considered also the Student–t distribution with finite degrees of freedom, however, he found out that when this distribution is used, infinite unconditional variance might occur. Therefore, the distribution function of errors must be chosen carefully.

Engle & Ng (1993) provide criticism of the EGARCH model that is based on empirical studies. They report that according to their results, the EGARCH model has a tendency to overweight the effect of shocks with greater magnitude on the volatility. As a result the EGARCH model should provide poorer results than the GARCH model. Since the topic of this thesis is to investigate several VaR models in times of increased volatility, own opinion on the criticism will be provided later on.

2.3.3 TARCH

Another model from the family of asymmetric ARCH models is the TARCH model. This model has been introduced by Zakoian (1994) and the idea behind it is based on the work of Davidian & Carroll (1987) who found out that not squared residuals, but absolute residuals provide better variance estimates. It is important to note that Zakoian (1994) estimates conditional standard deviation instead of conditional variance as it is the case with GARCH and EGARCH models. The idea behind the model is to divided innovations ε_t into two intervals that are disjunct and then to estimate parameters in the linear function of conditional standard deviation. The TARCH model is thus described by the Equation 2.22.

$$h_t^{1/2} = \alpha_0 + \sum_{i=1}^{q} (\alpha_i^+ \varepsilon_{t-i}^+ - \alpha_i^- \varepsilon_{t-i}^-) + \sum_{i=1}^{p} \beta_i h_{t-i}^{1/2} \quad , \tag{2.22}$$

where $\varepsilon_t^+ = \max(\varepsilon_t, 0)$ and $\varepsilon_t^- = \min(\varepsilon_t, 0)$.

Even though, it would be possible to leave out restrictions on values of parameters α_i and β_i, since the modeled conditional standard deviation can attain negative values, Zakoian (1994) restricts these parameters, so they have the same properties as in the standard GARCH model.

$$\begin{aligned} &\alpha_0 > 0, \quad \alpha_i^+ \geq 0, \quad \alpha_i^- \geq 0, \quad i = 1, \cdots, q \\ &\beta_i \geq 0, \quad\quad\quad\quad\quad\quad\quad\quad\quad i = 1, \cdots, p \end{aligned} \tag{2.23}$$

From the Equation 2.22 it is obvious that parameters α_i^+ and α_i^- allow to have different values for positive and negative shocks. On the contrary to the EGARCH model, the TARCH model allows for finer control over asymmetry of squared residuals. The reason lies in the fact that various lags might result in opposite contributions to conditional volatility; for example, $\alpha_1^+ - \alpha_1^- < 0$ and $\alpha_2^+ - \alpha_2^- > 0$. Second important difference between the two models stated by Zakoian (1994) is that TARCH employs additive modeling and volatility is a function of (non–normalized) innovations. Last but not least difference is that the $\ln h_t$ process requires an ARMA process and "*it does not provide any linear equation in any function of* ε" (Zakoian 1994, pp. 935).

Value–at–risk forecast

As well as in the two previous cases, the parameter estimation of the random variable z_t is obtained by employing the MLE method. The calculation of the VaR forecast at time $t+1$ conditional on the information up to time t, is based on the one–day–ahead VaR formula provided in the Equation 2.18. The calculation of $\hat{h}_{t+1|t}$, as in the case of the EGARCH model, differs from the GARCH case. The conditional standard deviation of the innovation process in the TARCH model at time $t+1$ based on the information set up to time t, is obtained from the Equation 2.24.

$$VaR_{t+1|t} = F(\alpha)\hat{h}_{t+1|t}^{1/2} \quad ,$$

where

$$\hat{h}^{1/2}_{t+1|t} = \alpha_0^{(t)} + \sum_{i=1}^{q}\Big([\alpha_i^{+}\varepsilon^{+}_{t-i+1}]^{(t)} - [\alpha_i^{-}\varepsilon^{-}_{t-i+1}]^{(t)}\Big) + \sum_{i=1}^{p}\beta_i^{(t)} h^{1/2}_{t-i+1}. \quad (2.24)$$

Model discussion

There exist several modifications of the TARCH model, such as the Double Threshold Autoregressive Conditional Heteroskedasticity (DTARCH) model developed by Li & Li (1996), which is a non–linear threshold model. The TARCH model is, on the contrary, linear in parameters, since it assumes that the value of the threshold parameter is equal to zero. Another example of a nonlinear TARCH model is the Tree–Structured Generalized Autoregressive Conditional Heteroskedasticity (STGARCH) model by Audrino & Buhlmann (2001). This thesis will, however, work solely with the two most popular asymmetric models, namely the EGARCH model and the TARCH model described in this section.

2.4 Back testing

Since the topic of the thesis is to investigate which VaR model is more appropriate in times of increased volatility, evaluation methods shall be discussed. The choice of a particular back–testing method is important decision and one particular problem while evaluating the various VaR models is caused by the fact that it is not possible to directly observe VaR. The back–testing methods should be based on several criteria. Firstly, it should be the number of cases when the actual market loss was greater than the forecasted VaR by a particular model — from now on the number of *violations*. Moreover, methods should also check that violations are i.i.d. i.e. that they are not correlated.

In the literature on VaR methods, several back–testing methods appear very often and can therefore be considered as standard. One of them is the (un)conditional coverage, as introduced by Christoffersen (1998). This method is able to catch both above mentioned types of forecasting errors, therefore, it will be employed in the thesis. Another approach to back–test VaR forecasts that is sometimes used in the literature is the Engle & Manganelli (1999; 2004) model, which tests for correlated VaR violations. This thesis, however, works solely with the (un)conditional coverage. In the case when two or more VaR methods prove to provide comparable forecasts, the distinction among them will be based on values of loss functions.

2.4.1 Conditional coverage

The first back–testing method described in this section is the conditional coverage approach. The main source of information in this section comes from the work of Christoffersen (1998), where the conditional coverage test was presented for the first time. In order to create a comprehensive test for the accuracy of a VaR model, the unconditional coverage test will first be introduced and the conditional coverage test will follow. Together, these two test form the so–called Christoffersen's framework.

In order to adhere to the naming conventions, let's denote y_t the sequence of realized returns. Forecasted VaR at time $t+1$, based on the information available at time t, is further on denoted as $VaR_{t+1|t}$ and p is the coverage probability. An indicator function I_{t+1}, which simply indicates whether or not the forecasted VaR value has been violated, has the following form:

$$I_{t+1} = \begin{cases} 1, & \text{when } y_{t+1} < VaR_{t+1|t} \\ 0, & \text{when } y_{t+1} \geq VaR_{t+1|t} \end{cases}. \tag{2.25}$$

The advantage of the above specified indicator function is clearly the lack of any assumptions on the distribution of the underlying data generating process. Therefore, it is usable even in situations when the actual distribution of the time series in mind is not known. This is particularly advantageous in the case of financial time series due to their non–standard properties as already discussed in previous sections. According to Christoffersen (1998), the series of indicator function values is i.i.d. and has a Bernoulli distribution with parameter p, ie. $\{I_t\} \overset{iid}{\sim} Bern(p)$.

However, the variable of interest for back–testing purposes is the total number of violations $N = \sum_{t=1}^{T} I_t$, which thanks to the properties of I_t follows a Binomial distribution, ie. $N \sim B(T, p)$. The idea of the unconditional coverage test is to test the actual violations ratio $\pi = \frac{N}{T}$ against the hypothetical ratio denoted as p. Therefore the null and alternative hypotheses of the unconditional coverage test are:

$$H_0: \quad p = \pi = \tfrac{N}{T}$$
$$H_A: \quad p \neq \pi = \tfrac{N}{T}$$

In order to test the hypotheses, likelihood functions for both hypotheses need to be specified as 2.26 and 2.27. The likelihood functions then form the unconditional Likelihood Ratio (LR) test as represented by the Equation 2.28. The test statistics LR_{UC} is asymptotically distributed according to the chi–squared distribution with one degree of freedom, ie. $LR_{UC} \overset{asy}{\sim} \chi^2(1)$. Therefore, critical values for particular significance levels can be obtained from this distribution.

$$L(p; I_1, \cdots, I_T) = (1-p)^{T-N} p^N \tag{2.26}$$
$$L(\pi; I_1, \cdots, I_T) = (1-\pi)^{T-N} \pi^N \tag{2.27}$$

$$LR_{UC} = -2\ln\left[\frac{L(p; I_1, \cdots, I_T)}{L(\pi; I_1, \cdots, I_T)}\right] = -2\ln\left[\frac{(1-p)^{T-N}p^N}{(1-\pi)^{T-N}\pi^N}\right] \tag{2.28}$$

Even thought, the unconditional coverage test provides information about whether or not the violations ratio actually equals the allowed violations ratio, it is not the complete information that is needed in order to decide on the adequacy of a VaR model. The problem is that the LR_{UC} test does not say anything about the time dependency of violations. Therefore, it is not possible to asses the possible effect of clustered violations. Angelidis *et al.* (2004) points out that the unconditional coverage test can rule out VaR models with both too high and too low violations ratio. On the other hand, Kupiec (1995) argues that the power of the test is generally low.

One possible way, how to test for independence is the Ljung–Box test proposed by Ljung & Box (1978) or the conditional coverage test by Christoffersen (1998). In addition to the unconditional coverage test, the conditional coverage test suggests a test for independence and a test for joint independence — *conditional coverage*. The main advantage of the conditional coverage test lies in a separation of a possible violations clustering effect from assumptions on their distribution. To start with the independence test, it should be noted that the test hypothesis of conditional coverage is tested against a binary first–order Markov chain, represented by the indicator function I_t, with a transition probability matrix Π_1, as in 2.29.

$$\Pi_1 = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix}, \qquad \pi_{ij} = P(I_t = j | I_{t-1} = i). \tag{2.29}$$

The likelihood function for the independence test of violations has the form presented in the Equation 2.30. The likelihood function is conditional on the first observation. Parameters of the likelihood function can easily be obtained by maximizing the log–likelihood function across its parameters. In this case, maximum log–likelihood estimates of parameters are simple ratios as in the matrix 2.31, where n_{ij} is a number of observations with value i that occur after the observation j; $i, j \in \{0, 1\}$.

$$L(\Pi_1; I_1, \cdots, I_T) = (1-\pi_{01})^{n_{00}} \pi_{01}^{n_{01}} (1-\pi_{11})^{n_{10}} \pi_{11}^{n_{11}} \tag{2.30}$$

$$\hat{\Pi}_1 = \begin{bmatrix} \frac{n_{00}}{n_{00}+n_{01}} & \frac{n_{01}}{n_{00}+n_{01}} \\ \frac{n_{10}}{n_{10}+n_{11}} & \frac{n_{11}}{n_{10}+n_{11}} \end{bmatrix} \tag{2.31}$$

Denoting Π_2 the independence matrix, the first–order Markov chain estimated on violations can be used to test the null–hypothesis that violations are independently distributed. Maximizing the log–likelihood function of Π_2 yields the estimate $\hat{\Pi}_2$.

$$\Pi_2 = \begin{bmatrix} 1-\pi_2 & \pi_2 \\ 1-\pi_2 & \pi_2 \end{bmatrix} \tag{2.32}$$

$$L(\Pi_2; I_1, \cdots, I_T) = (1-\pi_2)^{n_{00}+n_{01}} \pi_2^{n_{10}+n_{11}} \tag{2.33}$$

$$\hat{\Pi}_2 = \hat{\pi}_2 = \frac{n_{01}+n_{11}}{n_{00}+n_{01}+n_{10}+n_{11}} \tag{2.34}$$

As in the case of the unconditional coverage, the LR test denoted LR_{IND} is constructed from the likelihood functions 2.30 and 2.33. As well as LR_{UC}, the test statistics LR_{IND} is asymptotically distributed according to the chi–squared distribution with one degree of freedom, ie. $LR_{IND} \overset{asy}{\sim} \chi^2(1)$. Christoffersen (1998) notes that this test is independent of the actual value of p, thus it really tests only the independence of violations.

$$\begin{aligned} LR_{IND} &= -2 \ln \left[\frac{L(\hat{\Pi}_2; I_1, \cdots, I_T)}{L(\hat{\Pi}_1; I_1, \cdots, I_T)} \right] \\ &= -2 \ln \left[\frac{(1-\hat{\pi}_2)^{n_{00}+n_{01}} \hat{\pi}_2^{n_{10}+n_{11}}}{(1-\hat{\pi}_{01})^{n_{00}} \hat{\pi}_{01}^{n_{01}} (1-\hat{\pi}_{11})^{n_{10}} \hat{\pi}_{11}^{n_{11}}} \right] \end{aligned} \tag{2.35}$$

On behalf of the LR_{IND} test, Angelidis *et al.* (2004) summarize that the test can rule out models that "*generate either too many or too few clustered violations*", however, they also point out that at least several hundred observations should be available for the test to provide accurate results. The derivation of the conditional coverage test, denoted LR_{CC}, is then straightforward. The test in constructed in the way that the null–hypothesis of the unconditional coverage LR_{UC} is tested against the alternative of the independence test LR_{IND}. Since the distribution of both tests is chi–squared with one degree of freedom and the distribution is additive, the distribution of the conditional coverage test statistic is chi–squared with two degrees of freedom, ie. $LR_{CC} \overset{asy}{\sim} \chi^2(2)$. The formula for the conditional coverage test can be written in the form specified by the Equation 2.36.

$$\begin{aligned} LR_{CC} &= -2\ln\left[\frac{L(p; I_1, \cdots, I_T)}{L(\hat{\Pi}_1; I_1, \cdots, I_T)}\right] \\ &= LR_{UC} + LR_{IND} \end{aligned} \tag{2.36}$$

2.4.2 Loss function

Loss function is a simple tool that serves the purpose of measuring the accuracy of various calculation methods. As it has been already said, back–testing approaches can mark several VaR methods as adequate — capable of forecasting VaR at a given α-quantile. However, they do not say anything about the accuracy of such forecasts. Therefore, measuring the accuracy of VaR methods is the domain of the loss function.

A general example of a loss function is presented in the work of Lopez (1998), who suggested a loss function derived from the binomial state loss function of Kupiec (1995). The loss function is based on the squared difference between the forecasted VaR and the actual VaR, which means that not only the actual violation but also the magnitude of the violation is taken into account. The loss function as of Lopez (1998), denoted C_{t+1}, is defined by the Expression 2.37, where y_{t+1} is the actual realized return and $VaR_{t+1|t}$ is the forecasted VaR for time period $t+1$ based on the information set available at time t.

$$C_{t+1} = \begin{cases} 1 + (y_{t+1} - VaR_{t+1|t})^2, & \text{if } y_{t+1} < VaR_{t+1|t} \\ 0, & \text{if } y_{t+1} \geq VaR_{t+1|t} \end{cases} \tag{2.37}$$

The above presented definition is a definition for a loss at time $t+1$. In order to be able to compare the forecast accuracy over the entire sample size, C_{t+1} must be summed as $C = \sum_{t=1}^{T} C_t$. From the definition it is obvious that a VaR method that systematically underestimates the VaR will yield higher value of the loss function than a more accurate method. However, the defined loss function does not penalize systematic overestimation of the VaR. One possibility how to solve this problem is a so–called Quantile Loss (QL) function suggested by Angelidis *et al.* (2004). Another possibility how to overcome the systematic overestimation is to use loss functions defined below. This is the approach adopted by the thesis.

Loss functions defined by equations 2.38 – 2.43 represent several widely used loss functions, as noted by Wei *et al.* (2010). Decision, which one of the provided loss functions to use, is not simple. More information on the topic of choosing an appropriate loss function can be found in the work of Lopez (2001) or Patton (2006). Selected loss functions could be divided into several groups. First there are the Mean Squared Error (MSE) and the Mean Absolute Error (MAE), which are quite simple. Then there are the Heteroskedasticity Adjusted Mean Squared Error (HMSE) and the Heteroskedasticity Adjusted Mean Absolute Error (HMAE). These two loss functions are adjusted for heteroskedasticity. Another widely used loss function is the QLIKE loss function that is calculated as a loss implied by the Gaussian likelihood. The last mentioned loss function is the R^2LOG that "*is similar to the R^2 of the Mincer–Zarnowitz regressions*" (Wei *et al.* 2010, pp. 1480) .

$$MSE = \frac{1}{n}\sum_{t=1}^{T}\left(\sigma_t^2 - \hat{\sigma}_t^2\right)^2 \tag{2.38}$$

$$MAE = \frac{1}{n}\sum_{t=1}^{T}\left|\sigma_t^2 - \hat{\sigma}_t^2\right| \tag{2.39}$$

$$HMSE = \frac{1}{n}\sum_{t=1}^{T}\left(1 - \frac{\sigma_t^2}{\hat{\sigma}_t^2}\right)^2 \tag{2.40}$$

$$HMAE = \frac{1}{n}\sum_{t=1}^{T}\left|1 - \frac{\sigma_t^2}{\hat{\sigma}_t^2}\right| \tag{2.41}$$

$$QLIKE = \frac{1}{n}\sum_{t=1}^{T}\left(\ln\hat{\sigma}_t^2 + \frac{\sigma_t^2}{\hat{\sigma}_t^2}\right) \tag{2.42}$$

$$R^2LOG = \frac{1}{n}\sum_{t=1}^{T}\left[\ln(\frac{\sigma_t^2}{\hat{\sigma}_t^2})\right]^2, \tag{2.43}$$

where $\sigma_t^2 = r_t^2$ represents realized volatility at time t and the term $\hat{\sigma}_t^2$ denotes conditional volatility as predicted by one of the GARCH processes at time t, as suggested by Sadorsky (2006).

In order to be able to compare the loss function of various models, one might use the hypothesis test of Diebold & Mariano (2002), which serves the purpose of evaluating the forecasting ability of various models. The hypothesis test is able to work with various loss functions and violation distributions and its null–hypothesis is no difference in the accuracy of two models. To be able to use the hypothesis test, lets define a sequence $\{d\}_{t=1}^{T}$ as :

$$d_{t+1} = C_{t+1}^{A} - C_{t+1}^{B}, \tag{2.44}$$

where C_{t+1}^{A} is the loss function of a method A and C_{t+1}^{B} is the loss function of a method B. Since a lower value of the loss function is better then a higher value, $d_t < 0$ indicates that the model A was able to forecast the VaR more accurately than the model B. The idea of the hypothesis test lies in a regression of d_t on a constant and using its Student–t statistic value as the test statistics.

The theoretical introduction to the problematic of the VaR forecast calculation is now complete and it can be applied on real data. To summarize, VaR will be calculated using methods described in this chapter, namely the family of GARCH models – the GARCH, EGARCH, and TARCH models. The forecasts adequacy will be measured by a back–testing approach based on the conditional coverage test. In the case that the back–testing approach suggests that several VaR methods are adequate, the final decision on the accuracy of VaR methods will by provided by employing above introduced loss functions and testing them for significance.

Chapter 3

Value-at-risk methods application

In order to test the forecasting accuracy of selected value–at–risk models, a strategy has to be developed first. Models are tested on data from six stock indices that are divided into two groups, as it will be described later in more details. Moreover, models are evaluated first on the in–sample data and then their forecasting capabilities are tested on the out–of–sample data. The out–of–sample subset is further divided into four parts, each with 125 observations. This allows for periodical reestimateion of the models in an interval approximately equal to six months. The reason to include periodical re-estimation of the models lies in the fact that in reality models are re-estimated in certain periods in order to improve their quality and to let them to adapt to new situation on the market. Since the purpose of the thesis is to find out which models have the best forecasting accuracy, the models are ranked according to the theoretical framework described in the previous section.

3.1 Model specification

Even though the thesis tests several hundred models[1], they can all be considered dynamic models. Static models such as the HS method are not discussed in the thesis. The 648 dynamic models, on the other hand, represent a great number of tested models thanks to the variability of VaR model specification.

3.1.1 Dynamic models

The reason for such a high number of dynamic models that are to be tested in the thesis is that dynamic VaR models allow for great variability thanks to pos-

[1] Precisely 648 models for each stock index.

sible combinations of methods that model both mean and variance. Previous sections introduced theoretical frameworks of conditional variance and conditional mean, which are essential parts of dynamic models. The conditional mean is modeled exclusively by the ARMA process and conditional volatility by the GARCH family of processes. The Table 3.1 offers a quick illustration of the dynamic model composition possibilities.

Conditional Mean	Conditional Variance
	GARCH
ARMA	EGARCH
	TARCH

Table 3.1: Dynamic model composition

The decision to employ conditional mean while modeling log–returns was based on the fact that there are no other explanatory variables of the model, except for lagged log–returns. In other words, the mean is modeled as an AR process. The stationarity condition of log–returns is achieved by differencing logarithmic prices in the time series. Since the MA process might also be present in a particular time series, it is included in the mean equation as well. Therefore, modeling the mean using an ARMA process seems to be an optimal choice, since it allows for a great variability in the model selection. The actual choice of ARMA process parameters is subject to the following constraint:

$$(ar, ma) \in \{(0,1),(0,2),(1,0),(1,1),(1,2),(2,0),(2,1),(2,2)\}.$$

In the literature,[2] it is often assumed that the GARCH(1, 1) model is the most appropriate choice for the modeling of conditional variance. This thesis does not consider this assumption as a fact that is always true for every financial time series. Therefore, conditional variance in tested models will be modeled using three GARCH based models — GARCH, EGARCH, and TARCH. Moreover, each model will be subject to changing lagged parameters p and q, which will be drawn from the following set:

$$(p, q) \in \{(0,0),(1,0),(1,1),(1,2),(2,0),(2,1),(2,2)\}.$$

The distribution function that models residuals (error term) represents another aspect in the definition of dynamic models. Most statistical software allow

[2] Ghahramani & Thavaneswaran (2008); Drost & Klaassen (1997) and others.

to specify the distribution function for residuals in dynamic models. Typically, the normal distribution function, the Student–t distribution function, and the GED are available. Calculations in the thesis are performed using the STATA[3] software, which is able to work with all three mentioned distribution functions. Thus all possible combinations of parameters (ar, ma) and (p, q) used for the modeling of conditional mean and conditional variance are extended by the three distribution functions.

3.1.2 Ranking methodology

The theoretical framework in the previous chapter introduced several concepts that are used in the practical part of the thesis. So far only the modeling of conditional mean using the ARMA process and the modeling of conditional volatility using the GARCH family of models were described. The process of estimating and evaluating the various models on selected indices would not be complete without the description of the selection procedure that ranks models based on various criteria.

There are several different ways how to rank estimated models. Since models are first estimated on the in–sample subset, one could calculate the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC) and rank models according to information criteria. Some authors such as Pagan & Schwert (1990), however, discourage from evaluating models using the information criteria, since a superior in–sample performance of a particular model does not necessarily mean a superior out–of–sample performance. Another possibility is to rank models according to the value of their log–likelihood function, but then there is again the problem of the in–sample versus the out–of–sample performance as mentioned with the AIC and the BIC criteria. For this reasons, the evaluation procedure is based on the out–of–sample performance.

The main decision criteria used in the thesis depends on the percentage of failures given a particular confidence interval, which is simply a number of cases when the actual realized log–return $r_{t+1|t}$ is less than the forecasted $\mathrm{VaR}_{t+1|t}$. Since this measure does not address issues described in the section about unconditional and conditional coverage, p–values of the unconditional and conditional coverage are included in the ranking procedure as well. Therefore, the discussion about the ranking procedure lead to the decision to base

[3]The version of STATA used is 11.1. For more information about the software, visit `http://www.stata.com/`

the ranking on the unconditional and conditional coverage, since it is able to capture both the violations rate and the independence of violations.

By ranking models according to above mentioned variables, it is meant to simply sort models with respect to the p–value of the conditional coverage test in an ascending order and incrementally assigning each model a number from the set $\{1, 2, ..., N\}$, where N represents the total number of actually estimated models. In the case when several models obtain the same p–value of conditional coverage, values of selected loss functions provide final distinction between such models. The comparison of loss function values is performed according to the Diebold & Mariano (2002) framework described in the previous chapter. After all ranking variables have been used to rank the models, there should be one model left that is the most appropriate for the particular index and confidence interval.

By inclusion of periodical re-estimation, as examined in the 3.3, the ranking methodology must be further extended in order to grasp the performance of the models after all re-estimations. As described in the previous paragraph, the models are after each re-estimation ranked using the conditional coverage framework and incrementally assigned number based on the their rank among other models. To merge the individual performances into one overall performance indicator, a geometrical average is computed from the product of both the unconditional and conditional coverage for each model with respect to confidence intervals. The geometrical average then serves as the main indicator of the overall performance for each model after a series of re-estimations.

3.2 Model application

The analysis of the selected models is performed on six stock indices, which are listed below. Datasets have been selected with the requirement to grasp indices from different parts of the world and not just one or two usual US stock indices. Therefore, in addition to indices from the US, datasets from Germany, Japan, and the United Kingdom have been chosen. The indices are the following:

- Dow Jones Industrial Average (DJI) – US stock based index
- SPDR S&P 500 (GSPC) – US stock based index
- NASDAQ Composite (IXIC) – US stock based index
- FTSE 100 (FTSE) – UK stock based index

- German DAX Index (GDAXI) – German stock based index
- NIKKEY 225 (N225) – Japanese stock based index

3.2.1 Data analysis

Each dataset includes exactly 1500 observations from which the first 1000 represent the initial in–sample subset and the remaining 500 represent the out–of–sample subset. Observations in all datasets are centered around the end of the year 2007 – more precisely 31st December 2007. The reason is to have an artificial point for the division of each dataset into the in–sample and the out–of–sample subsets. Since the stock indices come from several countries and trading dates are not internationally standardized, the beginning of the in–sample period and the end of the out–of–sample period vary by several days across the indices. Thanks to the clear division point between the subsets and the exact number of observations for each index, it is possible to evaluate the models on all datasets under equal conditions.

In–sample subset

The descriptive summary of the log–returns in the in–sample subset for each index is presented in the Table A.1. Log–returns are based on daily returns. All indices have kurtosis higher than 3, which is the kurtosis of the normal distribution. Moreover, all indices are characterized with a negative skewness, mostly below -0.3, which is an indication of fat–tails on the left side of the distribution. Histogram plots of the indices on Figure B.3 also indicate that there are fat–tails on the left side of the distribution. Therefore, log–returns might not be distributed normally. The hypothesis that log–returns are normally distributed is tested using the Jarque–Bera test statistic and the results confirm that the null–hypothesis of normality is rejected at all significance levels for each index, except for the IXIC index, where the null–hypothesis cannot be rejected at 1% significance level. Values of the Jarque–Bera test statistic and their corresponding p–values are presented in the Table A.1.

Log–return plots of the indices on Figure B.1 and realized volatility plots on Figure B.2 provide a quick view at the volatility of the indices. The in–sample subset is more or less stable and its volatility does not seem to be very high either, which is in line with the data provided in the Table A.1. According to the figures and the summary table, it seems that the indices behave relatively similarly. Both minimum and maximum values are quite similar and also mean and variance values lie around the same values. Closer look at the Figure B.4 with autocorrelation and the Figure B.5 with partial autocorrelation of residuals[4] suggests that AR and MA processes are present in all indices. The order of AR and MA processes, however, cannot be clearly determined, since partial autocorrelation graphs do not contain a clear pattern. Therefore, the exact number of lags shall be determined by the log–likelihood maximization procedure of a particular dynamic model in the statistical software.

In order to test for the possibility of GARCH family of processes in the variance, figures with squared residuals plots represent a good starting point. Autocorrelation functions are presented in the Figure B.6 and partial autocorrelation functions in the Figure B.7. The IXIC and N225 indices do not seem to include any GARCH family of processes. On the other hand, figures for the rest of the indices do not reject the possibility of GARCH based processes in the in–sample subset. The presence of a GARCH process in the in–sample subset could be determined by maximizing the log–likelihood function of dynamic models, as in the case of the AR process.

The application of the AR process, the MA process and GARCH family of processes is conditioned by stationarity of the time series. Therefore, each in–sample subset has been tested by the Augmented Dickey–Fuller test for the presence of a unit root, which would effectively mean that the particular in–sample subset is not stationary. The application of the mentioned process in the presence of a unit root would not be possible. Tables A.3 – A.8 provide results of the Augmented Dickey–Fuller test together with corresponding critical values at three significance levels. From the tables it is clear that the differenced log–returns are stationary, since the null–hypothesis of a unit root up to the tenth lag has been rejected at all significance levels for each index, with the exception of the N225 index, where the null–hypothesis has been rejected at all significance levels only up to the fifth lag. Since models that are to be applied work with a maximum of two lags, higher order autocorrelations should not cause any problems.

Out–of–sample subset

The descriptive summary of the out–of–sample subset is presented in the Table A.2. The subset contains 500 observations for each index, starting on 1st Jan-

[4]Residuals were obtained by regressing the log–returns of each index on a constant.

uary 2008. From figures of log–returns and realized volatility, it is clear that the out–of–sample subset has much different characteristics than the in–sample subset. Both minimum and maximum values are more extreme and the mean of the indices has shifted from positive values to negative values for all indices. The skewness has actually improved in the sense that it is closer to the skewness of the normal distribution, which is equal to zero. The reason for that might lie in a lower number of observations and a higher number of both extreme positive and negative realized returns. On the other hand, the kurtosis of the indices in the out–of–sample subset has increased and reached values around 7. For most indices, it is almost twice as much as in the in–sample subset. The kurtosis is obvious from the histogram on the Figure B.8. To test the distribution of the out–of–sample subset for normality, the Jarque–Bera test statistic is employed. P–values of the test statistics reject the null–hypothesis of normality at all significance levels for each index. Actual values of the Jarque–Bera test statistics and their corresponding p–values are presented in the summary table, as well.

The variance of the out–of–sample has also increased. The increase amounts to approximately 5 to 8 times the variance of the in–sample subset, depending on the particular index. The volatility plot depicted on Figure B.2 provides a better illustration of the increase in the log–returns variance. The out–of–sample subset is therefore quite turbulent, which is very suitable for the purpose of the thesis, since the topic is to evaluate various VaR calculation methods in the period of higher volatility. There are also signs of volatility clustering, as indicated on the figures B.1 and B.2.

The autocorrelation and partial autocorrelation functions of residuals[5], depicted on figures B.9 and B.10 suggest that there might be AR and MA processes in the subset for each index, since autocorrelation functions are slowly decreasing and partial autocorrelation functions contain spikes at certain lags. As well as in the in–sample subset, the actual order of AR and MA processes varies index by index and it seems that AR and MA processes are of a high order, since there are typically up to ten significant lags. This could be interpreted as a long–time memory of the process, which means that the past realized log–returns might play a significant role in the calculation of the conditional mean.

Graphical analysis of squared residuals, depicted on the autocorrelation and partial autocorrelation function on figures B.11 and B.12, attempts to provide an insight on the behavior of the variance of the indices during the out–of–sample period. As well as in the in–sample subset, the GARCH family

[5]Residuals were obtained by regressing the log–returns of each index on a constant.

of processes seems to be present in the out–of–sample subset for all indices except for the GDAXI index. The highest order GARCH process seems to be present in the FTSE and the N225 indices.

From the analysis of the in–sample and the out–of–sample subsets it is clear that the data have quite different properties. Since the models are estimated on the in–sample data and then applied on the out–of–sample data, it is interesting to observe, whether or not the models are able to use the past realized log–returns at time t in order to provide an appropriate VaR forecast for time $t + 1$. Results of the models application are discussed in following sections. The results are analyzed on per index basis, since every model is estimated separately for each index. The evaluation of the models proceeds with the analysis of the top in–sample estimated models, according to their rank based on AIC and Log–Likelihood (LL) values. Then the top out–of–sample VaR model evaluation, based on the rank of the unconditional and conditional coverage, completes the analysis of the accuracy of the various models for a particular index.

As well as in the case of the in–sample subset, the out–of–sample subset must fulfill the condition of stationarity, when the AR process, the MA process and the GARCH family of processes are to be applied. Stationarity of the indices is tested using the Augmented Dickey–Fuller test and the results of the test for each out–of–sample subset are available in tables A.9 – A.14. From the results it is clear that the null–hypothesis of a unit root up to the tenth lag has been rejected at all significance levels for each out–of–sample subset, with the exception of the N225 index, as it was the case also for the in–sample subset. The unit root of the N225 index has been rejected at all significance levels up to the sixth lag. Since the considered models work with a maximum of two lags, the autocorrelation of squared residuals at higher lags should not represent a problem.

3.2.2 DJI index

In the in–sample subset summary section it is concluded that according to the Jarque–Bera statistic, the DJI timeseries is not normally distributed. The main reason is in the presence of a fat–tail on the left side of the distribution, where negative log–returns are depicted. It is therefore not surprising that when the estimated models are ranked by their AIC and LL values, the top ranking models are the ones that assume either the Student-t or the GED for the distribution of the residuals. The residuals' estimated degrees of freedom parameter of the Student-t distribution lies in most cases between 7 and 8 and the shape parameter for the GED lies between 1.4 and 1.5.

The results of the Portmanteau Q test for serial autocorrelation are presented in the Table 3.2. Even thought the test does not reject the null–hypothesis of no serial autocorrelation of the squared residuals at 2 lags at 3% significance level, the best in–sample ranking models employ either the TARCH(2,2) process or the EGARCH(2,1) for conditional variance. The list of the top eight in–sample models is in the Table 3.3 and the estimated parameters of the model with the highest AIC are presented in the Table C.1. From the Table 3.2 it is obvious that the application of the conditional volatility process TARCH(2,2) has a positive effect on the serial autocorrelation of squared residuals. The null–hypothesis of the Portmanteau Q test at the selected lags cannot be rejected at 6.54% significance level in all cases.

	Test statistic value	p-value
In–sample before model application		
Portmanteau Q(1)	1.5563	0.2122
Portmanteau Q(2)	6.7867	0.0336
Portmanteau Q(3)	16.7373	0.0008
Portmanteau Q(5)	66.3841	0.0000
Portmanteau Q(10)	164.6121	0.0000
In–sample after model application		
Portmanteau Q(1)	0.0847	0.7711
Portmanteau Q(2)	0.3818	0.8262
Portmanteau Q(3)	0.5157	0.9154
Portmanteau Q(5)	6.6990	0.2440
Portmanteau Q(10)	17.4277	0.0654

Table 3.2: Serial autocorrelation tests for the DJI index

Quite surprising is the fact that the top ranking in–sample models are dominated by the order of the particular TARCH or EGARCH process and the orders of the AR and MA processes seem to play a minor role. In other words, the most dominant factor, while estimating the in–sample models on the DJI index, is the conditional volatility. It might, however, be possible that the process, which models the conditional volatility, is of a higher order than in the estimated models, since the models are estimated only up to $p = q = 2$. To summarize, the in–sample model estimates prove that the distribution of the log–returns is not normally distributed and that the log–returns are not homoskedastic.

In–sample model	AIC	LL
AR(2)-MA(1)-TARCH(2,2)-GED	-7172.359	3598.180
AR(2)-MA(2)-EGARCH(1,2)-GED	-7164.760	3593.380
AR(2)-MA(2)-TARCH(2,2)-GED	-7164.157	3595.079
AR(0)-MA(1)-TARCH(2,2)-GED	-7163.590	3591.795
AR(1)-MA(0)-TARCH(2,2)-GED	-7163.579	3591.790
AR(1)-MA(1)-EGARCH(2,1)-T	-7163.299	3591.649
AR(2)-MA(2)-EGARCH(2,1)-GED	-7162.218	3593.109
AR(2)-MA(0)-TARCH(2,2)-GED	-7161.605	3591.803

Table 3.3: Top eight in–sample models for the DJI index

To test the accuracy of the one–day–ahead VaR forecast, the estimated models are applied to the out–of–sample data. Since the out–of–sample data exhibit higher volatility, it is not surprising that the best performing out–of–sample models are different from the best performing in–sample models. The evaluation of the accuracy of the models is calculated for the following confidence intervals $\alpha = \{0.90, 0.95, 0.99\}$. The top eight performing models for the VaR forecasts for the DJI index are presented in the tables A.15, A.16 and A.17, according to the selected confidence level. It is interesting that even though the TARCH and EGARCH models are selected during the in–sample estimation, they exhibit very poor results in the out–of–sample evaluation. None of the out–of–sample top performing models uses the TARCH and EGARCH process for the conditional volatility.

Evaluation at $\alpha = 0.90$

The most accurate forecasts of the one–day–ahead VaR are achieved with models that assume the Student-t distribution with 8 degrees of freedom as the distribution for the error term. The degrees of freedom parameter is estimated using the statistical software while estimating the models. The low number of degrees of freedom corresponds to the presence of fat–tails discovered during the in–sample data analysis. The top eight out–of–sample models are presented in the Table A.15 in the Appendix A. Unlike the in–sample models, the out–of–sample models are dominated by models that model conditional volatility using the GARCH(2,1) process, as it is the case with five out of eight models. The remaining three models use GARCH(1,1) process for modeling the conditional volatility.

All of the top eight models underestimate the VaR by a small amount, as the percentage of cases, when the VaR forecast does not exceed the realized log–return, lies between 12.4% and 12.6%. Since the unconditional coverage measures whether the percentage of failures corresponds to the desired α level and punishes deviations to both sides equally, the p–value of the unconditional coverage, p_{uc}, is quite small. Nevertheless, at the significance level of 5% one cannot reject the null–hypothesis in either of the eight models. Similar situation is with the p–value of the conditional coverage, p_{cc}. When this value is used as the main selection criterion, there are only three models left that cannot be rejected at the 5% significance level and all three of them have the same value of p_{cc}. The differences among them are quite small, however, the priority is given to the model which performs better in means of the loss function values. Therefore, the most accurate VaR model for the DJI index at confidence interval $\alpha = 0.90$ is the AR(1)-MA(2)-GARCH(1,1)-T model.

Evaluation at $\alpha = 0.95$

As well as for the confidence interval $\alpha = 0.90$, the most accurate VaR forecasts are achieved with the Student-t distribution with 8 degrees of freedom as the assumed distribution for the error term. The top eight out–of–sample models are listed in the Table A.16. The prevailing process for modeling the conditional volatility at this confidence level is the GARCH(1,1) process, with four out of eight cases. The remaining conditional volatility processes are GARCH(2,1) and GARCH(2,2). The conditional mean process does not seem to follow any particular pattern, as the order of the AR processes varies from 0 to 2. The same situation is with the MA process.

The models at this confidence level also underestimate the VaR. Quick look at the failure rate proves that the failure rate, ranging from 7.0% to 7.4%, is more than the expected rate of 5%. The result of such underestimation is projected to the p–value of the unconditional coverage, which is higher than 5% only in two cases. The value of the conditional coverage p–value is somewhat better, since six out of eight models pass the selection criterion at 5% significance level, however, based on the value of p_{cc}, there are only two superior models. To select a single model with the best performance, the realized values of the loss functions are compared. The best performing VaR model for the

DJI index at confidence interval $\alpha = 0.95$ is the AR(1)-MA(0)-GARCH(2,2)-T model.

Evaluation at $\alpha = 0.99$

Concerning the distribution of the error term, the evaluation at the confidence interval $\alpha = 0.99$ does not differ from the previous confidence intervals. The top eight performing out–of–sample models are presented in the Table A.17. All of the eight best performing models model the conditional volatility as a GARCH(1,1) process. The AR and MA processes do not seem to follow a specific pattern and their orders range from 0 to 2.

At this confidence level, all eight models achieve the same failure rate of 1%, which is exactly the expected value of the failure rate. The exact performance is projected to the value of the conditional coverage p_{cc}, as well as to the value of the unconditional coverage, p_{uc}, which is equal to one. Moreover, each model passes the selection criteria of unconditional and conditional coverage at quite high significance levels. Since the models attain the same p–values of the conditional coverage p_{cc} they are compared using the values of the loss functions, which leads to the conclusion that at the confidence interval $\alpha = 0.99$ the model AR(1)-MA(1)-GARCH(1,1)-T is the best performing one for the DJI index.

Note

There are three models that appear in all three tables (A.15, A.16 and A.17) with the best performing models for the given confidence intervals. These models are the AR(1)-MA(0)-GARCH(1,1)-T model, then the AR(1)-MA(2)-GARCH(1,1)-T model and the AR(2)-MA(1)-GARCH(1,1)-T model. This could be interpreted as three possibilities how to model the log–returns for the DJI index, when the confidence interval is abstracted from. On the other hand, at two of the three confidence intervals there is at least one model, which outperforms the three mentioned models, in terms of the p_{cc} value. Thus the decision on whether or not to consider one of these models as an appropriate general model for the DJI index, should be considered carefully.

The results prove that in the case of the DJI index it is not optimal to calculate one–day–ahead VaR using models that attain the highest values of either AIC or LL. Even though these models might perform relatively well on the in–sample subset, they do not provide adequate results when applied to the out–of–sample subset. Conditional volatility processes such as EGARCH and TARCH seem to be adequate in quite stable times in terms of volatility. On the other hand, when a period with higher volatility is expected, the results suggest to use a simple GARCH model, for example the GARCH(1,1) model, since it outperforms the other models in terms of the conditional coverage test, which captures both accuracy of the VaR forecast and the independence of the realized violations of the predicted VaR.

3.2.3 GSPC index

Looking at the GSPC index histogram plot depicted on the Figure B.3, it is clear that the log–returns distribution cannot be considered as the normal distribution. There seems to be a quite high number of extremely negative log–returns, which lie above the line representing the border of the normal distribution. The hypothesis of normally distributed log–returns is therefore rejected by the Jarque–Bera test, as presented in the in–sample summary table. The value of the test statistic is even higher than for the DJI index. Thus the best performing in–sample models, in terms of AIC and LL values, are the ones working with either the Student-t or the GED as the distribution for the error term. The estimated degrees of freedom parameter for the Student-t distribution lies at 7.5 and the estimated shape parameter for the GED is approximately 1.4. Both values indicate fat–tails in the distribution.

	Test statistic value	p-value
In–sample before model application		
Portmanteau Q(1)	4.9255	0.0265
Portmanteau Q(2)	15.5650	0.0004
Portmanteau Q(3)	33.0938	0.0000
Portmanteau Q(5)	100.3342	0.0000
Portmanteau Q(10)	248.9185	0.0000
In–sample after model application		
Portmanteau Q(1)	0.0001	0.9915
Portmanteau Q(2)	0.0001	0.9999
Portmanteau Q(3)	0.0003	1.0000
Portmanteau Q(5)	5.4464	0.3639
Portmanteau Q(10)	12.6551	0.2436

Table 3.4: Serial autocorrelation tests for the GSPC index

The Table 3.4 with results of the Portmanteau Q test for serial autocorrelation of the squared residuals indicates that a serial autocorrelation process is most likely part of the time series. Since the test for serial autocorrelation of order two and more rejects the null–hypothesis of no serial autocorrelation of the specified order, the best in–sample models should include a conditional volatility modeling process. This is indeed the case, as all the best in–sample models use the EGARCH(2,2) process in order to model conditional volatility, as presented in the Table 3.5 and estimated parameters of the model with the highest AIC are presented in the Table C.2. The Table 3.4 provides results of the Portmanteau Q test after the application of the conditional volatility process EGARCH(2,2). The process effectively captures the serial autocorrelation, since the null–hypothesis of the Portmanteau Q test at selected lags cannot be rejected even at 24.36% significance level in all cases.

In–sample model	AIC	LL
AR(1)-MA(1)-EGARCH(2,2)-T	-7102.982	3562.491
AR(1)-MA(1)-EGARCH(2,2)-GED	-7102.257	3562.129
AR(2)-MA(0)-EGARCH(2,2)-T	-7101.815	3561.907
AR(0)-MA(2)-EGARCH(2,2)-GED	-7100.620	3561.310
AR(0)-MA(1)-EGARCH(2,2)-GED	-7100.537	3560.269
AR(2)-MA(0)-EGARCH(2,2)-GED	-7100.476	3561.238
AR(1)-MA(2)-EGARCH(2,2)-T	-7100.182	3562.091
AR(2)-MA(1)-EGARCH(2,2)-T	-7100.044	3562.022

Table 3.5: Top eight in–sample models for the GSPC index

The best in–sample VaR models are dominated by the conditional volatility part. The conditional mean is modeled with multiple combinations of AR and MA processes with orders ranging from 0 to 2. Therefore, it is obvious that volatility plays the most important role in the prediction of the one–day–ahead VaR. It is possible that orders of the EGARCH process are not the most optimal ones, since the tested models are of a maximum order $p = q = 2$. The presence of the conditional volatility process corresponds with the outcomes of the performed Portmanteau Q tests for serial autocorrelation.

The application of the estimated models on the out–of–sample data is a good stress test, since the out–of–sample subset is more volatile than the in–sample subset. This might be the reason, why the best in–sample models are all outperformed by models with different conditional volatility modeling process. In fact, none of the top eight in–sample models appears in the result tables for the out–of–sample period. All models are calculated for three confidence intervals $\alpha \in \{0.90, 0.95, 0.99\}$ and the best performing out–of–sample models are summarized in the tables A.18, A.19 and A.20.

Evaluation at $\alpha = 0.90$

The best out–of–sample results of the GSPC index at the confidence interval $\alpha = 0.90$ are achieved under the assumption that the error term is distributed according to the Student–t distribution with 7 to 8 degrees of freedom. This is in line with the results of the DJI index, where the distribution and its parameters are very similar. The top eight models for this confidence interval treat the conditional volatility as a GARCH process, five out of them as GARCH(1,1) process, one as a GARCH(1,2) and two as a GARCH(2,2). The top eight models are summarized in the Table A.18.

All eight best performing models underestimate the VaR, since the failure rate lies between 12.0% and 13.2%, even thought the theoretical failure rate should be 10%. The difference between the lowest and the highest values amounts to 1.2%, which is the highest difference among the selected indices and confidence intervals. On the other hand, the p–values of the unconditional coverage test, p_{uc}, indicate that one cannot reject the null–hypothesis of the failure rate being equal to 10% at a significance level 5% for five models. The p–values of the conditional coverage test, however, are not as high. The null–hypothesis cannot be rejected at the significance level of 1% at the most. It seems that violations of the predicted VaR are not randomly distributed and they appear repeatedly one after another. To conclude the evaluation at the $\alpha = 0.90$ confidence interval, the most accurate model seems to be the AR(2)-MA(2)-GARCH(1,1)-T model, since it exhibits the highest values of both conditional and unconditional coverage and the values of its loss functions are the lowest among the compared models.

Evaluation at $\alpha = 0.95$

Although the best results at this confidence interval are achieved with the Student–t distribution such as in the previous confidence interval, the estimated number of degrees of freedom is closer to 7, for all models. The conditional variance process differs from the process at the $\alpha = 0.90$ confidence interval. Even though the conditional variance is still modeled using the GARCH process, as opposed to the in–sample EGARCH process, its parameters are different.

Five out of eight models are modeled using the GARCH(2,1) process, two with GARCH(1,1) process and one with GARCH(2,2) process.

The top eight out–of-sample models underestimate the VaR, as the failure rates lie between 7.2% and 8.2%. This fact is reflected in the p–value of the unconditional coverage that suggests, that the null–hypothesis cannot be rejected at the 3.37% significance level only for two models. Even worse numbers attains the p–value of the conditional coverage. The values do not even reach 1%, which indicates that even the best performing models are not very precise. As well as in the previous case, the violations of the log–returns are followed one by another, which worsens the value of the conditional coverage. According to the discussed p–values and the values of the loss functions, the best performing model for the GSPC index at the $\alpha = 0.95$ confidence interval is the AR(1)-MA(0)-GARCH(2,1)-T model.

Evaluation at $\alpha = 0.99$

The results of the evaluation at the highest confidence interval considered are quite similar to the results of the evaluation at the same confidence interval for the DJI index. The error term is assumed to be distributed according to the Student–t distribution with 7 degrees of freedom. This is in line with the presence of a fat–tail on the negative side of the log–returns for this stock index. Six out of eight models take advantage of the conditional variance process modeled as a GARCH(2,1) process. The remaining two models are modeled using the GARCH(1,1) process. The models vary in the orders of the AR and the MA process, as their orders attain values from 0 to 2. The table with the results is presented in the appendix as the Table A.20.

None of the models overestimates the VaR, which means that the failure rates are greater than 1% for all top eight models. Quite interesting is the fact that all models have the same values of both the unconditional and conditional coverage. Thus one cannot reject the null–hypothesis of either test even at the significance level of 66.3%. On the other hand, the p-values of the tests in mind are usually quite high for the $\alpha = 0.99$ confidence interval. The decision to select one model out of the eight as the best one therefore depends only on the values of the loss functions. Surprising is the fact that according to the values of the loss functions, the wining model is using the GARCH(1,1) process for conditional volatility, even though the majority of the top eight models uses GARCH(2,1) process for the conditional volatility. To conclude, the best performing model for the GSPC index at the $\alpha = 0.99$ confidence interval is the AR(1)-MA(0)-GARCH(1,1)-T model. One must, however, keep in mind that all eight models exhibit the same values of both the conditional and unconditional volatility.

Note

Unlike as in the case of the DJI index, there is not any particular model that is present in each table with the best scoring models for a given confidence interval. Important fact from the analysis of the GSPC index is that all models at the $\alpha = 0.95$ confidence interval are rejected based on the p–values of the conditional coverage, since the p–values are less than 1%.

As well as for the DJI index, the most accurate one–day–ahead VaR forecasts are not achieved with in–sample models that prove to be the best performing ones based on AIC and LL values. The top in–sample models model conditional volatility exclusively with the EGARCH process; however, the out–of–sample data are quite different from the in–sample data and therefore the conditional volatility process is not the same. Since the best results for the one–day–ahead VaR are achieved using the GARCH process, this process is the recommended conditional volatility process for the GSPC index in period with higher volatility.

3.2.4 IXIC index

The histogram plot of IXIC index log–returns is depicted on the Figure B.3. The histogram indicates kurtosis greater than 3 and a negative skewness. The negative skewness is an indicator of a fat–tail on the left side of the distribution, which means that negative log–returns occur more often than they should appear according to the normal distribution. The summary table of the in–sample subset confirms the information from the figure and provides exact numbers. The normality of the log-returns of the IXIC index is tested using the Jarque–Bera test statistic in the same way as by the previous indices. Interesting result is that the normality test cannot be rejected at the 1% significance level. The 5% significance level, however, already rejects the null–hypothesis. Therefore it could be expected that some of the best performing in–sample models should use normal distribution as the distribution of the error term. This proves to be true, since two out of the eight best performing models actually use normal distribution for the error term. The other six models use the Student–t distribution with 16 to 21 degrees of freedom.

	Test statistic value	p-value
In–sample before model application		
Portmanteau Q(1)	0.4496	0.5025
Portmanteau Q(2)	12.6987	0.0017
Portmanteau Q(3)	16.6745	0.0008
Portmanteau Q(5)	41.7432	0.0000
Portmanteau Q(10)	111.6832	0.0000
In–sample after model application		
Portmanteau Q(1)	0.4185	0.5177
Portmanteau Q(2)	0.4216	0.8100
Portmanteau Q(3)	0.9758	0.8071
Portmanteau Q(5)	1.8497	0.8695
Portmanteau Q(10)	9.4724	0.4879

Table 3.6: Serial autocorrelation tests for the IXIC index

The results of the Portmanteau test for serial autocorrelation of the squared residuals are presented in the Table 3.6. There does not seem to be any serial autocorrelation at lag one, since the null–hypothesis cannot be rejected at 50% significance level. The tests for serial autocorrelation at 2 or more lags, however, already reject the null–hypothesis of no serial autocorrelation. Therefore it can be expected that the best performing in–sample models should be the ones that model conditional variance with a process of at least two orders. Indeed, the expectation is confirmed as all of the top eight in–sample models employ an EGARCH process with two lags for the autocorrelation term as the conditional volatility process. The table with the top eight in–sample models is provided in Table 3.7 and the estimated parameters of the model with the highest AIC are presented in the Table C.3. The top in–sample model seems to be an appropriate one, considering the application of the conditional volatility process, as presented in the Table 3.6, which provides the results of the Portmanteau Q test after the application of the conditional volatility process EGARCH(2,2). The serial autocorrelation is effectively captured, since the null–hypothesis of the Portmanteau Q test at the selected lags cannot be rejected even at 48.79% significance level in all cases.

The top eight in–sample models are all based on the EGARCH(2,1) or the EGARCH(2,2) processes as the conditional volatility modeling part. It is important to mention that the models are estimated for $p = q = 2$ at the most. Thus it might be possible that the actual conditional volatility process is of a

In–sample model	AIC	LL
AR(0)-MA(2)-EGARCH(2,2)-T	-6525.862	3273.931
AR(1)-MA(1)-EGARCH(2,2)-N	-6523.866	3271.933
AR(0)-MA(2)-EGARCH(2,2)-N	-6523.205	3271.602
AR(0)-MA(1)-EGARCH(2,1)-T	-6516.296	3267.148
AR(1)-MA(0)-EGARCH(2,1)-T	-6516.290	3267.145
AR(0)-MA(2)-EGARCH(2,1)-T	-6515.787	3267.893
AR(2)-MA(0)-EGARCH(2,1)-T	-6515.675	3267.837
AR(1)-MA(2)-EGARCH(2,1)-T	-6515.362	3268.681

Table 3.7: Top eight in–sample models for the IXIC index

higher order than the estimated ones. It has been noted that the Jarque–Bera test did not reject the null–hypothesis of normality of the log–returns at the 1% significance level. Therefore it is not surprising that two out of the eight models actually consider normal distribution as the distribution for the error term.

Since the out–of–sample subset has greater volatility and kurtosis, the application of the models on the out–of-sample subset might change the order of the best performing models. A likely outcome is that the models that use normal distribution for the error term will not score among the top eight models. The models are evaluated at three confidence intervals $\alpha \in \{0.90, 0.95, 0.99\}$ and the best performing out–of–sample models are summarized in the tables A.21, A.22 and A.23.

Evaluation at $\alpha = 0.90$

The models evaluation results on the out–of–sample subset are quite consistent in terms of the conditional volatility process. Six out of the top eight out–of–sample models are modeled using the GARCH(1,1) process for the conditional volatility with Student–t as the optimal distribution for the error term. The degrees of freedom parameter ranges from 16 to 17. The other two models take advantage of the GARCH(2,2) process with the same distribution function for the error term as the GARCH(1,1) process. The results for the IXIC index at this confidence interval are similar to the results for the two previous indices, when the conditional volatility process is mostly GARCH(1,1).

A quick look at the table with the top eight models at this confidence interval provides an indication that even the best performing models underestimate the VaR. The failure rate is quite high, as it ranges from 13.2% to 14.2%, even though the required failure rate lies at 10%. The underlying conditional volatility process is not able to capture the behavior of the out–of–sample subset very precisely. This is also observable on the p–values of the unconditional and conditional coverage. Ranking the models based on the unconditional coverage, the models cannot be rejected only at the 2.23% significance level, at the highest. The conditional coverage values provide even less convincing information about the predicting accuracy of the models, as the p–values are less then 1%. Even though the results are not as good as expected, the best model among the eight models at the $\alpha = 0.90$ confidence interval can still be selected. Based on the values of the conditional coverage and the loss functions, it is the AR(1)-MA(1)-GARCH(1,1)-T model.

Evaluation at $\alpha = 0.95$

Among the successful models at the $\alpha = 0.95$ confidence interval there are models with two types of the conditional volatility process. There are six models that employ the GARCH(1,1) process for the conditional volatility, such as at the $\alpha = 0.90$ confidence interval. Moreover there are two models that take advantage of the GARCH(2,2) process. All of the models use the Student–*t* distribution for the error term, with 16 to 17 degrees of freedom. Should the number of top performing models be greater than 20, there would also be several models that use the normal distribution for the error term. This observation is in line with the suggestion made in the introductory part of the IXIC index.

As well as at the previous confidence interval, the results of models applications on the out–of–sample subset do not provide satisfactory results. The failure rates range from 7.8% to 8.0%, which is significantly higher than the expected 5% failure rate. Therefore, the p–values of the unconditional coverage lie all below the 1% value. The null–hypothesis of the failure rate being equal to 5% can thus be rejected even at the significance level of 1%. The values of the conditional coverage are even smaller, which means that the null–hypothesis can be rejected at the significance level of 1% for all models. The four models with the highest values of the conditional coverage are then compared based on loss functions values. Based on this criteria, the AR(1)-MA(0)-GARCH(1,1)-T model is selected as a model with the best performance at the confidence interval of $\alpha = 0.95$. It is important to note, however, that based on the values of the conditional coverage, none of the models proves to be an adequate one.

Evaluation at $\alpha = 0.99$

The top eight performing out–of–sample models at the confidence interval of $\alpha = 0.99$ are the same models as for the confidence interval of $\alpha = 0.99$ in the DJI index, which is an interesting outcome. The distribution of the error term is assumed to be the Student–t with 16 to 17 degrees of freedom. At the first sight, it might seam that the behavior of the IXIC index can be captured with the GARCH(1,1) process for the conditional volatility. On the other hand, it must not be forgotten that the values of the unconditional and conditional coverage — the main selection criteria — are not very convincing.

This is true also for the results of the evaluation at the $\alpha = 0.99$ confidence interval. Even though the expected failure rate of the one–day–ahead VaR forecast is equal to 1%, the observed failure rates range from 1.8% to 2.0%. This corresponds to the p–values of the unconditional coverage which indicate that the null–hypothesis of a failure rate equal to 1% cannot be rejected at the significance level of 10% only in two cases. In comparison with the p–values of the unconditional coverage for DJI and GSPC indices, where the same p–values are mostly over 60%, the p–values for the IXIC index are not very convincing. The same situation is with the p–values of the conditional coverage, since only two models reach a value greater than 20%. Based on this information and the loss functions values, the most accurate model for the IXIC index at the confidence interval $\alpha = 0.99$ is the AR(1)-MA(0)-GARCH(1,1)-T model, even though its p–values of both the unconditional and conditional coverage are not very persuasive.

Note

The distribution of the IXIC index log–returns resembles the normal distribution at the most among the selected indices. On the other hand, the shape of the out–of–sample distribution is not the same, which is the most likely reason why the results of the out–of–sample evaluation do not provide convincing results. The models achieve low p–values of both unconditional and conditional coverage and the failure rates are quite high. The AR(1)-MA(0)-GARCH(1,1)-T model is selected as the best performing one for the confidence intervals $\alpha = 0.95$ and $\alpha = 0.99$ and the AR(1)-MA(1)-GARCH(1,1)-T for the $\alpha = 0.90$. It seems that the underlying process for the conditional mean consists of an autoregressive part of order one and that the conditional volatility is at the best modeled using the GARCH(1,1) process. Based on the selection criteria, the decision to employ the wining models should be carefully considered.

The outcome of the analysis is quite similar to the outcomes of the previous two indices. Even though the EGARCH process performs very well for the in–sample subset, which exhibits lower volatility, the one–day–ahead VaR forecasts based on the out–of–sample subset provide different results. The best accuracy is achieved when the conditional volatility process is modeled using the GARCH process. Thus also the results for the IXIC index confirm that the best in–sample models based on the AIC and LL values are not the suggested models for periods with higher volatility.

3.2.5 FTSE index

The properties of the in–sample subset of the FTSE index are provided in the Table A.1. The histogram on Figure B.3 indicates that the FTSE index log–returns exhibit a significant kurtosis and most likely also skewness. This suspicion is confirmed by the properties table referenced above, from which it is obvious that the FTSE index has the highest kurtosis of all selected indices. The same situation is with the skewness, which indicates a fat–tail on the side of the negative log–returns. In order to test the log–returns for normality, the Jarque–Bera test is employed. According to the test statistic, the null–hypothesis of normality is rejected, which is in line with the observed characteristics from the histogram. This is also the reason why the top eight in–sample models, which are ranked according to the values of AIC and LL, work with the assumption that the error term is distributed either according to Student–t distribution with 16 to 22 degrees of freedom or the GED distribution with the shape parameter equal to 1.8. Both parameters indicate fat–tails in the log–returns distribution.

To test for the possibility of serially correlated squared residuals, the Portmanteau Q test is employed. The results of the test, as presented in the Table 3.8, provide an insight on the likely outcome of the top in–sample models. The test at lag one rejects the null–hypothesis of no serial autocorrelation at lag one and the subsequent tests at higher lags also reject the null–hypothesis. It is therefore expected that the top in–sample models take advantage of a conditional volatility process. The expectation is confirmed, since the top eight in–sample models, as presented in the Table 3.9, all model volatility with either the EGARCH or the TARCH process. The estimated parameters of the model

	Test statistic value	p-value
In–sample before model application		
Portmanteau Q(1)	78.6563	0.0000
Portmanteau Q(2)	109.4573	0.0000
Portmanteau Q(3)	172.0104	0.0000
Portmanteau Q(5)	329.7434	0.0000
Portmanteau Q(10)	466.1323	0.0000
In–sample after model application		
Portmanteau Q(1)	0.0609	0.8052
Portmanteau Q(2)	0.2388	0.8874
Portmanteau Q(3)	0.2538	0.9685
Portmanteau Q(5)	0.3649	0.9962
Portmanteau Q(10)	4.1620	0.9397

Table 3.8: Serial autocorrelation tests for the FTSE index

with the highest AIC are presented in the Table C.4. The application of the EGARCH(1,1) process while dealing with the serial autocorrelation of squared residuals is able to remove the serial autocorrelation quite well. The Table 3.8 provides the results of the Portmanteau Q test. The serial autocorrelation is effectively captured, since the null–hypothesis of the Portmanteau Q test at the selected lags cannot be rejected even at 80.52% significance level in all cases.

In–sample model	AIC	LL
AR(2)-MA(2)-EGARCH(1,1)-T	-7249.064	3634.532
AR(2)-MA(1)-TARCH(2,2)-T	-7067.065	3545.533
AR(1)-MA(2)-TARCH(2,2)-T	-7067.033	3545.517
AR(2)-MA(2)-TARCH(2,1)-T	-7066.537	3545.268
AR(2)-MA(2)-TARCH(2,2)-T	-7066.464	3546.232
AR(2)-MA(1)-TARCH(1,1)-T	-7066.083	3542.042
AR(2)-MA(2)-TARCH(2,2)-GED	-7066.015	3546.007
AR(1)-MA(0)-TARCH(1,1)-T	-7065.887	3539.944

Table 3.9: Top eight in–sample models for the FTSE index

The top eight in–sample models are mostly the ones using the TARCH process for the conditional volatility process, with one exception that is represented by the EGARCH process. The parameters of the conditional volatility processes, p and q, range from one to two, which is in line with the outcomes of the Portmanteau Q test that rejects the null–hypothesis of no serial autocorrelation of the squared residuals at lags one and greater. It is, however, possible that the maximum order of $p = q = 2$ is not large enough to capture the underlying volatility process. Seven out of the eight models work under the assumption of the Student–t distribution for the error term and one is taking advantage of the GED distribution. As already mentioned, this is in line with the outcome of the Jarque–Bera test statistic, which rejects normality of the log–returns.

Although the out–of–sample subset of the FTSE index exhibits a value of skewness closer to zero, the kurtosis is higher. The value of the Jarque–Bera test statistic is even higher than in the in–sample subset, which means that the out–of–sample subset is not normally distributed. The variance of the log–returns is also significantly higher. It is therefore expected that the order of the best performing models might be quite different from order of the in–sample models. The models are evaluated at three confidence intervals $\alpha \in \{0.90, 0.95, 0.99\}$ and the best performing out–of–sample models are summarized in the tables A.24, A.25 and A.26.

Evaluation at $\alpha = 0.90$

The top out–of–sample models for the FTSE index at the $\alpha = 0.90$ confidence interval are all working under the assumption of the Student–t distribution with 11 degrees of freedom for the error term. Quite good results achieve also models that assume the normal distribution for the error term, since they account for almost half of the models in the top 50. The conditional volatility process is among the top eight models modeled five times using the GARCH(1,1) process, two models use the GARCH(2,1) process and one model the GARCH(2,2) process.

The failure rates at this confidence interval are similar to the failure rates of the IXIC index, since the failure rates range from 13.0% to 13.4%. According to the p–values of the unconditional coverage, the expected failure rate of 10% for the one–day–ahead VaR forecast cannot be rejected at 3.17% at the highest. This indicates that even the top eight out–of–sample models are not able to capture the behavior of the log–returns very well. The p–values of the conditional coverage select three models with the same p–value $p_{cc} = 0.0979$, which means that the null–hypothesis cannot be rejected at the significance level 9.79%. These results are better than the results of the DJI, GSPC and IXIC indices. The values of the loss functions further distinguish the three models. Based on their values, the best performing model for the FTSE index at the confidence interval of $\alpha = 0.90$ is the AR(1)-MA(2)-GARCH(1,1)-T model.

Evaluation at $\alpha = 0.95$

The results of the models evaluation on the out–of–sample data at this confidence interval provide mixed results. Even though all models assume the Student–t distribution for the error term, the degrees of freedom parameter is estimated to be equal to 12, as opposed to 11 for the previous confidence interval. The models indicate a presence of an autoregressive process in the conditional volatility process and the order of the process ranges from one to two.

Looking at the failure rates, it is obvious that the models are not able to predict the one–day–ahead VaR very accurately. The failure rates range from 7.8% to 8.0%, which is quite large. Therefore the unconditional coverage test rejects the null–hypothesis of the failure rate equal to 5% even at the 1% significance level. It is interesting to note that the p–values are similar to the p–values of the IXIC index at the same confidence interval. On the other hand, the p–values of the conditional coverage test are higher and four models achieve the same value of 2.87%. When the models are further on compared based on the values of their loss functions, there is one model with the lowest achieved value in two of the three selected loss functions. Therefore, the AR(0)-MA(2)-GARCH(2,1)-T model seems to be the most adequate one for the FTSE index at the $\alpha = 0.95$ confidence interval.

Evaluation at $\alpha = 0.99$

As well as in the previous confidence interval, the top eight models for the confidence interval $\alpha = 0.95$ differ in the orders of the underlying conditional volatility process. The Table A.26 provides a list of these models together with the values of the selection criteria. The best results are achieve using the Student–t distribution for the error term with 11 degrees of freedom. Models assuming the normal distribution begin to appear in the results table staring on the 35^{th} position.

In comparison with the other five indices, the failure rates for the FTSE index at the $\alpha = 0.99$ confidence interval are the highest. Even the best performing models are not able to forecast the one–day–ahead VaR with a failure rate lower than 2.6%. To be precise, the failure rates range from 2.6% to 2.8%. This is the reason why none of the models attains the p–value of the unconditional coverage greater than 1%, which means that the null–hypothesis of the failure rate being equal to 1% is rejected for all models at the 1% significance level. The same applies to the values of the conditional coverage, as the highest achieved value is equal to 0.8%. Nevertheless, to continue with the same procedure as with the previous indices, the values of the loss functions provide the final decision on the models performance. The lowest loss functions values are achieved with the AR(1)-MA(2)-GARCH(1,2)-T model, however, the model is still rejected at the significance level of 1%.

Note

The results of the application of the models on the out–of–sample subset do not provide satisfactory results. The reason is the same as for the IXIC index. The in–sample subset is quite different from the out–of–sample subset and the selected models are not able to accurately capture the difference. This is apparent in the $\alpha = 0.99$ confidence interval, where all models are rejected at 1% significance level. The AR(0)-MA(2)-GARCH(2,1)-T model is present among the top eight models for all three confidence intervals, however, it is selected as the best performing one only for the $\alpha = 0.95$ confidence interval. Selecting this model as a general model for the FTSE index when the confidence interval is abstracted from, is thus discouraged.

The application of the top AIC and LL based models on the FTSE index out–of–sample data is not an optimal choice. None of the in–sample models is able to accurately forecast the one–day–ahead VaR. Interesting observation is that even for the FTSE index it is optimal to employ models that take advantage of a GARCH conditional volatility process, as such models provide the best forecasts. It is important to note, however, that even those models do not provide acceptable forecasts. Thus it is obvious that a simple estimation of a model in an in–sample subset does not ensure an adequate forecasting accuracy on the out–of–sample subset.

3.2.6 GDAXI index

The log–returns histogram of the GDAXI index is presented on the Figure B.3, such as the other indices. From the histogram it is apparent that the distributions suffers from a quite strong fat–tail on the side of the negative log–returns. The kurtosis seems to be well above the kurtosis of the normal distribution.

The detailed analysis of the in–sample subset in the Table A.1 confirms that both the kurtosis and skewness deviate from the values of these properties in the normal distribution. The Jarque–Bera test statistic rejects the null–hypothesis of normally distributed log–returns at the 1% significance level. It is therefore not surprising that the top in–sample models are not the ones that assume normal distribution for the error term. The order of the in–sample models is obtained in the same way as for the previous indices — the models are ranked according to their AIC and LL values.

	Test statistic value	p-value
In–sample before model application		
Portmanteau Q(1)	3.3375	0.0677
Portmanteau Q(2)	26.4240	0.0000
Portmanteau Q(3)	49.6646	0.0000
Portmanteau Q(5)	72.8182	0.0000
Portmanteau Q(10)	131.1777	0.0000
In–sample after model application		
Portmanteau Q(1)	0.5895	0.4426
Portmanteau Q(2)	0.6652	0.7170
Portmanteau Q(3)	0.7002	0.8731
Portmanteau Q(5)	1.2781	0.9372
Portmanteau Q(10)	6.7858	0.7455

Table 3.10: Serial autocorrelation tests for the GDAXI index

The Table 3.10 provides the results of the Portmanteau Q test for serial autocorrelation of the squared residuals for the GDAXI index. The test cannot reject the null–hypothesis of no serial autocorrelation at lag one at 5% significance level. On the other hand, when the test is applied on lags greater than two, the value of the test statistic rejects the null–hypothesis. The interpretation of the test results is that the best in–sample models most likely contain a conditional volatility process with order two. It might be possible that the actual order of the serial autocorrelation is greater than two, however, the thesis works with conditional volatility processes of a maximum order $p = q = 2$, so a higher order serial autocorrelation will not be estimated for. The Table 3.10 provides the results of the Portmanteau Q test after the application of the conditional volatility process TARCH(2,2). The serial autocorrelation is effectively captured, since the null–hypothesis of the Portmanteau Q test at the selected lags cannot be rejected even at 44.26% significance level in all cases.

In–sample model	AIC	LL
AR(2)-MA(2)-TARCH(2,2)-GED	-6632.773	3329.387
AR(2)-MA(2)-TARCH(2,1)-GED	-6631.912	3327.956
AR(2)-MA(2)-TARCH(1,1)-GED	-6631.805	3325.902
AR(1)-MA(1)-TARCH(2,2)-GED	-6630.230	3326.115
AR(1)-MA(0)-TARCH(2,2)-GED	-6629.695	3324.848
AR(0)-MA(1)-TARCH(2,2)-GED	-6629.664	3324.832
AR(1)-MA(2)-TARCH(2,2)-GED	-6628.365	3326.182
AR(2)-MA(1)-TARCH(2,2)-GED	-6628.364	3326.182

Table 3.11: Top eight in–sample models for the GDAXI index

As listed in the Table 3.11, all of the top performing models in terms of their AIC value take advantage of the conditional volatility modeled as a TARCH process. Interesting observation is that the models work exclusively with the GED distribution as the distribution for the error term. The shape parameter of the distribution is estimated to be approximately 1.5. As it is suggested by the Portmanteau Q test, the order of the autoregressive part of the TARCH process is equal to two in seven out of eight cases. There is one model though that models the conditional volatility as a TARCH(1,1) process. The estimated parameters of the model with the highest AIC are presented in the Table C.5.

The Table A.2 with the properties of the out–of–sample subset provides an insight on the shape of the out–of–sample log–returns. Both the skewness and kurtosis differ significantly and so does the volatility of the subset. Based on the results of the evaluation of the previous indices, it is expected that the top in–sample models might not score among the top out–of–sample models, simply due to the changed properties of the subset. The models are again evaluated at three confidence intervals $\alpha \in \{0.90, 0.95, 0.99\}$ and the best performing out–of–sample models are summarized in the tables A.27, A.28 and A.29.

Evaluation at $\alpha = 0.90$

The top eight out–of–sample models for the GDAXI index at the $\alpha = 0.90$ confidence interval assume the Student–t distribution for the error term. The estimated degrees of freedom parameter lies around 8. There does not seem to be present any specific process for the conditional volatility, as the top eight models employ four different conditional volatility process. On the other hand, the presence of the GARCH process indicates that the decision to model the conditional volatility is not out of question.

The failure rates for this confidence interval range from 12.2% to 13.2%. It is not surprising that the best p–values for the unconditional coverage are achieved by the models with the failure rate equal to 12.2%. Quite interesting is the fact that these three models are outperformed by the models with the highest failure rates in terms of the conditional coverage p–value. The reason lies on less independently distributed failures for the two models. To continue with the evaluation, the values of the loss functions distinguish between the two models with the highest p–values of the conditional coverage and select the AR(1)-MA(2)-GARCH(2,2)-T model as the model with the best performance for the GDAXI index at the $\alpha = 0.90$ confidence interval. It is important to note, however, that there are models with slightly lower p–values of the conditional coverage, but also with lower values of the loss functions.

Evaluation at $\alpha = 0.95$

The results of the application of the models at the $\alpha = 0.95$ confidence interval are not satisfactory, as it is described further on. The selected models employ the Student–t distribution for the error term with 8 degrees of freedom. The conditional volatility process is modeled by a GARCH processes with multiple orders. None of the EGARCH or TARCH processes appears among the top eight results for this confidence interval.

The one–day–ahead VaR forecasts are violated approximately in 8.0% to 8.2% cases, which is quite consistent. On the other hand, the rates are still well above the expected failure rate of 5%. Therefore, the unconditional coverage test rejects the null–hypothesis at the 1% significance level for all models. The p–values of the conditional coverage test are also low and the test rejects the null–hypothesis also at the 1% significance level. The values of the loss functions select the AR(1)-MA(2)-GARCH(1,1)-T model as the one with the lowest realized loss and highest p–value of the conditional coverage test for the GDAXI index at the $\alpha = 0.95$ confidence interval.

Evaluation at $\alpha = 0.99$

The models at the $\alpha = 0.99$ confidence interval work with the assumption of the error term distributed according to the Student–t distribution with 8 degrees of freedom. The conditional volatility processes is modeled using the GARCH(1,1) process in seven cases and with the GARCH(2,2) process in the eight case. The other tested conditional volatility process — EGARCH and

TARCH — do not appear among the top eight models. The same outcome is observed at the other indices as well.

An interesting observation at the $\alpha = 0.99$ confidence interval is that all models achieve the same failure rate of 1.6%. Considering the expected failure rate of 1%, it is not an unsatisfactory result. This is the reason why the top eight models share the same p–value of the unconditional coverage and even of the conditional coverage. The p–values do not reject the models at 40% significance level. Therefore, the models must be ranked according to the values of the loss functions. The preferred model for the GDAXI index at the $\alpha = 0.99$ confidence interval is thus the AR(1)-MA(1)-GARCH(1,2)-T.

Note

None of the models appears among the top eight performing models for all confidence intervals. Moreover, the evaluation of the models at the $\alpha = 0.95$ confidence interval does not provide satisfactory results, since all of the models are rejected by both the unconditional and conditional coverage tests. Therefore it cannot be confirmed that there is a one particular model that is able to capture the behavior of the GDAXI index at the best.

Although the in–sample estimation suggest to use models that take advantage of the TARCH process with GED as the distribution for the error term, the application of these models on the out–of–sample subset provides quite different results. The best relative accuracy is achieved with models that model conditional volatility with the GARCH process and the Student–t distribution. Thus the conclusion is that even the GDAXI index confirms that top in–sample performance does not automatically mean top out–of–sample performance.

3.2.7 N225 index

The Figure B.3 with the histogram of the log–returns of the N225 index indicates that there is a fat–tail on the side of the negative log–returns. This is confirmed by the detailed statistics in the in–sample summary Table A.1. The kurtosis of the log–returns also seems to be above the value of 3, which is the kurtosis of the normal distribution and it is confirmed by the statistics as well. In order to test for the normality of the log–returns, the Jarque–Bera test is employed. The test rejects the null–hypothesis of normality at the 1% significance level. Even though the normality is rejected, two of the top eight in–sample models work with the assumption of normally distributed error term, as listed in the Table 3.13.

	Test statistic value	p-value
In–sample before model application		
Portmanteau Q(1)	7.5726	0.0059
Portmanteau Q(2)	21.1184	0.0000
Portmanteau Q(3)	43.3973	0.0000
Portmanteau Q(5)	68.3412	0.0000
Portmanteau Q(10)	93.9835	0.0000
In–sample after model application		
Portmanteau Q(1)	0.0569	0.8115
Portmanteau Q(2)	3.0480	0.2178
Portmanteau Q(3)	3.3156	0.3455
Portmanteau Q(5)	7.8548	0.1644
Portmanteau Q(10)	15.5731	0.1125

Table 3.12: Serial autocorrelation tests for the N225 index

The Portmanteau Q test for serial autocorrelation of the squared residuals at lag one rejects the null–hypothesis of no serial autocorrelation at 1% significance level. The results are provided in the Table 3.12. The tests for serial autocorrelation at higher orders also reject the null–hypothesis, which leads to the conclusion that the conditional volatility should be modeled using one of the available conditional volatility processes. The Table 3.13 with the top in–sample models proves that it is the case, since all models take advantage of the EGARCH process for conditional volatility. The estimated parameters of the model with the highest AIC are presented in the Table C.6. The Table 3.12 proves that after the application of the conditional volatility process EGARCH(2,1), the serial autocorrelation is effectively captured. The null–hypothesis of the Portmanteau Q test at the selected lags cannot be rejected even at the 11.25% significance level in all cases.

Even though the conditional volatility process, EGARCH, is the same for all top in–sample models, its parameters differ and so does the assumed distribution of the error term. Five models assume the EGARCH(2,2) process, while the other three assume the EGARCH(2,1) process. Quite interesting observation is that there are two models that work with the normal distribution, three models that work with the GED distribution and three with the Student–t distribution. The shape parameter for the GED distribution is approximately 1.6

In–sample model	AIC	LL
AR(0)-MA(1)-EGARCH(2,1)-T	-35779.33	17898.67
AR(2)-MA(1)-EGARCH(2,1)-T	-6470.151	3246.075
AR(1)-MA(1)-EGARCH(2,2)-T	-6396.750	3209.375
AR(1)-MA(0)-EGARCH(2,2)-GED	-6310.031	3165.015
AR(0)-MA(2)-EGARCH(2,2)-GED	-6308.577	3165.288
AR(1)-MA(0)-EGARCH(2,2)-N	-6304.140	3161.070
AR(1)-MA(2)-EGARCH(2,2)-N	-6301.853	3161.927
AR(1)-MA(2)-EGARCH(2,1)-GED	-6285.056	3153.528

Table 3.13: Top eight in–sample models for the N225 index

and the degrees of freedom parameter for the Student–t distribution is greater than 10^{13}.

The skewness of out–of–sample subset is very similar to the skewness of the in–sample subset. On the other hand, the kurtosis is significantly higher. Moreover, the volatility on Figure B.2 is a clear indicator that the out–of–sample subset of the N225 index is more turbulent. Therefore the top performing in–sample models might most likely fail to capture the change in the parameters of the subset and it is expected that the normal and GED distributions shall be outperformed by the Student–t distribution, as it is the case with the previously analyzes indices. The models are evaluated at three confidence intervals $\alpha \in \{0.90, 0.95, 0.99\}$ and the best performing out–of–sample models are summarized in the tables A.30, A.31 and A.32.

Evaluation at $\alpha = 0.90$

Neither the N225 index is an exception in the assumed distribution of the error term. The best results at this confidence interval are achieved with the Student–t distribution with estimated 11 degrees of freedom. This is an indication of quite fat–tails of the log–returns. The only conditional volatility process among the top eight out–of–sample models is the GARCH process. The parameters of the process vary and are presented in the Table A.30.

The failure rate ranges from 12.6% to 13.0%. In comparison with the other tested indices, these failure rates are not among the best one, neither are they among the worst ones. According to the p–values of the unconditional coverage, the null–hypothesis of the failure rates being equal to 10% cannot be rejected at 6.14% at the best. The p–values of the conditional coverage are quite similar to the values of the unconditional coverage and according to them, the null–hypothesis cannot be rejected at 4.19%. The values of the loss functions indicate that the model with the lowest achieved values of the loss functions is the AR(1)-MA(1)-GARCH(2,1)-T model. It is also one of the models with the failure rate closest to the expected value of 10%.

Evaluation at $\alpha = 0.95$

The top eight out–of–sample models at the $\alpha = 0.95$ confidence interval employ the Student–t distribution with estimated 10 degrees of freedom for the error term. The degrees of freedom parameter is estimated using the statistical software during the evaluation of the models. The conditional volatility processes are modeled using the GARCH process, such as in the previous confidence interval. Interesting observation is that models assuming the normal distribution occupy the last third of the top 50 models.

The failure rates for the one–day–ahead VaR at the $\alpha = 0.95$ confidence interval are the lowest failure rates from the selected indices, as they range from 6.2% to 6.6%. This fact is also transformed to the p–values of the unconditional coverage, since the highest achieved p–value is $p_{uc} = 23.46\%$. Moreover, even the other models in the top eight cannot be rejected even at 10% significance level. The values of the conditional coverage are also quite satisfactory. Based on these values, there is only one model with the highest score, namely the AR(2)-MA(1)-GARCH(1,1)-T model. This model has also the highest realized p–value of the conditional coverage of all tested models and all indices.

Evaluation at $\alpha = 0.99$

The Student–t distribution with 10 degrees of freedom as the assumed distribution for the error term provides the best results for the N225 index at the $\alpha = 0.99$ confidence interval. The conditional volatility process is in all eight models modeled exclusively by the GARCH process. The EGARCH and TARCH processes do not score even among the top 50 models. It is quite interesting that the GARCH process dominates the EGARCH process, when the models are applied on the out–of–sample subset.

The failure rates are equal for all of the top eight models, which is the same situation as with the GDAXI index. The realized failure rate is equal to 1.4% as opposed to the expected rate of 1%. The difference is captured by the p–value of the unconditional coverage test. The p–value of the unconditional coverage test suggest that the null–hypothesis cannot be rejected 39.66% significance level for all models. Since even the p–values of the conditional coverage are the same and attain a value of 64.11%, the decision on which model is the most appropriate one for the N225 index at the $\alpha = 0.99$ confidence interval is based on the realized values of the loss functions. The lowest values were achieved by the AR(2)-MA(1)-GARCH(1,1)-T model.

Note

None of the models in the top eight models for the particular confidence interval appear among the results for all confidence intervals. Therefore it cannot be suggested to use one model as a general model for the N225 index when the confidence interval is abstracted from. Interesting observation occurs in the $\alpha = 0.95$ confidence interval, where the closest match to the expected failure rate of 5% is achieved among all indices.

The EGARCH process for conditional volatility is suggested by the in–sample estimation procedure. However, the application of the estimated models on the out–of–sample subset indicates that there are other models that are able to achieve better VaR forecasting accuracy. As well as for the other indices, the best accuracy is achieved when the conditional volatility is modeled using the GARCH process. It is quite interesting that the GARCH process offers the best performance in all indices. The likely reason lies in the fact that the GARCH process treats both positive and negative shocks in the same way and that the situation on the markets during the out–of–sample period probably exhibited this exact behavior.

3.2.8 Summary

This section described the evaluation of 648 models based on the ARMA process for modeling conditional mean and GARCH family of processes for modeling conditional volatility. Models parameters were first estimated on the in–sample subset, which consisted from data observed during the years 2004–2007, and top eight models according to their AIC values were selected. The top AIC value models were mostly EGARCH or TARCH based with either Student–t or the GED as the most appropriate distribution for the error term. These estimates proved that models that work with conditional volatility modeled as an asymmetric process provided more accurate and significant estimates than typically used models with a symmetric GARCH process. On the other hand, orders of AR and MA processes did not seem to be of particular importance, as their orders quite varied and seemed to play a minor role in the actual models specification. The implication of this result is that markets treat volatility of a particular stock index as a dominant factor, rather than the previously realized value of the stock index. For each stock index there were several models that failed to be estimated by the statistical software. The number of failed models, however, was quite small.

Quite interesting outcome of the analysis was the fact that the conditional volatility of in–sample subsets was best modeled with either the EGARCH or the TARCH process and in the majority of cases the Student–t distribution with approximately 7–8 degrees of freedom. The interpretation of this outcome is that markets treat positive shocks in a different manner than negative shocks. Considering the fact that the in–sample subsets were less volatile than the out–of–sample subsets, it seems that in periods with relatively stable volatility, asymmetric conditional volatility processes provide better estimates than symmetric conditional volatility processes. This is, on the other hand, not true in periods with relatively higher volatility, as it is described further on.

All estimated models were then applied on the out–of–sample subset with data from years 2008–2009, in order to test the predictive accuracy of the one–day–ahead VaR. Mdels were not periodically re–estimated. For each estimated model there were 500 forecasts calculated and models were then compared based on values of the conditional coverage test. The advantage of applying the conditional coverage lies in the fact that it does not only test whether the number of extreme cases corresponds to the selected confidence interval, it also tests whether violations are clustered or not. This is an important factor in the evaluation of the one–day–ahead VaR, since the occurrence probability of a violation should not depend on the occurrence of previous violations. Therefore the evaluation used in the thesis should have ruled out models with clustered violations.

The models evaluation was performed for three most common confidence intervals, namely $\alpha \in \{0.90, 0.95, 0.99\}$. Results of the models application on the out–of–sample subset were quite different from in–sample results. All of the top scoring models were based on modeling conditional volatility using the GARCH process. Even though most models underestimated the VaR by some amount, as it is described in summary tables in appendices, the majority of models was not rejected based on the p–value of the conditional coverage test.

3.3 Model re-estimation

The models evaluated in the previous section suffered from the fact that the process of estimating their parameters was performed only once on the initial in–sample subset that contained 1000 observations and the estimated models were then used to predict the time series behavior for the next 500 observations. Moreover, as the aim of this work is to be able to tell which model has capability to most accurately simulate the real behavior of a time series in times of increased volatility, the absence of periodical re-estimation put further stress on the tested models.

When value–at–risk models are used in the financial sector, the models are being periodically re-estimated in order to adapt to the real market situation as precisely as possible. Therefore there is a need to re-estimate the models more often than just once. This section summarizes the analysis of the models introduced earlier with the difference that the models are re-estimated after 125 observations, which in the context of the available dataset represents four re-estimations within the range of 500 out–of–sample observations.

3.3.1 Data analysis

As well as in the previous section, each dataset includes exactly 1500 observations. However, the periodical re-estimation of the models is performed on a moving window of observations, where the in–sample subset has always 1000 observation and the out–of–sample subset has 125 observations immediately following the end of the in–sample subset. The beginning of the in–sample subset shifts by 125 observation for each re-estimation attempt. Thanks to such division, each model will be re-estimated four times within the available dataset.

For brevity the in–depth in–sample and out–of–sample dataset analysis is not described in detail in this section. The primary focus is put on the analysis of re-estimated models with respect to the smaller out–of–sample dataset sizes. The motivation for periodical re-estimation is to answer the question whether the reduction in size of the out–of–sample subset and models re-estimation lead to more consolidated outcomes compared to the analysis in the previous section. Further on, the analysis of re-estimated models is for all indices discussed for the confidence interval $\alpha = 0.95$.

3.3.2 DJI index

The results of the analysis of the DJI index in the previous section showed that the market can be at the best modeled using a GARCH process for the conditional volatility part of the model with four models having GARCH parameters equal to $p = q = 1$. Two other models actually showed tendency to model the index with parameters $p = q = 2$. Therefore it was not a clear, which model is a unique most adequate one for the time series. However, all best performing models used the Student–t distribution for modeling the error term.

Results of the DJI index analysis performed on smaller out–of–sample datasets and with periodically re-estimated parameters can be found in the Table A.33 in the Appendix A. Eight top performing models, based on the geometrical average of their values of conditional coverage, exhibit a strong tendency to the GARCH model with parameters $p = 1$ and $q = 1$. Six out of eight models share these parameters, while the other two differ in the parameter $q = 2$. It is therefore obvious that the overall performance of GARCH(1,1) based model was not rejected by periodical re-estimation and shorter out–of–sample size.

The results, however, revealed quite interesting information. Taking look at models' characteristics as represented by the failure rate and values of both conditional and unconditional coverage, it can be concluded that models did not perform very well on the data from the interval $0 - 125$, since all models suffered from almost twice as high failure rates than the expected failure rate of 5%. This observation is even more interesting given the fact that the out–of–sample subset exhibits only a slightly higher volatility than the in–sample subset, on which models were estimated. This result is naturally projected into values of both unconditional and conditional coverage that were calculated as $p_{uc} = 0.0351$ and $p_{cc} = 0.1084$ respectively.

The second re-estimation with benchmark data coming from the range $126-250$ is in most case better than in the first part, which is quite interesting given the fact that the second range contains the highest share of high volatility on the market. The most interesting outcome, however, comes from results of the third period. Models were estimated on high volatility in–sample subset, which included also the data from the range $126-250$ and this is most likely the reason why models after third re-estimation performed with the best relative results. Failure rates lie around the desired value of 5%, which consequently improves values of unconditional and conditional coverage. Finally, the fourth re-estimation with benchmark out–of–sample subset in the range $376 - 500$

exhibits the lowest failure rate for all models. Even though low failure rates might seem desirable, the opposite is true, since models are supposed the fail with certain rate and a departure in either direction decreases the predictive potential of the model i.e. increases uncertainty in its behavior.

To conclude the analysis of the periodical model re-estimation on the DJI index, the following results should be noted. The re-estimation did not shatter the leading performance of the GARCH conditional volatility process with parameters $p = q = 1$ as the most appropriate conditional volatility process for the DJI index. The results also showed that the Student–t distribution is an appropriate distribution function for the error term.

3.3.3 GSPC index

The results of the analysis for the GSPC index in the previous section suggested that the time series exhibited an autoregressive process in the form of the GARCH process with parameters $p = 2$ and $q = 1$, since this was the case in six out of eight best performing models. As well as in the case of the DJI index, the Student–t distribution of the error term was the preferred distribution.

Periodically re-estimated models tested on the GSPC index are summarized in the Table A.34, where eight models with the best overall performance measured as a geometrical average of their conditional coverage, are listed together with their parameters and statistics. The results seem to prefer the GARCH process for conditional volatility, however, it is not clear which parameters deliver the best performance. Five out of eight models worked with parameters $p = 1$ and $q = 2$, which was not observed in the previous analysis of the large out–of–sample. The other three models prefer parameters $p = q = 1$. All of the models provided best results when working with Student–t as the error term distribution just as in the previous analysis.

The results exhibit quite similar tendency as the DJI index to lower failure rates with increasing shift in the position on the in–sample and out–of–sample subsets. Failure rates in the first range of observations $0 - 125$ are almost twice as high as it was expected according to the model setup with respect to the confidence interval of $\alpha = 0.95$. The unsatisfactory results are projected into the values of the unconditional and conditional coverage, which attain quite low values $p_{uc} = 0.0351$ and $p_{cc} = 0.0335$. The second subset shows only a minor improvement in the failure rate, however, the failure rate still remains quite high.

The third re-estimation finally provided more promising results than previous two estimations. The failure rate lies approximately around the desired failure rate of 5% as required by the selected confidence interval. Even though failure rates are above the 5% threshold, values of both the unconditional and conditional coverage improve dramatically. It is safe to assume that the presence of data with higher volatility in the third and fourth re-estimation process led to the adaption of models to the volatility in the time series. The adaption went well, since after the fourth re-estimation, models achieved the highest accuracy so far. The actual failure rate was in six cases only by 0.2% points lower than expected, which is a great result. Values of the unconditional coverage confirm that the violations were close to the expected failure rate and the conditional coverage supports the fact that the violations were not clustered, since $p_{uc} = 0.9178$ and $p_{cc} = 0.7330$.

The analysis with periodical re-estimation showed that in the case of the GSPC index it is recommended to model the time series using the GARCH model with parameters $p = q = 1$. The preferred distribution for the error term is the Student–t distribution, just as in the analysis in the previous section.

3.3.4 IXIC index

In the analysis of the IXIC index in the previous section, it has been detected that the time series suffers from clustered violations and conditional volatility. The analysis and application of models over the entire out–of–sample subset showed that the time series should be modeled using the GARCH process with parameters $p = q = 1$ in six out of eight cases and with parameters $p = q = 2$ in two remaining cases. The IXIC index is therefore not different from other indices. The same statement can be said about the distribution of the error term that was at the best modeled using the Student–t distribution.

The Table A.35 lists top eight models with the best overall performance as geometrically averaged over the four re-estimation periods. From the results it is obvious that the GARCH process with parameters $p = q = 1$ is the most adequate conditional volatility process for the IXIC time series. From the results, however, it is not clear what is the role of both autoregressive and moving average processes in the mean of the time series. Since parameters of the ARMA processes do not show any preferred combination of parameters and even the values of conditional coverage do not seem to help in discovering the answer, the significance of the ARMA process is likely insignificant.

When individual out–of–sample results are analyzed, it provides quite similar outcomes to outcomes of two previous indices. Failure rates tend to decrease with the advance in the out–of–sample subset. Nevertheless, results of the second part of the out–of–sample subset, represented by the range $126 - 250$, exhibit worse failure rates than results in the range $0 - 125$. In comparison to previous indices, failure rates move in the opposite direction and reach almost twice as high failure rate than the expected failure rate of 5%. The explanation lies in the fact that models applied on the second range of out–of–sample data were estimated on data that were not as volatile.

The predictive accuracy of models in last two ranges of out–of–sample data indicate that re-estimations were able to accurately grasp the underlying process in the time series, since failure rates tend to lie around the expected failure rate of 5%. Relatively successful estimates are also supported by values of both unconditional and conditional coverage, since both p_{uc} and p_{cc} values rose sharply and reached values of 0.7330 and 0.9178 respectively.

The conclusion for the IXIC index is quite straightforward. The time series suffered from a conditional volatility process, which was successfully removed using the GARCH process with parameters $p = q = 1$. Moreover, distribution of the error term random variable can be described at the best by the Student–t distribution, just as in the case of two previous indices. The implication of this result is that the error term is not normally distributed and there are fat tails. Therefore the Student–t distribution is recommended.

3.3.5 FTSE index

The analysis of the FTSE index based on the entire out–of–sample subset was not clear about the dominant process for conditional volatility. Even though conditional volatility process has been detected, the best performing models were in four cases based on the GARCH model with parameters $p = q = 1$, however, other four models had each a different combination of parameters. The distribution of the error term was by the analysis suggested as Student–t.

After the application of models on partial out–of–sample subsets, results seem to be more concise. As it can be seen from the Table A.36, top eight overall performing models are still modeled using the GARCH process with parameters $p = q = 1$ in four cases and the other four cases have consolidated to the GARCH process with parameters $p = 2$ and $q = 1$. It is quite interesting that top eight overall performing models as suggested by the analysis, all share the same failure rate, the same unconditional coverage and the same conditional coverage. It is therefore obvious that models behave consistently over the analyzed out–of–sample subset.

It should be noted that the performance of all eight models declined in the second range (observations $126 - 250$), as the failure rate increased by 0.8% points from the failure rate in the first range with out–of–sample data. The increase in failure rate was caused by the increase in volatility in the second out–of–sample range, which models were not able to capture successfully. The third period provides results that are the closest ones to desired results. Values of unconditional coverage reached the highest values for the FTSE index, when they all reached the value $p_{uc} = 0.4904$.

The fourth re-estimation caused models to decrease the failure rate even more to as low level as 3.2%. Unfortunately, such low failure rate is not desirable, since models are supposed to fail with 5% probability. Therefore, the value of the unconditional coverage test has decreased. On the other hand, the value of conditional coverage has improved relatively to previous applications and reached the highest level for the FTSE index, equal to $p_{cc} = 0.5383$.

The periodical re-estimation performed on the FTSE index suggested that the time series should be modeled with a consideration of including a conditional volatility process. Two top performing conditional volatility processes were both GARCH processes with two different sets of parameters. Four models were using parameters $p = q = 1$ and the other four models $p = 2$ and $q = 1$. These two sets of parameters had the same effect on failure rates and on values of both unconditional and conditional coverage. The FTSE index behavior is not different from the other indices in the sense that the recommended distribution for the error term is the Student–t distribution.

3.3.6 GDAXI index

Results of the application of various models on the GDAXI index suggested, that the time series can be modeled with the highest accuracy with the GARCH process. In other words, the time series contains conditional volatility process. The suggested GARCH process is in seven out of eight top performing cases estimated with parameters $p = q = 1$. The recommended distribution for the error term is the Student–t distribution, just as by the other considered indices.

When the out–of–sample subset was further on divided into four parts and periodically re-estimated, results turned to be quite different from results presented in the previous section. From the Table A.37 with top overall performing models it is obvious that the periodical re-estimation had a significant effect on models performance. The new analysis indicated that the actual conditional volatility process should be GARCH with parameters $p = 1$ and $q = 2$, since models that took advantage of this process performed overall very well.

The predictive accuracy of models in the first part of the out–of–sample dataset was approximately equal to the performance of models in the previous section. Even though failure rates were not twice as high as it was the case with previously analyzed indices, the failure rate still reached 8%, which is well over the threshold of 5%. Values of both unconditional and conditional coverage reflect high failure rates in their values as well.

The second part of the out–of–sample subset performed even worse than the first part, since failure rates reached 8.8%. Naturally conditional coverage value is quite small, nevertheless it is still better than values realized on the entire out–of–sample subset in the previous section. This is a result of better model adaption on the current market situation in the case of periodical re-estimation.

A positive observation can be deducted from a closer look at results of the application on the third and the fourth part of the out–of–sample subset. Failure rates lie very close to the desired value of 5% and also values of conditional coverage tests suggest that re-estimated models were successfully able to capture the underlying behavior of the GDAXI time series. To put results in numbers, the failure rate in the fourth part of the out–of–sample subset reached 4.8% and values of unconditional coverage and conditional coverage reached $p_{uc} = 0.9178$, $p_{cc} = 0.7330$ respectively. It is also worth noting that the GARCH process with parameters $p = 1$ and $q = 2$ yields better results than the same process with parameters $p = q = 2$ or parameters $p = 2$ and $q = 1$.

From the analysis it is clear that the GDAXI index benefits from the periodical re-estimation. Even though models suggested by the periodical re-estimation differ from models suggested in the previous section, the important benchmark is the predictive accuracy of VaR, which has improved by periodical re-estimation. On the other hand, both analysis provided better results when using the Student–t distribution as the preferred distribution for the error term.

3.3.7 N225 index

In the case of the N225 index, the previous analysis could not suggest a single model that would dominate other models in performance. Although all models took advantage of the GARCH process for modeling conditional volatility, parameters of the GARCH process varied. However, all models had one common point and that was the suggestion to use the Student–t distribution for the distribution of the error term.

Results of the re-estimation after certain periods in time, as presented in the Table A.38, indicate that in the case of the N225 index the re-estimation improves the quality of the VaR application. With the exception of the second part of the out–of–sample subset, where the failure rate reached 8%, its results are better than results from the analysis in the previous section. Failure rates in the first part of the out–of–sample ranged from 5.6% to 6.4%, which is better than the 6.6% achieved in the previous section. Also values of both unconditional and conditional coverage reach quite satisfactory values.

Just as it was the case with all other five preceding indices, also for the N225 index it holds that the further the out–of–sample subset begins, the better the forecasting accuracy of the estimated models. From the table with results it can be observed that the failure rate of 5.6% in the third part of the subset and 4.8% in the fourth part of the subset are very good results. This is also confirmed by values of both tests from the conditional coverage framework. The unconditional coverage reached value $p_{uc} = 0.9178$ and the value of the conditional coverage reached $p_{cc} = 0.5343$. This is quite an improvement compared to values $p_{uc} = 0.2346$ and $p_{cc} = 0.3775$ achieved by the analysis in the previous section.

The analysis showed that periodical estimation has positive effect on the accuracy of models applied on the N225 index. The overall performance measured by the geometrical average of the unconditional and conditional coverage has increased dramatically and failure rates narrowed to the desired level of 5% as required by the chosen confidence interval. As well as other indices, the N225 index suffers from fat tails, which consequently ranks the Student–t distribution as the most appropriate distribution for the error term from three considered distribution functions.

3.3.8 Summary

Now that the analysis of the periodically re-estimated models has been completed, it is appropriate to summarize achieved results. This section described outcomes of periodical re-estimation on a divided out–of–sample subset with the aim to compare the overall model performance and predictive accuracy with respect to the analysis performed on the entire out–of–sample in the previous section. Each model has been estimated on 1000 in–sample observations and tested on immediately following 125 observations. This process has been repeated four times, since the entire out–of–sample subset had 500 observations in total. The models performance was compared based on the geometrical average of their unconditional and conditional coverage, which is able to capture both failure rate and clustering of violations.

Results showed several interesting outcomes. Firstly, quite often were selected top performing models with same parameters as models in the previous analysis of the entire out–of–sample. This leads to a conclusion that the underlying conditional variance process has been successfully captured by both approaches. Mean modeling processes were, however, not clearly identified, just as it was the case in the previous analysis. One possible explanation is that the mean process is not of much importance while calculating VaR on selected stock indices.

The second interesting outcome is that models improved their accuracy over time. Results indicate that the first estimation performed on data with relatively low volatility was not very successful when applied on data with relatively high volatility. However, as the data used for re-estimation of models contained more observations with high volatility, the forecasting accuracy of models has increased. For all indices, results in the third and the fourth part of the out–of–sample subset seemed to be quite accurate thanks to the inclusion of period two, which contained most volatility in the entire out–of–sample subset.

The third outcome of the analysis is concerned with the failure rate in the fourth period. From the analysis it is obvious that when models are estimated on data with relatively high volatility and the models are then applied on data with relatively low volatility, models tend to fail in less occasions than expected. In other words, models are more conservative about volatility. This is actually the inverse of the situation on the second out–of–sample period, where models systematically underestimated volatility thanks to being estimated on data with low volatility.

To conclude, periodical re-estimation of VaR models based on a combination of ARMA-GARCH processes has a positive effect on the overall performance of models. It is, however, crucial to estimate models on a set of data which behave similarly to the expected behavior of the analyzed time series. Departure in either direction has the effect of either underestimating or overestimating the risk.

Chapter 4

Conclusion

The value–at–risk forecasting theory has been applied on six selected stock indices. Even though the most popular choice of stock indices includes either the indices from central and eastern European economies or the US–based indices, the thesis strived to provide a broader picture and based the analysis on other less often used indices. At the same time, these indices (DJI, GSPC, IXIC, FTSE, GDAXI, N225) were chosen as to incorporate indices from the most important world stock exchanges. The data for the one–day–ahead VaR forecast came from the years 2004 – 2007 for the in–sample subset, which serves the purpose of estimating the coefficients of the models, and the years 2008 – 2009 for the out–of–sample subset, which served both as a source of data with higher volatility and as a benchmark.

The decision to divide the dataset into two subsets was a logical step considering the aim of the thesis, which was to estimate various value–at–risk forecasting models in the times of a relative mild volatility and to reevaluate the same models based on data with higher volatility, such as during the recent financial crisis. The methodology used in the thesis was similar to the methodology commonly used by commercial banks with the exception that the models were not periodically re-estimated. Another important aspect of the thesis was that in comparison to the methodology used by commercial banks, the thesis worked with more advanced methods, such as the conditional volatility process.

Even though it is quite common in the literature to work with the assumption of normally distributed log–returns, the thesis did not adhered to such simplification and tested the log–returns for a variety of distributions including the popular Student–t and the GED distributions. The results proved that the log–returns of the selected indices cannot be considered as distributed according to the normal distribution. In fact, the more appropriate distribution seemed to be the Student–t distribution with 7 to 8 degrees of freedom or the GED with the shape parameter lying between 1.4 – 1.5. Thus the hypothesis of normally distributed log–returns has been rejected at 5% significance level for all selected indices and the decision to reject the null–hypothesis of normality was based on the values of the Jarque–Bera test statistic and checked against the histogram plots. Therefore in addition to the commonly assumed normal distribution, the applied VaR models were parametrized with the Student–t and the GED distribution functions, which proved to result in more accurate estimates.

One of the most significant aspects of the thesis was the variability of employed models, as the thesis attempted to estimate 648 dynamic models. Moreover, all of the models were estimated on the six considered indices and evaluated separately. The large number of tested models was a result of the dynamic approach followed by the thesis, since the models were composed from two components — the conditional mean, which was modeled with the AR and the MA processes with up to two lags, and the conditional variance, which was modeled using one of the GARCH, EGARCH or TARCH processes with up to two lags. Moreover, each model was estimated for the three afore mentioned distributions. The decision to include conditional volatility process was backed by the results of the Portmanteau Q test statistic at several lags.

Although many authors suggest using the GARCH(1,1) process for modeling the conditional volatility of stock indices, the thesis did not work with such apriori assumption and tested a vast number of models. From the analysis it stems that the suggestion of using GARCH(1,1) process for the conditional volatility process while predicting VaR could not be rejected. Each one of the suggested out–of–sample models took advantage of the GARCH process, however, the parameters of the GARCH process varied. It is therefore an interesting outcome that even though the asymmetric models achieved the best results in terms of the AIC and LL values, the actual forecast capabilities were dominated by a symmetric conditional volatility process. A likely interpretation is that in times of increased volatility, the markets treat both positive and negative shocks in the same way, which is quite interesting. This outcome therefore confirms the hypothesis that an asymmetric EGARCH is outperformed by a GARCH based model when the magnitude of shocks, volatility, is high.

Therefore the suggestion to forecast VaR only with models that have the highest value of the AIC was not backed by the results of the thesis. None of the models with the highest in–sample AIC values was among the top out–of–sample models. The results obtained in the thesis suggest to take advantage of the commonly used GARCH process, which saves a lot of computation time and provides satisfactory results. In order to improve the VaR forecasts, the most obvious step was to periodically re-estimate the models. Another analysis was therefore performed on the same set of data, however, with four periodical reestimations. The reestimation confirmed the dominant role of the GARCH(1,1) process, which was suggested by the analysis earlier.

To conclude, the thesis loosely followed a typical one–day–ahead VaR evaluation procedure with a number of improvements introduced into the procedure. Such improvements included a no apriori assumption on the distribution of the log–returns, which proved to be a step in the right direction. Further on, the thesis estimated a quite large number of models that allowed to compare the models with various conditional mean and conditional volatility processes, as well as with three distribution functions for the error term. The final part of the evaluation took advantage of a less known framework that is used to measure the accuracy of the forecasted models. Thanks to this, the thesis was able to provide a new insight on the topic, which certainly belongs to the most discussed topics in the financial sector.

Bibliography

Akgiray, V. (1989): "Conditional heteroscedasticity in time series of stock returns: Evidence and forecasts." *The Journal of Business* **62**: pp. 55–80.

Alexander, C. (2009): *Market Risk Analysis, Value at Risk Models*, volume 4. Wiley, 1 edition.

Angelidis, T., A. Benos, & S. Degiannakis (2004): "The use of garch models in var estimation." *Statistical Methodology* **1(1-2)**: pp. 105 – 128.

Audrino, F. & P. Buhlmann (2001): "Tree-structured generalized autoregressive conditional heteroscedastic models." *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **63(4)**: pp. 727–744.

Black, F. (1976): "Studies of stock market volatility changes." *Proceedings of the American Statistical Association, Business and Economic Statistics Section* pp. 177–181.

Bollerslev, T. (1986): "Generalized autoregressive conditional heteroskedasticity." *Journal of Econometrics* **31**: pp. 307–327.

Bollerslev, T., R. F. Engle, & D. B. Nelson (1994): "Chapter 49 arch models." volume 4 of *Handbook of Econometrics*, pp. 2959 – 3038. Elsevier.

Brooks, C. & G. Persand (2003): "The effect of asymmetries on stock index return value-at-risk estimates." *The Journal of Risk Finance* **4**: pp. 29–42.

Christoffersen, P. F. (1998): "Evaluating interval forecasts." *International Economic Review* **39(4)**: pp. 841–862.

Davidian, M. & R. J. Carroll (1987): "Variance function estimation." *Journal of the American Statistical Association* **82(400)**: pp. 1079–1091.

DICKEY, D. A. & W. A. FULLER (1979): "Distribution of the estimators for autoregressive time series with a unit root." *Journal of the American Statistical Association* **74(366)**: pp. pp. 427–431.

DIEBOLD, F. & R. S. MARIANO (2002): "Comparing predictive accuracy." *Journal of Business & Economic Statistics* **20(1)**: pp. 134–44.

DROST, F. C. & C. A. J. KLAASSEN (1997): "Efficient estimation in semiparametric garch models." *Journal of Econometrics* **81(1)**: pp. 193 – 221.

ENGLE, R. F. (1982): "Autogregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation." *Econometrica* **50**: pp. 987–1008.

ENGLE, R. F. & S. MANGANELLI (1999): "Caviar: Conditional value at risk by quantile regression." *NBER Working Papers 7341*, National Bureau of Economic Research, Inc.

ENGLE, R. F. & S. MANGANELLI (2004): "Caviar: Conditional autoregressive value at risk by regression quantiles." *Journal of Business and Economic Statistics* **22**: pp. 367–381.

ENGLE, R. F. & V. K. NG (1993): "Measuring and testing the impact of news on volatility." *Journal Of Finance* **48**: pp. 1749–1778.

GHAHRAMANI, M. & A. THAVANESWARAN (2008): "A note on garch model identification." *Computers & Mathematics with Applications* **55(11)**: pp. 2469 – 2475.

GHYSELS, E., A. C. HARVEY, & E. RENAULT (1996): "5 stochastic volatility." In G. MADDALA & C. RAO (editors), "Statistical Methods in Finance," volume 14 of *Handbook of Statistics*, pp. 119 – 191. Elsevier.

GUERMAT, C. & R. D. F. HARRIS (2002): "Forecasting value at risk allowing for time variation in the variance and kurtosis of portfolio returns." *International Journal of Forecasting* **18(3)**: pp. 409 – 419.

HE, C., T. TERÄSVIRTA, & H. MALMSTEN (2002): "Moment structure of a family of first-order exponential garch models." *Econometric Theory* **18**: pp. 868–885.

Holton, G. A. (2003): *Value at Risk: Theory and Practice*, volume 1. Academic Press.

Huang, A. Y. (2010): "An optimization process in value-at-risk estimation." *Review of Financial Economics* **19(3)**: pp. 109–116.

Jorion, P. (2007): *Value at risk: The New Benchmark for Managing Financial Risk*. New York, NY [u.a.]: McGraw-Hill, 3. ed. edition.

Kupiec, P. H. (1995): "Techniques for verifying the accuracy of risk measurement models." *Finance and Economics Discussion Series 95-24*, Board of Governors of the Federal Reserve System (U.S.).

Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, & Y. Shin (1992): "Testing the null hypothesis of stationarity against the alternative of a unit root : How sure are we that economic time series have a unit root?" *Journal of Econometrics* **54(1-3)**: pp. 159 – 178.

Lambert, P. & S. Laurent (2001): *Modelling financial time series using GARCH-type models and a skewed student density.* Master's thesis, Universite de Liege.

Li, C. W. & W. K. Li (1996): "On a double-threshold autoregressive heteroscedastic time series model." *Journal of Applied Econometrics* **11(3)**: pp. 253–274.

Ljung, G. M. & G. E. P. Box (1978): "On a measure of lack of fit in time series models." *Biometrika* **65(2)**: pp. 297–303.

Lopez, J. A. (1998): "Methods for evaluating value-at-risk estimates." *Economic Policy Review* **(Oct)**: pp. 119–124.

Lopez, J. A. (2001): "Evaluating the predictive accuracy of volatility models." *Journal of Forecasting* **20(2)**: pp. 87–109.

Malmsten, H. & T. Teräsvirta (2004): "Stylized facts of financial time series and three popular models of volatility." *SSE/EFI working Paper Series in Economics and Finance* **563**.

Mandelbrot, B. (1963): "The variation of certain speculative prices." *The Journal of Business* **36**: pp. 394–419.

MANDELBROT, B. (1967): "The variation of some other speculative prices." *The Journal of Business* **40**: pp. 393–413.

NELSON, D. B. (1991): "Conditional heteroskedasticity in asset returns: A new approach." *Econometrica* **59**: pp. 347–370.

PAGAN, A. R. & G. W. SCHWERT (1990): "Alternative models for conditional stock volatility." *Journal of Econometrics* **45(1-2)**: pp. 267 – 290.

PATTON, A. (2006): "Volatility forecast comparison using imperfect volatility proxies." *Research Paper Series 175*, Quantitative Finance Research Centre, University of Technology, Sydney.

PHILLIPS, P. C. B. & P. PERRON (1988): "Testing for a unit root in time series regression." *Biometrika* **75(2)**: pp. pp. 335–346.

RABEMANANJARA, R. & J.-M. ZAKOÏAN (1993): "Threshold arch models and asymmetries in volatility." *Journal of Applied Econometrics* **8**: pp. 31–49.

SADORSKY, P. (2006): "Modeling and forecasting petroleum futures volatility." *Energy Economics* **28(4)**: pp. 467 – 488.

SAID, S. E. & D. A. DICKEY (1984): "Testing for unit roots in autoregressive-moving average models of unknown order." *Biometrika* **71(3)**: pp. pp. 599–607.

TERÄSVIRTA, T. (2006): "An introduction to univariate garch models." *Working papers in Economics and Finance* **646**: p. 30.

WEI, Y., Y. WANG, & D. HUANG (2010): "Forecasting crude oil market volatility: Further evidence using garch-class models." *Energy Economics* **32(6)**: pp. 1477 – 1484.

WOOLDRIDGE, J. (2008): *Introductory Econometrics: A Modern Approach (with Economic Applications, Data Sets, Student Solutions Manual Printed Access Card)*. South-Western College Pub, 4 edition.

ZAKOIAN, J.-M. (1994): "Threshold heteroskedastic models." *Journal of Economic Dynamics and Control* **18(5)**: pp. 931 – 955.

Appendix A

Tables

	DJI	GSPC	IXIC	FTSE	GDAXI	N225
Observations	1000	1000	1000	1000	1000	1000
Minimum	-0.0334876	-0.0353427	-0.0393587	-0.0418503	-0.0351624	-0.0556955
Maximum	0.0252232	0.0287896	0.0340557	0.0344408	0.0260508	0.0360312
Median	0.0004412	0.0007385	0.0010444	0.0008293	0.0012235	0.0004349
Mean	0.0002377	0.0002692	0.0002397	0.0003638	0.0006870	0.0003891
Variance	0.0000527	0.0000580	0.0000938	0.0000644	0.0000870	0.0001263
Standard deviation	0.0072608	0.0076173	0.0096851	0.0080266	0.0093260	0.0112390
Skewness	-0.3311214	-0.3099436	-0.1641089	-0.4295073	-0.4018695	-0.3622378
Kurtosis	4.5036750	4.8043270	3.4968970	5.7610460	3.7716550	4.3441480
Jarque–Bera	79.466	121.596	8.261	283.256	12.058	58.351
p–value	0.000	0.000	0.016	0.000	0.002	0.000

Table A.1: Properties of the in–sample subsets (log–returns)

	DJI	GSPC	IXIC	FTSE	GDAXI	N225
Observations	500	500	500	500	500	500
Minimum	-0.0820051	-0.0946951	0.0958770	-0.0926454	-0.0743346	-0.1211102
Maximum	0.1050835	0.1095720	0.1115944	0.0938424	0.1079747	0.1323459
Median	0.0001969	0.0009805	0.0006197	-0.0002057	0.0000000	0.0004258
Mean	-0.0004739	-0.0005406	-0.0003116	-0.0003972	-0.0006591	-0.0006850
Variance	0.0004043	0.0004873	0.0005005	0.0003932	0.0004489	0.0005737
Standard deviation	0.0201077	0.0220739	0.0223708	0.0198284	0.0211864	0.0239522
Skewness	0.1293708	-0.1143677	-0.0654192	-0.0006206	0.2978910	-0.3142532
Kurtosis	7.5003250	7.2374020	6.3187880	7.1332600	7.4221750	7.9021570
Jarque–Bera	419.003	372.217	228.834	355.043	402.522	493.053
p–value	0.000	0.000	0.000	0.000	0.000	0.000

Table A.2: Properties of the out–of–sample subsets (log–returns)

lags	DF-GLS	1% Critical value	5% Critical value	10% Critical value
10	-8.119	-3.480	-2.845	-2.558
9	-8.650	-3.480	-2.846	-2.559
8	-9.663	-3.480	-2.848	-2.560
7	-10.464	-3.480	-2.849	-2.562
6	-10.945	-3.480	-2.850	-2.563
5	-12.003	-3.480	-2.852	-2.564
4	-12.970	-3.480	-2.853	-2.565
3	-14.684	-3.480	-2.854	-2.567
2	-16.761	-3.480	-2.856	-2.568
1	-21.911	-3.480	-2.857	-2.569

Table A.3: In–sample DF-GLS stationarity test for DJI

lags	DF-GLS	1% Critical value	5% Critical value	10% Critical value
10	-6.482	-3.480	-2.845	-2.558
9	-7.028	-3.480	-2.846	-2.559
8	-8.030	-3.480	-2.848	-2.560
7	-8.890	-3.480	-2.849	-2.562
6	-9.491	-3.480	-2.850	-2.563
5	-10.488	-3.480	-2.852	-2.564
4	-11.702	-3.480	-2.853	-2.565
3	-13.692	-3.480	-2.854	-2.567
2	-15.798	-3.480	-2.856	-2.568
1	-21.321	-3.480	-2.857	-2.569

Table A.4: In–sample DF-GLS stationarity test for GSPC

lags	DF-GLS	1% Critical value	5% Critical value	10% Critical value
10	-4.085	-3.480	-2.845	-2.558
9	-4.563	-3.480	-2.846	-2.559
8	-5.209	-3.480	-2.848	-2.560
7	-5.923	-3.480	-2.849	-2.562
6	-6.375	-3.480	-2.850	-2.563
5	-7.119	-3.480	-2.852	-2.564
4	-8.223	-3.480	-2.853	-2.565
3	-9.997	-3.480	-2.854	-2.567
2	-11.822	-3.480	-2.856	-2.568
1	-16.291	-3.480	-2.857	-2.569

Table A.5: In–sample DF-GLS stationarity test for IXIC

lags	DF-GLS	1% Critical value	5% Critical value	10% Critical value
10	-5.807	-3.480	-2.845	-2.558
9	-6.359	-3.480	-2.846	-2.559
8	-7.140	-3.480	-2.848	-2.560
7	-7.782	-3.480	-2.849	-2.562
6	-8.502	-3.480	-2.850	-2.563
5	-9.373	-3.480	-2.852	-2.564
4	-10.684	-3.480	-2.853	-2.565
3	-12.501	-3.480	-2.854	-2.567
2	-15.001	-3.480	-2.856	-2.568
1	-19.301	-3.480	-2.857	-2.569

Table A.6: In–sample DF-GLS stationarity test for FTSE

lags	DF-GLS	1% Critical value	5% Critical value	10% Critical value
10	-8.910	-3.480	-2.845	-2.558
9	-9.710	-3.480	-2.846	-2.559
8	-10.228	-3.480	-2.848	-2.560
7	-10.924	-3.480	-2.849	-2.562
6	-11.465	-3.480	-2.850	-2.563
5	-12.327	-3.480	-2.852	-2.564
4	-14.001	-3.480	-2.853	-2.565
3	-15.397	-3.480	-2.854	-2.567
2	-17.252	-3.480	-2.856	-2.568
1	-20.702	-3.480	-2.857	-2.569

Table A.7: In–sample DF-GLS stationarity test for GDAXI

lags	DF-GLS	1% Critical value	5% Critical value	10% Critical value
10	-2.391	-3.480	-2.845	-2.558
9	-2.526	-3.480	-2.846	-2.559
8	-2.796	-3.480	-2.848	-2.560
7	-2.951	-3.480	-2.849	-2.562
6	-3.159	-3.480	-2.850	-2.563
5	-3.560	-3.480	-2.852	-2.564
4	-3.947	-3.480	-2.853	-2.565
3	-4.978	-3.480	-2.854	-2.567
2	-5.951	-3.480	-2.856	-2.568
1	-8.548	-3.480	-2.857	-2.569

Table A.8: In–sample DF-GLS stationarity test for N225

lags	DF-GLS	1% Critical value	5% Critical value	10% Critical value
10	-4.037	-3.480	-2.849	-2.563
9	-4.138	-3.480	-2.852	-2.567
8	-4.718	-3.480	-2.856	-2.570
7	-5.134	-3.480	-2.859	-2.572
6	-6.097	-3.480	-2.862	-2.575
5	-6.562	-3.480	-2.865	-2.578
4	-7.716	-3.480	-2.868	-2.581
3	-8.931	-3.480	-2.871	-2.583
2	-10.769	-3.480	-2.874	-2.586
1	-16.370	-3.480	-2.876	-2.588

Table A.9: Out–of–sample DF-GLS stationarity test for DJI

lags	DF-GLS	1% Critical value	5% Critical value	10% Critical value
10	-4.805	-3.480	-2.849	-2.563
9	-4.826	-3.480	-2.852	-2.567
8	-5.488	-3.480	-2.856	-2.570
7	-5.899	-3.480	-2.859	-2.572
6	-6.898	-3.480	-2.862	-2.575
5	-7.415	-3.480	-2.865	-2.578
4	-8.682	-3.480	-2.868	-2.581
3	-10.004	-3.480	-2.871	-2.583
2	-11.870	-3.480	-2.874	-2.586
1	-17.421	-3.480	-2.876	-2.588

Table A.10: Out–of–sample DF-GLS stationarity test for GSPC

lags	DF-GLS	1% Critical value	5% Critical value	10% Critical value
10	-4.481	-3.480	-2.849	-2.563
9	-4.590	-3.480	-2.852	-2.567
8	-5.095	-3.480	-2.856	-2.570
7	-5.518	-3.480	-2.859	-2.572
6	-6.333	-3.480	-2.862	-2.575
5	-7.081	-3.480	-2.865	-2.578
4	-8.232	-3.480	-2.868	-2.581
3	-9.627	-3.480	-2.871	-2.583
2	-11.267	-3.480	-2.874	-2.586
1	-16.545	-3.480	-2.876	-2.588

Table A.11: Out–of–sample DF-GLS stationarity test for IXIC

lags	DF-GLS	1% Critical value	5% Critical value	10% Critical value
10	-6.049	-3.480	-2.849	-2.563
9	-5.916	-3.480	-2.852	-2.567
8	-6.652	-3.480	-2.856	-2.570
7	-6.991	-3.480	-2.859	-2.572
6	-8.066	-3.480	-2.862	-2.575
5	-9.643	-3.480	-2.865	-2.578
4	-10.204	-3.480	-2.868	-2.581
3	-10.418	-3.480	-2.871	-2.583
2	-14.991	-3.480	-2.874	-2.586
1	-17.316	-3.480	-2.876	-2.588

Table A.12: Out–of–sample DF-GLS stationarity test for FTSE

lags	DF-GLS	1% Critical value	5% Critical value	10% Critical value
10	-4.971	-3.480	-2.849	-2.563
9	-5.294	-3.480	-2.852	-2.567
8	-6.031	-3.480	-2.856	-2.570
7	-6.211	-3.480	-2.859	-2.572
6	-6.725	-3.480	-2.862	-2.575
5	-7.238	-3.480	-2.865	-2.578
4	-8.571	-3.480	-2.868	-2.581
3	-9.260	-3.480	-2.871	-2.583
2	-12.573	-3.480	-2.874	-2.586
1	-15.386	-3.480	-2.876	-2.588

Table A.13: Out–of–sample DF-GLS stationarity test for GDAXI

lags	DF-GLS	1% Critical value	5% Critical value	10% Critical value
10	-2.227	-3.480	-2.849	-2.563
9	-2.529	-3.480	-2.852	-2.567
8	-2.926	-3.480	-2.856	-2.570
7	-3.036	-3.480	-2.859	-2.572
6	-3.497	-3.480	-2.862	-2.575
5	-4.240	-3.480	-2.865	-2.578
4	-5.002	-3.480	-2.868	-2.581
3	-5.752	-3.480	-2.871	-2.583
2	-7.754	-3.480	-2.874	-2.586
1	-9.965	-3.480	-2.876	-2.588

Table A.14: Out–of–sample DF-GLS stationarity test for N225

Model	LL	AIC	Failures	p_{uc}	p_{cc}	HMSE	HMAE	QLIKE
AR(1)-MA(2)-GARCH(1,1)-T	3564.386	-7112.772	12.4%	0.0834	0.0626	4.1841	1.2775	-7.2237
AR(2)-MA(1)-GARCH(1,1)-T	3564.409	-7112.817	12.4%	0.0834	0.0626	4.1908	1.2781	-7.2234
AR(2)-MA(1)-GARCH(2,1)-T	3565.928	-7113.856	12.4%	0.0834	0.0626	4.3819	1.2994	-7.2351
AR(1)-MA(0)-GARCH(1,1)-T	3562.726	-7113.452	12.6%	0.0614	0.0419	4.1958	1.2803	-7.2222
AR(2)-MA(0)-GARCH(2,1)-T	3565.285	-7114.571	12.6%	0.0614	0.0419	4.3731	1.2991	-7.2351
AR(1)-MA(2)-GARCH(2,1)-T	3565.911	-7113.821	12.6%	0.0614	0.0419	4.3733	1.2986	-7.2356
AR(2)-MA(2)-GARCH(2,1)-T	3565.939	-7111.878	12.6%	0.0614	0.0419	4.3946	1.3006	-7.2345
AR(1)-MA(0)-GARCH(2,1)-T	3564.423	-7114.846	12.6%	0.0614	0.0419	4.4387	1.3068	-7.2327

Table A.15: Best performing models for the DJI index at $\alpha = 0.9$

Model	LL	AIC	Failures	p_{uc}	p_{cc}	HMSE	HMAE	QLIKE
AR(1)-MA(0)-GARCH(2,2)-T	3564.402	-7112.804	7.0%	0.0523	0.0855	4.4548	1.3156	-7.2173
AR(0)-MA(1)-GARCH(2,2)-T	3564.453	-7112.905	7.0%	0.0523	0.0855	4.4573	1.3160	-7.2172
AR(1)-MA(0)-GARCH(1,1)-T	3562.726	-7113.452	7.2%	0.0337	0.0536	4.1958	1.2803	-7.2222
AR(0)-MA(1)-GARCH(1,1)-T	3562.785	-7113.571	7.2%	0.0337	0.0536	4.2015	1.2810	-7.2220
AR(2)-MA(1)-GARCH(2,1)-T	3565.928	-7113.856	7.2%	0.0337	0.0536	4.3819	1.2994	-7.2351
AR(2)-MA(2)-GARCH(2,1)-T	3565.939	-7111.878	7.2%	0.0337	0.0536	4.3946	1.3006	-7.2345
AR(1)-MA(2)-GARCH(1,1)-T	3564.386	-7112.772	7.4%	0.0211	0.0324	4.1841	1.2775	-7.2237
AR(2)-MA(1)-GARCH(1,1)-T	3564.409	-7112.817	7.4%	0.0211	0.0324	4.1908	1.2781	-7.2234

Table A.16: Best performing models for the DJI index at $\alpha = 0.95$

Model	LL	AIC	Failures	p_{uc}	p_{cc}	HMSE	HMAE	QLIKE
AR(1)-MA(1)-GARCH(1,1)-T	3564.297	-7114.593	1.0%	1.0000	0.9507	4.0868	1.2686	-7.2284
AR(0)-MA(2)-GARCH(1,1)-T	3563.622	-7113.244	1.0%	1.0000	0.9507	4.1526	1.2746	-7.2241
AR(2)-MA(0)-GARCH(1,1)-T	3563.636	-7113.273	1.0%	1.0000	0.9507	4.1632	1.2758	-7.2236
AR(1)-MA(2)-GARCH(1,1)-T	3564.386	-7112.772	1.0%	1.0000	0.9507	4.1841	1.2775	-7.2237
AR(2)-MA(1)-GARCH(1,1)-T	3564.409	-7112.817	1.0%	1.0000	0.9507	4.1908	1.2781	-7.2234
AR(1)-MA(0)-GARCH(1,1)-T	3562.726	-7113.452	1.0%	1.0000	0.9507	4.1958	1.2803	-7.2222
AR(2)-MA(2)-GARCH(1,1)-T	3564.421	-7110.842	1.0%	1.0000	0.9507	4.1986	1.2788	-7.2229
AR(0)-MA(1)-GARCH(1,1)-T	3562.785	-7113.571	1.0%	1.0000	0.9507	4.2015	1.2810	-7.2220

Table A.17: Best performing models for the DJI index at $\alpha = 0.99$

Model	LL	AIC	Failures	p_{uc}	p_{cc}	HMSE	HMAE	QLIKE
AR(2)-MA(2)-GARCH(1,1)-T	3533.573	-7049.146	12.0%	0.1471	0.0161	4.5987	1.2918	-7.0449
AR(1)-MA(2)-GARCH(1,1)-T	3532.033	-7048.066	12.0%	0.1471	0.0161	4.6963	1.2996	-7.0378
AR(2)-MA(1)-GARCH(1,1)-T	3531.902	-7047.805	12.0%	0.1471	0.0161	4.7054	1.3012	-7.0371
AR(1)-MA(1)-GARCH(2,2)-T	3537.306	-7056.612	12.6%	0.0614	0.0158	5.4108	1.4039	-7.0287
AR(0)-MA(2)-GARCH(2,2)-T	3536.429	-7054.857	13.2%	0.0223	0.0107	5.4762	1.4134	-7.0216
AR(2)-MA(0)-GARCH(1,1)-T	3531.308	-7048.616	12.2%	0.1116	0.0105	4.6663	1.2964	-7.0382
AR(1)-MA(1)-GARCH(1,2)-T	3536.402	-7056.804	12.2%	0.1116	0.0105	5.7563	1.4316	-7.0002
AR(1)-MA(1)-GARCH(1,1)-T	3532.643	-7051.286	12.4%	0.0834	0.0067	4.6086	1.2946	-7.0433

Table A.18: Best performing models for the GSPC index at $\alpha = 0.9$

Model	LL	AIC	Failures	p_{uc}	p_{cc}	HMSE	HMAE	QLIKE
AR(1)-MA(0)-GARCH(2,1)-T	3531.676	-7049.351	7.2%	0.0337	0.0069	4.9684	1.3413	-7.0450
AR(0)-MA(1)-GARCH(2,1)-T	3531.980	-7049.960	7.2%	0.0337	0.0069	5.0013	1.3437	-7.0440
AR(0)-MA(2)-GARCH(2,2)-T	3536.429	-7054.857	8.2%	0.0025	0.0030	5.4762	1.4134	-7.0216
AR(0)-MA(1)-GARCH(1,1)-T	3529.239	-7046.477	7.6%	0.0129	0.0022	4.6886	1.3046	-7.0364
AR(2)-MA(1)-GARCH(1,1)-T	3531.902	-7047.805	7.6%	0.0129	0.0022	4.7054	1.3012	-7.0371
AR(0)-MA(2)-GARCH(2,1)-T	3534.246	-7052.493	7.6%	0.0129	0.0022	4.8936	1.3284	-7.0489
AR(2)-MA(0)-GARCH(2,1)-T	3534.020	-7052.041	7.6%	0.0129	0.0022	4.9215	1.3320	-7.0471
AR(1)-MA(2)-GARCH(2,1)-T	3534.672	-7051.345	7.6%	0.0129	0.0022	4.9319	1.3327	-7.0473

Table A.19: Best performing models for the GSPC index at $\alpha = 0.95$

Model	LL	AIC	Failures	p_{uc}	p_{cc}	HMSE	HMAE	QLIKE
AR(1)-MA(0)-GARCH(1,1)-T	3528.987	-7045.974	1.2%	0.6630	0.8454	4.6752	1.3038	-7.0368
AR(0)-MA(1)-GARCH(1,1)-T	3529.239	-7046.477	1.2%	0.6630	0.8454	4.6886	1.3046	-7.0364
AR(2)-MA(2)-GARCH(2,1)-T	3535.996	-7051.992	1.2%	0.6630	0.8454	4.7781	1.3171	-7.0574
AR(1)-MA(1)-GARCH(2,1)-T	3535.184	-7054.368	1.2%	0.6630	0.8454	4.8441	1.3244	-7.0540
AR(0)-MA(2)-GARCH(2,1)-T	3534.246	-7052.493	1.2%	0.6630	0.8454	4.8936	1.3284	-7.0489
AR(2)-MA(0)-GARCH(2,1)-T	3534.020	-7052.041	1.2%	0.6630	0.8454	4.9215	1.3320	-7.0471
AR(1)-MA(2)-GARCH(2,1)-T	3534.672	-7051.345	1.2%	0.6630	0.8454	4.9319	1.3327	-7.0473
AR(2)-MA(1)-GARCH(2,1)-T	3534.537	-7051.073	1.2%	0.6630	0.8454	4.9452	1.3348	-7.0463

Table A.20: Best performing models for the GSPC index at $\alpha = 0.99$

Model	LL	AIC	Failures	p_{uc}	p_{cc}	HMSE	HMAE	QLIKE
AR(1)-MA(1)-GARCH(1,1)-T	3251.832	-6489.665	13.2%	0.0223	0.0036	4.5699	1.2338	-6.9376
AR(2)-MA(0)-GARCH(1,1)-T	3251.822	-6489.645	13.2%	0.0223	0.0036	4.5971	1.2332	-6.9365
AR(0)-MA(2)-GARCH(1,1)-T	3251.951	-6489.902	13.2%	0.0223	0.0036	4.5997	1.2337	-6.9367
AR(2)-MA(1)-GARCH(1,1)-T	3252.555	-6489.110	13.2%	0.0223	0.0036	4.6106	1.2343	-6.9365
AR(1)-MA(2)-GARCH(1,1)-T	3252.563	-6489.127	13.2%	0.0223	0.0036	4.6122	1.2345	-6.9366
AR(2)-MA(2)-GARCH(2,2)-T	3259.344	-6496.688	13.8%	0.0070	0.0022	5.3939	1.3671	-6.8973
AR(2)-MA(2)-GARCH(1,1)-T	3253.207	-6488.414	13.4%	0.0154	0.0021	4.5797	1.2351	-6.9390
AR(1)-MA(0)-GARCH(2,2)-T	3254.370	-6492.740	14.2%	0.0030	0.0018	5.6463	1.3124	-6.9053

Table A.21: Best performing models for the IXIC index at $\alpha = 0.9$

Model	LL	AIC	Failures	p_{uc}	p_{cc}	HMSE	HMAE	QLIKE
AR(1)-MA(0)-GARCH(1,1)-T	3250.677	-6489.353	7.8%	0.0077	0.0010	4.5522	1.2298	-6.9365
AR(0)-MA(1)-GARCH(1,1)-T	3250.682	-6489.365	7.8%	0.0077	0.0010	4.5534	1.2299	-6.9365
AR(1)-MA(0)-GARCH(2,2)-T	3254.370	-6492.740	7.8%	0.0077	0.0010	5.6463	1.3124	-6.9053
AR(0)-MA(1)-GARCH(2,2)-T	3254.400	-6492.799	7.8%	0.0077	0.0010	5.7314	1.3150	-6.9031
AR(1)-MA(1)-GARCH(1,1)-T	3251.832	-6489.665	8.0%	0.0045	0.0005	4.5699	1.2338	-6.9376
AR(2)-MA(0)-GARCH(1,1)-T	3251.822	-6489.645	8.0%	0.0045	0.0005	4.5971	1.2332	-6.9365
AR(0)-MA(2)-GARCH(1,1)-T	3251.951	-6489.902	8.0%	0.0045	0.0005	4.5997	1.2337	-6.9367
AR(2)-MA(1)-GARCH(1,1)-T	3252.555	-6489.110	8.0%	0.0045	0.0005	4.6106	1.2343	-6.9365

Table A.22: Best performing models for the IXIC index at $\alpha = 0.95$

Model	LL	AIC	Failures	p_{uc}	p_{cc}	HMSE	HMAE	QLIKE
AR(1)-MA(0)-GARCH(1,1)-T	3250.677	-6489.353	1.8%	0.1060	0.2296	4.5522	1.2298	-6.9365
AR(0)-MA(1)-GARCH(1,1)-T	3250.682	-6489.365	1.8%	0.1060	0.2296	4.5534	1.2299	-6.9365
AR(1)-MA(1)-GARCH(1,1)-T	3251.832	-6489.665	2.0%	0.0479	0.1152	4.5699	1.2338	-6.9376
AR(2)-MA(2)-GARCH(1,1)-T	3253.207	-6488.414	2.0%	0.0479	0.1152	4.5797	1.2351	-6.9390
AR(2)-MA(0)-GARCH(1,1)-T	3251.822	-6489.645	2.0%	0.0479	0.1152	4.5971	1.2332	-6.9365
AR(0)-MA(2)-GARCH(1,1)-T	3251.951	-6489.902	2.0%	0.0479	0.1152	4.5997	1.2337	-6.9367
AR(2)-MA(1)-GARCH(1,1)-T	3252.555	-6489.110	2.0%	0.0479	0.1152	4.6106	1.2343	-6.9365
AR(1)-MA(2)-GARCH(1,1)-T	3252.563	-6489.127	2.0%	0.0479	0.1152	4.6122	1.2345	-6.9366

Table A.23: Best performing models for the IXIC index at $\alpha = 0.99$

Model	LL	AIC	Failures	p_{uc}	p_{cc}	HMSE	HMAE	QLIKE
AR(1)-MA(2)-GARCH(1,1)-T	3530.794	-7045.588	13.0%	0.0317	0.0979	5.9519	1.3747	-7.0856
AR(2)-MA(2)-GARCH(1,1)-T	3530.817	-7043.635	13.0%	0.0317	0.0979	5.9612	1.3753	-7.0853
AR(0)-MA(2)-GARCH(1,1)-T	3525.995	-7037.990	13.0%	0.0317	0.0979	6.4857	1.4013	-7.0628
AR(0)-MA(2)-GARCH(2,1)-T	3526.516	-7037.032	13.2%	0.0223	0.0730	6.5590	1.4187	-7.0589
AR(2)-MA(0)-GARCH(2,1)-T	3526.504	-7037.009	13.2%	0.0223	0.0730	6.5670	1.4188	-7.0586
AR(2)-MA(0)-GARCH(1,1)-T	3525.990	-7037.981	13.2%	0.0223	0.0704	6.4919	1.4014	-7.0626
AR(2)-MA(1)-GARCH(1,1)-T	3526.030	-7036.060	13.2%	0.0223	0.0704	6.4925	1.4008	-7.0627
AR(0)-MA(2)-GARCH(2,2)-T	3526.715	-7035.430	13.4%	0.0154	0.0532	6.5694	1.4244	-7.0567

Table A.24: Best performing models for the FTSE index at $\alpha = 0.9$

Model	LL	AIC	Failures	p_{uc}	p_{cc}	HMSE	HMAE	QLIKE
AR(0)-MA(2)-GARCH(2,1)-T	3526.516	-7037.032	7.8%	0.0077	0.0287	6.5590	1.4187	-7.0589
AR(2)-MA(0)-GARCH(2,1)-T	3526.504	-7037.009	7.8%	0.0077	0.0287	6.5670	1.4188	-7.0586
AR(0)-MA(1)-GARCH(2,2)-T	3526.547	-7037.094	7.8%	0.0077	0.0287	6.5882	1.4242	-7.0565
AR(0)-MA(1)-GARCH(1,2)-T	3526.538	-7039.076	7.8%	0.0077	0.0287	6.5984	1.4246	-7.0559
AR(0)-MA(2)-GARCH(1,1)-T	3525.995	-7037.990	8.0%	0.0045	0.0175	6.4857	1.4013	-7.0628
AR(2)-MA(0)-GARCH(1,1)-T	3525.990	-7037.981	8.0%	0.0045	0.0175	6.4919	1.4014	-7.0626
AR(2)-MA(1)-GARCH(1,1)-T	3526.030	-7036.060	8.0%	0.0045	0.0175	6.4925	1.4008	-7.0627
AR(0)-MA(1)-GARCH(1,1)-T	3525.929	-7039.858	8.0%	0.0045	0.0175	6.5122	1.4023	-7.0621

Table A.25: Best performing models for the FTSE index at $\alpha = 0.95$

Model	LL	AIC	Failures	p_{uc}	p_{cc}	HMSE	HMAE	QLIKE
AR(1)-MA(2)-GARCH(1,2)-T	3531.426	-7044.851	2.6%	0.0027	0.0080	6.0272	1.3919	-7.0815
AR(2)-MA(2)-GARCH(1,2)-T	3531.429	-7042.857	2.6%	0.0027	0.0080	6.0290	1.3919	-7.0814
AR(2)-MA(1)-GARCH(1,2)-T	3531.428	-7044.857	2.6%	0.0027	0.0080	6.0301	1.3920	-7.0814
AR(2)-MA(1)-GARCH(2,2)-T	3531.487	-7042.974	2.6%	0.0027	0.0080	6.0324	1.3919	-7.0822
AR(1)-MA(2)-GARCH(2,1)-T	3531.435	-7044.871	2.6%	0.0027	0.0080	6.0587	1.3923	-7.0817
AR(2)-MA(2)-GARCH(2,1)-T	3531.436	-7042.872	2.6%	0.0027	0.0080	6.0597	1.3924	-7.0816
AR(1)-MA(2)-GARCH(1,1)-T	3530.794	-7045.588	2.8%	0.0009	0.0027	5.9519	1.3747	-7.0856
AR(2)-MA(2)-GARCH(1,1)-T	3530.817	-7043.635	2.8%	0.0009	0.0027	5.9612	1.3753	-7.0853

Table A.26: Best performing models for the FTSE index at $\alpha = 0.99$

Model	LL	AIC	Failures	p_{uc}	p_{cc}	HMSE	HMAE	QLIKE
AR(1)-MA(2)-GARCH(2,2)-T	3300.650	-6581.299	13.2%	0.0223	0.0417	8.3695	1.3811	-6.8956
AR(2)-MA(1)-GARCH(2,2)-T	3300.644	-6581.289	13.2%	0.0223	0.0417	8.3780	1.3814	-6.8954
AR(1)-MA(2)-GARCH(1,1)-T	3300.204	-6584.408	12.2%	0.1116	0.0374	8.1662	1.3620	-6.8950
AR(2)-MA(1)-GARCH(1,1)-T	3300.199	-6584.397	12.2%	0.1116	0.0374	8.1750	1.3622	-6.8949
AR(2)-MA(2)-GARCH(1,2)-T	3301.195	-6582.389	12.2%	0.1116	0.0374	8.4665	1.3857	-6.8896
AR(1)-MA(2)-GARCH(2,1)-T	3300.645	-6583.290	13.0%	0.0317	0.0370	8.3652	1.3784	-6.8965
AR(1)-MA(2)-GARCH(1,2)-T	3300.562	-6583.124	13.0%	0.0317	0.0370	8.3713	1.3863	-6.8908
AR(2)-MA(1)-GARCH(2,1)-T	3300.640	-6583.279	13.0%	0.0317	0.0370	8.3737	1.3787	-6.8964

Table A.27: Best performing models for the GDAXI index at $\alpha = 0.9$

Model	LL	AIC	Failures	p_{uc}	p_{cc}	HMSE	HMAE	QLIKE
AR(1)-MA(2)-GARCH(1,1)-T	3300.204	-6584.408	8.0%	0.0045	0.0054	8.1662	1.3620	-6.8950
AR(2)-MA(1)-GARCH(1,1)-T	3300.199	-6584.397	8.0%	0.0045	0.0054	8.1750	1.3622	-6.8949
AR(2)-MA(2)-GARCH(2,2)-T	3301.347	-6580.694	8.0%	0.0045	0.0054	8.4644	1.3799	-6.8956
AR(0)-MA(1)-GARCH(1,1)-T	3299.768	-6587.536	8.2%	0.0025	0.0028	8.0758	1.3555	-6.8978
AR(1)-MA(0)-GARCH(1,1)-T	3299.770	-6587.540	8.2%	0.0025	0.0028	8.0862	1.3558	-6.8976
AR(0)-MA(2)-GARCH(1,1)-T	3299.769	-6585.538	8.2%	0.0025	0.0028	8.0864	1.3558	-6.8976
AR(2)-MA(0)-GARCH(1,1)-T	3299.771	-6585.541	8.2%	0.0025	0.0028	8.0963	1.3561	-6.8974
AR(1)-MA(1)-GARCH(1,1)-T	3299.773	-6585.545	8.2%	0.0025	0.0028	8.1134	1.3566	-6.8971

Table A.28: Best performing models for the GDAXI index at $\alpha = 0.95$

Model	LL	AIC	Failures	p_{uc}	p_{cc}	HMSE	HMAE	QLIKE
AR(1)-MA(1)-GARCH(1,2)-T	3302.260	-6588.520	1.6%	0.2149	0.4068	7.5561	1.3678	-6.9077
AR(0)-MA(1)-GARCH(1,1)-T	3299.768	-6587.536	1.6%	0.2149	0.4068	8.0758	1.3555	-6.8978
AR(1)-MA(0)-GARCH(1,1)-T	3299.770	-6587.540	1.6%	0.2149	0.4068	8.0862	1.3558	-6.8976
AR(0)-MA(2)-GARCH(1,1)-T	3299.769	-6585.538	1.6%	0.2149	0.4068	8.0864	1.3558	-6.8976
AR(2)-MA(0)-GARCH(1,1)-T	3299.771	-6585.541	1.6%	0.2149	0.4068	8.0963	1.3561	-6.8974
AR(1)-MA(1)-GARCH(1,1)-T	3299.773	-6585.545	1.6%	0.2149	0.4068	8.1134	1.3566	-6.8971
AR(1)-MA(2)-GARCH(1,1)-T	3300.204	-6584.408	1.6%	0.2149	0.4068	8.1662	1.3620	-6.8950
AR(2)-MA(1)-GARCH(1,1)-T	3300.199	-6584.397	1.6%	0.2149	0.4068	8.1750	1.3622	-6.8949

Table A.29: Best performing models for the GDAXI index at $\alpha = 0.99$

Model	LL	AIC	Failures	p_{uc}	p_{cc}	HMSE	HMAE	QLIKE
AR(1)-MA(1)-GARCH(2,1)-T	3128.610	-6241.220	12.6%	0.0614	0.0419	3.5379	1.1337	-6.8480
AR(1)-MA(1)-GARCH(1,2)-T	3128.556	-6241.112	12.6%	0.0614	0.0419	3.5575	1.1362	-6.8454
AR(1)-MA(1)-GARCH(2,2)-T	3128.679	-6239.357	12.6%	0.0614	0.0419	3.5781	1.1361	-6.8467
AR(0)-MA(1)-GARCH(1,1)-T	3127.961	-6243.921	13.0%	0.0317	0.0370	3.4964	1.1379	-6.8454
AR(1)-MA(0)-GARCH(1,1)-T	3127.949	-6243.899	13.0%	0.0317	0.0370	3.4986	1.1381	-6.8454
AR(0)-MA(2)-GARCH(2,1)-T	3128.730	-6241.460	13.0%	0.0317	0.0370	3.5478	1.1388	-6.8492
AR(2)-MA(0)-GARCH(2,1)-T	3128.734	-6241.468	13.0%	0.0317	0.0370	3.5499	1.1389	-6.8491
AR(0)-MA(2)-GARCH(1,2)-T	3128.652	-6241.304	13.0%	0.0317	0.0370	3.5672	1.1417	-6.8463

Table A.30: Best performing models for the N225 index at $\alpha = 0.9$

Model	LL	AIC	Failures	p_{uc}	p_{cc}	HMSE	HMAE	QLIKE
AR(2)-MA(1)-GARCH(1,1)-T	3129.208	-6242.417	6.2%	0.2346	0.3775	3.3440	1.1209	-6.8485
AR(2)-MA(1)-GARCH(2,1)-T	3129.682	-6241.364	6.6%	0.1168	0.2913	3.4366	1.1256	-6.8505
AR(1)-MA(2)-GARCH(2,1)-T	3129.626	-6241.251	6.6%	0.1168	0.2913	3.4417	1.1262	-6.8505
AR(0)-MA(1)-GARCH(1,1)-T	3127.961	-6243.921	6.6%	0.1168	0.2913	3.4964	1.1379	-6.8454
AR(1)-MA(0)-GARCH(1,1)-T	3127.949	-6243.899	6.6%	0.1168	0.2913	3.4986	1.1381	-6.8454
AR(2)-MA(2)-GARCH(1,2)-T	3130.699	-6241.398	6.6%	0.1168	0.2913	3.5496	1.1374	-6.8457
AR(0)-MA(2)-GARCH(1,2)-T	3128.652	-6241.304	6.6%	0.1168	0.2913	3.5672	1.1417	-6.8463
AR(2)-MA(0)-GARCH(1,2)-T	3128.656	-6241.313	6.6%	0.1168	0.2913	3.5698	1.1418	-6.8462

Table A.31: Best performing models for the N225 index at $\alpha = 0.95$

Model	LL	AIC	Failures	p_{uc}	p_{cc}	HMSE	HMAE	QLIKE
AR(2)-MA(1)-GARCH(1,1)-T	3129.208	-6242.417	1.4%	0.3966	0.6411	3.3440	1.1209	-6.8485
AR(1)-MA(2)-GARCH(1,1)-T	3129.146	-6242.293	1.4%	0.3966	0.6411	3.3479	1.1215	-6.8484
AR(1)-MA(2)-GARCH(2,2)-T	3129.162	-6238.324	1.4%	0.3966	0.6411	3.3518	1.1212	-6.8481
AR(2)-MA(2)-GARCH(2,1)-T	3129.346	-6238.693	1.4%	0.3966	0.6411	3.3972	1.1252	-6.8541
AR(2)-MA(1)-GARCH(2,2)-T	3129.687	-6239.375	1.4%	0.3966	0.6411	3.4204	1.1213	-6.8488
AR(2)-MA(2)-GARCH(1,1)-T	3130.185	-6242.371	1.4%	0.3966	0.6411	3.4248	1.1294	-6.8466
AR(2)-MA(1)-GARCH(2,1)-T	3129.682	-6241.364	1.4%	0.3966	0.6411	3.4366	1.1256	-6.8505
AR(1)-MA(1)-GARCH(1,1)-T	3128.111	-6242.221	1.4%	0.3966	0.6411	3.4399	1.1290	-6.8460

Table A.32: Best performing models for the N225 index at $\alpha = 0.99$

Model	Out–of–sample	Failures	p_{uc}	p_{cc}
AR(2)-MA(1)-GARCH(1,1)-T	1 to 125	9.6%	0.0351	0.1084
	126 to 250	8.8%	0.0768	0.0715
	251 to 375	5.6%	0.7624	0.6283
	376 to 500	4.0%	0.5956	0.7039
AR(0)-MA(2)-GARCH(1,2)-T	1 to 125	11.2%	0.0058	0.0201
	126 to 250	7.2%	0.2882	0.2811
	251 to 375	5.6%	0.7624	0.6283
	376 to 500	4.8%	0.9177	0.7330
AR(0)-MA(1)-GARCH(1,1)-T	1 to 125	9.6%	0.0351	0.1084
	126 to 250	9.6%	0.0351	0.0299
	251 to 375	4.8%	0.9177	0.7330
	376 to 500	4.0%	0.5956	0.7039
AR(1)-MA(0)-GARCH(1,1)-T	1 to 125	9.6%	0.0351	0.1084
	126 to 250	9.6%	0.0351	0.0299
	251 to 375	4.8%	0.9177	0.7330
	376 to 500	4.0%	0.5956	0.7039
AR(1)-MA(2)-GARCH(1,1)-T	1 to 125	9.6%	0.0351	0.1084
	126 to 250	9.6%	0.0351	0.0299
	251 to 375	4.8%	0.9177	0.7330
	376 to 500	4.0%	0.5956	0.7039
AR(2)-MA(2)-GARCH(1,1)-T	1 to 125	9.6%	0.0351	0.1084
	126 to 250	9.6%	0.0351	0.0299
	251 to 375	4.8%	0.9177	0.7330
	376 to 500	4.0%	0.5956	0.7039
AR(0)-MA(1)-GARCH(1,2)-T	1 to 125	11.2%	0.0058	0.0201
	126 to 250	7.2%	0.2882	0.2811
	251 to 375	5.6%	0.7624	0.6283
	376 to 500	3.2%	0.3241	0.5382
AR(1)-MA(0)-GARCH(1,2)-T	1 to 125	11.2%	0.0058	0.0201
	126 to 250	7.2%	0.2882	0.2811
	251 to 375	5.6%	0.7624	0.6283
	376 to 500	3.2%	0.3241	0.5382

Table A.33: Top overall models for the DJI index at $\alpha = 0.95$

Model	Out–of–sample	Failures	p_{uc}	p_{cc}
AR(2)-MA(0)-GARCH(1,2)-T	1 to 125	9.6%	0.0351	0.0335
	126 to 250	8.8%	0.0769	0.0716
	251 to 375	5.6%	0.7625	0.6283
	376 to 500	4.8%	0.9178	0.7330
AR(1)-MA(1)-GARCH(1,2)-T	1 to 125	9.6%	0.0351	0.0335
	126 to 250	8.8%	0.0769	0.0716
	251 to 375	6.4%	0.4904	0.4538
	376 to 500	4.8%	0.9178	0.7330
AR(1)-MA(1)-GARCH(1,1)-T	1 to 125	9.6%	0.0351	0.0335
	126 to 250	9.6%	0.0351	0.0300
	251 to 375	5.6%	0.7625	0.6283
	376 to 500	4.8%	0.9178	0.7330
AR(0)-MA(1)-GARCH(1,2)-T	1 to 125	9.6%	0.0351	0.0335
	126 to 250	9.6%	0.0351	0.0300
	251 to 375	5.6%	0.7625	0.6283
	376 to 500	4.0%	0.5956	0.7040
AR(1)-MA(0)-GARCH(1,2)-T	1 to 125	9.6%	0.0351	0.0335
	126 to 250	9.6%	0.0351	0.0300
	251 to 375	5.6%	0.7625	0.6283
	376 to 500	4.0%	0.5956	0.7040
AR(2)-MA(2)-GARCH(1,2)-T	1 to 125	10.4%	0.0149	0.0127
	126 to 250	8.8%	0.0769	0.0716
	251 to 375	6.4%	0.4904	0.4538
	376 to 500	4.8%	0.9178	0.7330
AR(0)-MA(2)-GARCH(1,1)-T	1 to 125	10.4%	0.0149	0.0127
	126 to 250	9.6%	0.0351	0.0300
	251 to 375	5.6%	0.7625	0.6283
	376 to 500	4.8%	0.9178	0.7330
AR(1)-MA(2)-GARCH(1,1)-T	1 to 125	10.4%	0.0149	0.0127
	126 to 250	9.6%	0.0351	0.0300
	251 to 375	5.6%	0.7625	0.6283
	376 to 500	4.8%	0.9178	0.7330

Table A.34: Top overall models for the GSPC index at $\alpha = 0.95$

Model	Out–of–sample	Failures	p_{uc}	p_{cc}
AR(1)-MA(2)-GARCH(1,1)-T	1 to 125	8.8%	0.0769	0.0716
	126 to 250	9.6%	0.0351	0.0300
	251 to 375	6.4%	0.4904	0.4538
	376 to 500	4.8%	0.9178	0.7330
AR(2)-MA(1)-GARCH(1,1)-T	1 to 125	8.8%	0.0769	0.0716
	126 to 250	9.6%	0.0351	0.0300
	251 to 375	6.4%	0.4904	0.4538
	376 to 500	4.8%	0.9178	0.7330
AR(1)-MA(0)-GARCH(1,1)-T	1 to 125	8.8%	0.0769	0.0716
	126 to 250	10.4%	0.0149	0.0112
	251 to 375	5.6%	0.7625	0.6283
	376 to 500	4.0%	0.5956	0.7040
AR(0)-MA(1)-GARCH(1,1)-T	1 to 125	8.8%	0.0769	0.0716
	126 to 250	10.4%	0.0149	0.0112
	251 to 375	6.4%	0.4904	0.4538
	376 to 500	4.0%	0.5956	0.7040
AR(1)-MA(1)-GARCH(1,1)-T	1 to 125	8.8%	0.0769	0.0716
	126 to 250	10.4%	0.0149	0.0112
	251 to 375	6.4%	0.4904	0.4538
	376 to 500	4.0%	0.5956	0.7040
AR(0)-MA(2)-GARCH(1,1)-T	1 to 125	8.8%	0.0769	0.0716
	126 to 250	10.4%	0.0149	0.0112
	251 to 375	7.2%	0.2883	0.2811
	376 to 500	4.8%	0.9178	0.7330
AR(2)-MA(0)-GARCH(1,1)-T	1 to 125	8.8%	0.0769	0.0716
	126 to 250	10.4%	0.0149	0.0112
	251 to 375	7.2%	0.2883	0.2811
	376 to 500	4.8%	0.9178	0.7330
AR(2)-MA(2)-GARCH(1,1)-T	1 to 125	9.6%	0.0351	0.0300
	126 to 250	10.4%	0.0149	0.0112
	251 to 375	6.4%	0.4904	0.4538
	376 to 500	4.8%	0.9178	0.7330

Table A.35: Top overall models for the IXIC index at $\alpha = 0.95$

Model	Out–of–sample	Failures	p_{uc}	p_{cc}
AR(0)-MA(1)-GARCH(1,1)-T	1 to 125	7.2%	0.2883	0.3049
	126 to 250	8.0%	0.1553	0.0350
	251 to 375	6.4%	0.4904	0.4538
	376 to 500	3.2%	0.3242	0.5383
AR(0)-MA(1)-GARCH(2,1)-T	1 to 125	7.2%	0.2883	0.3049
	126 to 250	8.0%	0.1553	0.0350
	251 to 375	6.4%	0.4904	0.4538
	376 to 500	3.2%	0.3242	0.5383
AR(0)-MA(2)-GARCH(1,1)-T	1 to 125	7.2%	0.2883	0.3049
	126 to 250	8.0%	0.1553	0.0350
	251 to 375	6.4%	0.4904	0.4538
	376 to 500	3.2%	0.3242	0.5383
AR(0)-MA(2)-GARCH(2,1)-T	1 to 125	7.2%	0.2883	0.3049
	126 to 250	8.0%	0.1553	0.0350
	251 to 375	6.4%	0.4904	0.4538
	376 to 500	3.2%	0.3242	0.5383
AR(1)-MA(0)-GARCH(1,1)-T	1 to 125	7.2%	0.2883	0.3049
	126 to 250	8.0%	0.1553	0.0350
	251 to 375	6.4%	0.4904	0.4538
	376 to 500	3.2%	0.3242	0.5383
AR(1)-MA(0)-GARCH(2,1)-T	1 to 125	7.2%	0.2883	0.3049
	126 to 250	8.0%	0.1553	0.0350
	251 to 375	6.4%	0.4904	0.4538
	376 to 500	3.2%	0.3242	0.5383
AR(2)-MA(0)-GARCH(1,1)-T	1 to 125	7.2%	0.2883	0.3049
	126 to 250	8.0%	0.1553	0.0350
	251 to 375	6.4%	0.4904	0.4538
	376 to 500	3.2%	0.3242	0.5383
AR(2)-MA(0)-GARCH(2,1)-T	1 to 125	7.2%	0.2883	0.3049
	126 to 250	8.0%	0.1553	0.0350
	251 to 375	6.4%	0.4904	0.4538
	376 to 500	3.2%	0.3242	0.5383

Table A.36: Top overall models for the FTSE index at $\alpha = 0.95$

Model	Out–of–sample	Failures	p_{uc}	p_{cc}
AR(2)-MA(2)-GARCH(2,2)-T	1 to 125	7.2%	0.2883	0.2811
	126 to 250	8.8%	0.0769	0.2091
	251 to 375	5.6%	0.7625	0.6681
	376 to 500	4.0%	0.5956	0.7040
AR(0)-MA(1)-GARCH(1,2)-T	1 to 125	8.0%	0.1553	0.1513
	126 to 250	8.8%	0.0769	0.2091
	251 to 375	5.6%	0.7625	0.6681
	376 to 500	4.8%	0.9178	0.7330
AR(0)-MA(2)-GARCH(1,2)-T	1 to 125	8.0%	0.1553	0.1513
	126 to 250	8.8%	0.0769	0.2091
	251 to 375	5.6%	0.7625	0.6681
	376 to 500	4.8%	0.9178	0.7330
AR(1)-MA(0)-GARCH(1,2)-T	1 to 125	8.0%	0.1553	0.1513
	126 to 250	8.8%	0.0769	0.2091
	251 to 375	5.6%	0.7625	0.6681
	376 to 500	4.8%	0.9178	0.7330
AR(2)-MA(0)-GARCH(1,2)-T	1 to 125	8.0%	0.1553	0.1513
	126 to 250	8.8%	0.0769	0.2091
	251 to 375	5.6%	0.7625	0.6681
	376 to 500	4.8%	0.9178	0.7330
AR(2)-MA(2)-GARCH(1,2)-T	1 to 125	8.0%	0.1553	0.1513
	126 to 250	8.8%	0.0769	0.2091
	251 to 375	5.6%	0.7625	0.6283
	376 to 500	4.8%	0.9178	0.7330
AR(0)-MA(1)-GARCH(2,1)-T	1 to 125	8.0%	0.1553	0.1513
	126 to 250	8.8%	0.0769	0.2091
	251 to 375	5.6%	0.7625	0.6681
	376 to 500	5.6%	0.7625	0.6283
AR(0)-MA(1)-GARCH(2,2)-T	1 to 125	8.0%	0.1553	0.1513
	126 to 250	8.8%	0.0769	0.2091
	251 to 375	5.6%	0.7625	0.6681
	376 to 500	5.6%	0.7625	0.6283

Table A.37: Top overall models for the GDAXI index at $\alpha = 0.95$

Model	Out–of–sample	Failures	p_{uc}	p_{cc}
AR(0)-MA(1)-GARCH(2,1)-T	1 to 125	5.6%	0.7625	0.5809
	126 to 250	8.0%	0.1553	0.1800
	251 to 375	5.6%	0.7625	0.6283
	376 to 500	4.8%	0.9178	0.5343
AR(0)-MA(2)-GARCH(2,2)-T	1 to 125	5.6%	0.7625	0.5809
	126 to 250	8.0%	0.1553	0.1800
	251 to 375	5.6%	0.7625	0.6283
	376 to 500	4.8%	0.9178	0.5343
AR(1)-MA(0)-GARCH(2,1)-T	1 to 125	5.6%	0.7625	0.5809
	126 to 250	8.0%	0.1553	0.1800
	251 to 375	5.6%	0.7625	0.6283
	376 to 500	4.8%	0.9178	0.5343
AR(1)-MA(0)-GARCH(2,2)-T	1 to 125	5.6%	0.7625	0.5809
	126 to 250	8.0%	0.1553	0.1800
	251 to 375	5.6%	0.7625	0.6283
	376 to 500	4.8%	0.9178	0.5343
AR(0)-MA(2)-GARCH(2,1)-T	1 to 125	6.4%	0.4904	0.5887
	126 to 250	8.0%	0.1553	0.1800
	251 to 375	5.6%	0.7625	0.6283
	376 to 500	4.8%	0.9178	0.5343
AR(1)-MA(1)-GARCH(2,2)-T	1 to 125	6.4%	0.4904	0.5887
	126 to 250	8.0%	0.1553	0.1800
	251 to 375	5.6%	0.7625	0.6283
	376 to 500	4.8%	0.9178	0.5343
AR(1)-MA(2)-GARCH(2,2)-T	1 to 125	6.4%	0.4904	0.5887
	126 to 250	8.0%	0.1553	0.1800
	251 to 375	5.6%	0.7625	0.6283
	376 to 500	4.8%	0.9178	0.5343
AR(2)-MA(0)-GARCH(2,1)-T	1 to 125	6.4%	0.4904	0.5887
	126 to 250	8.0%	0.1553	0.1800
	251 to 375	5.6%	0.7625	0.6283
	376 to 500	4.8%	0.9178	0.5343

Table A.38: Top overall models for the N225 index at $\alpha = 0.95$

Appendix B

Figures

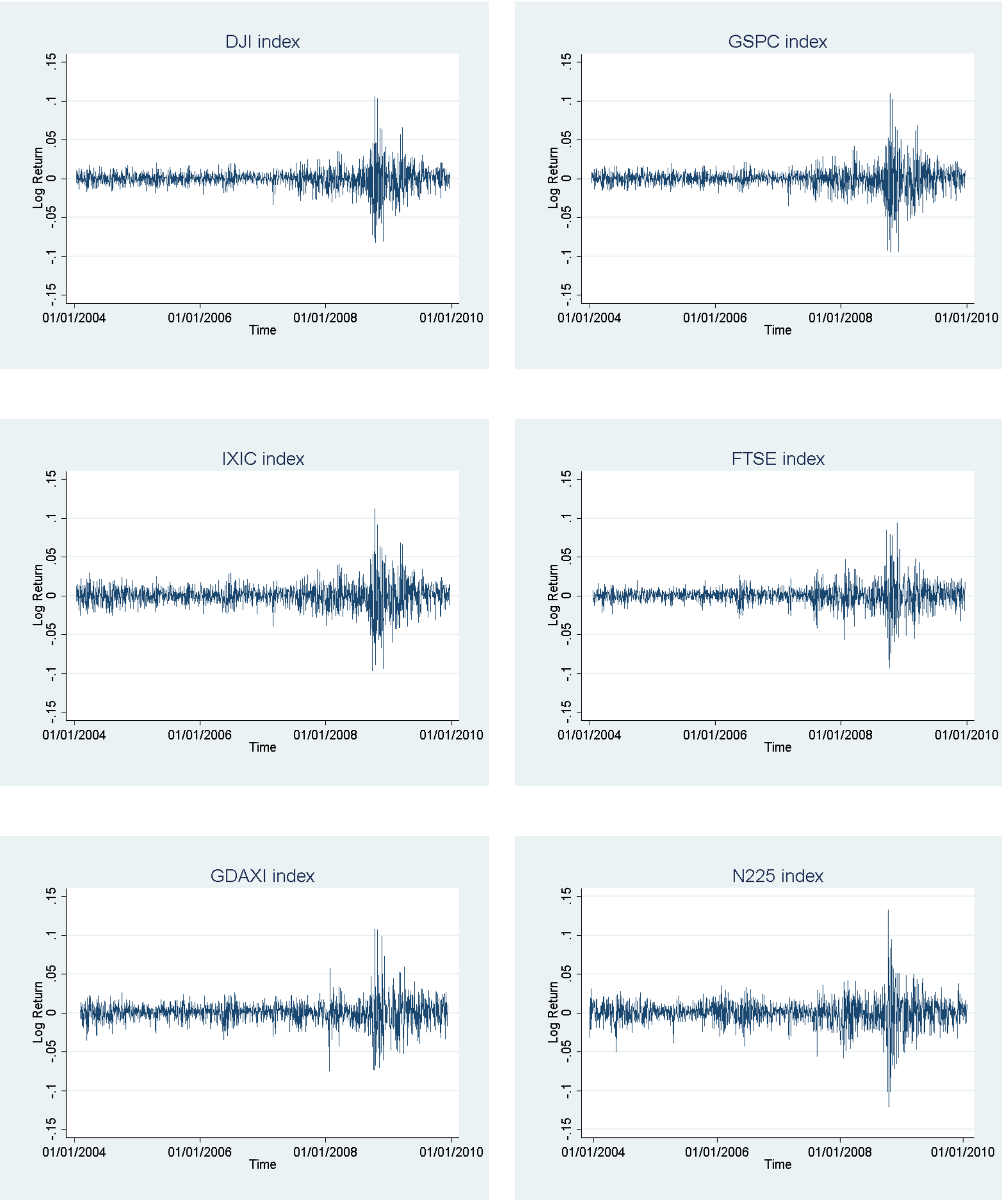

Figure B.1: Log–returns plots of the indices

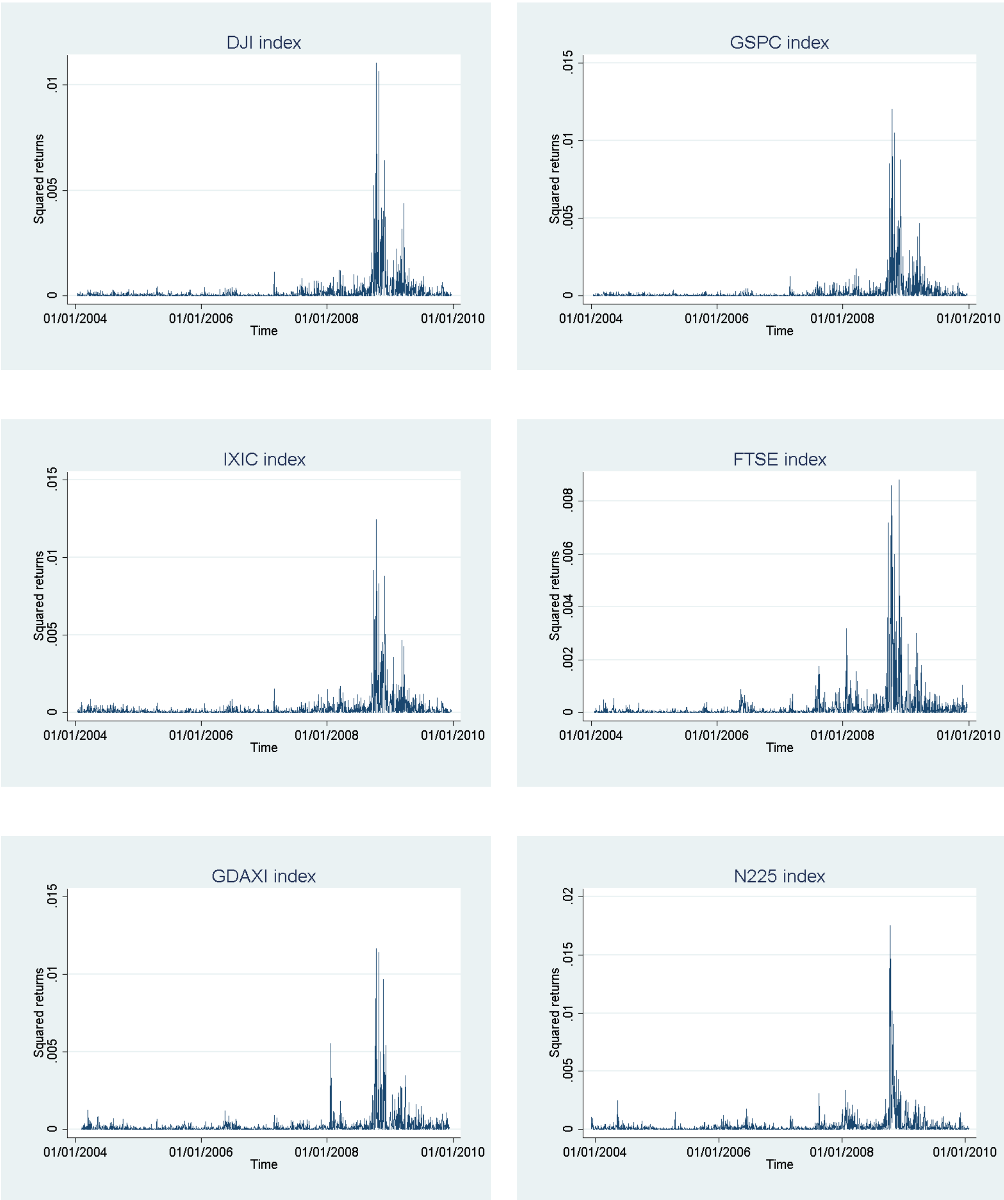

Figure B.2: Volatility of the log–returns

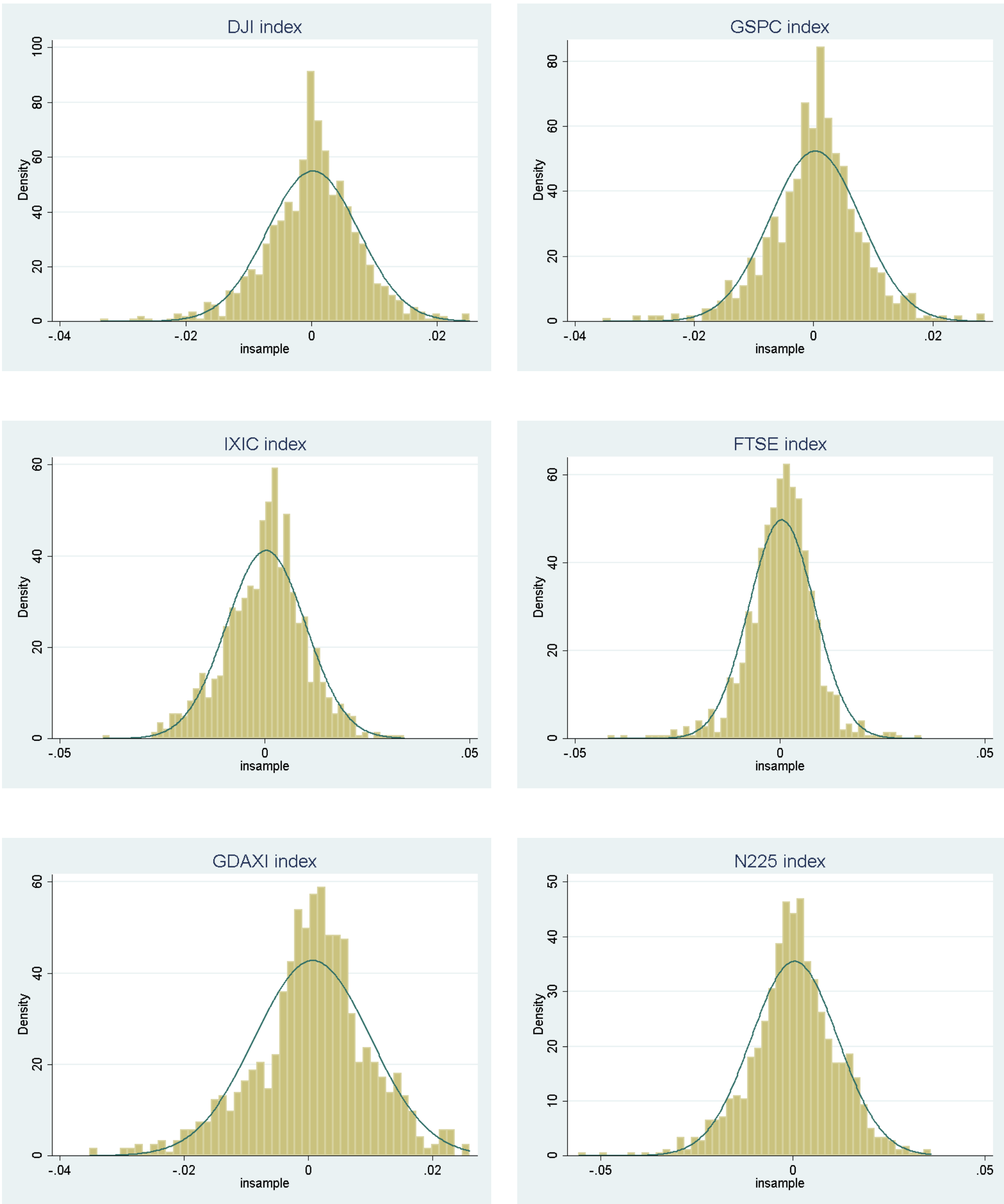

Figure B.3: In–sample histogram plots of the log–returns

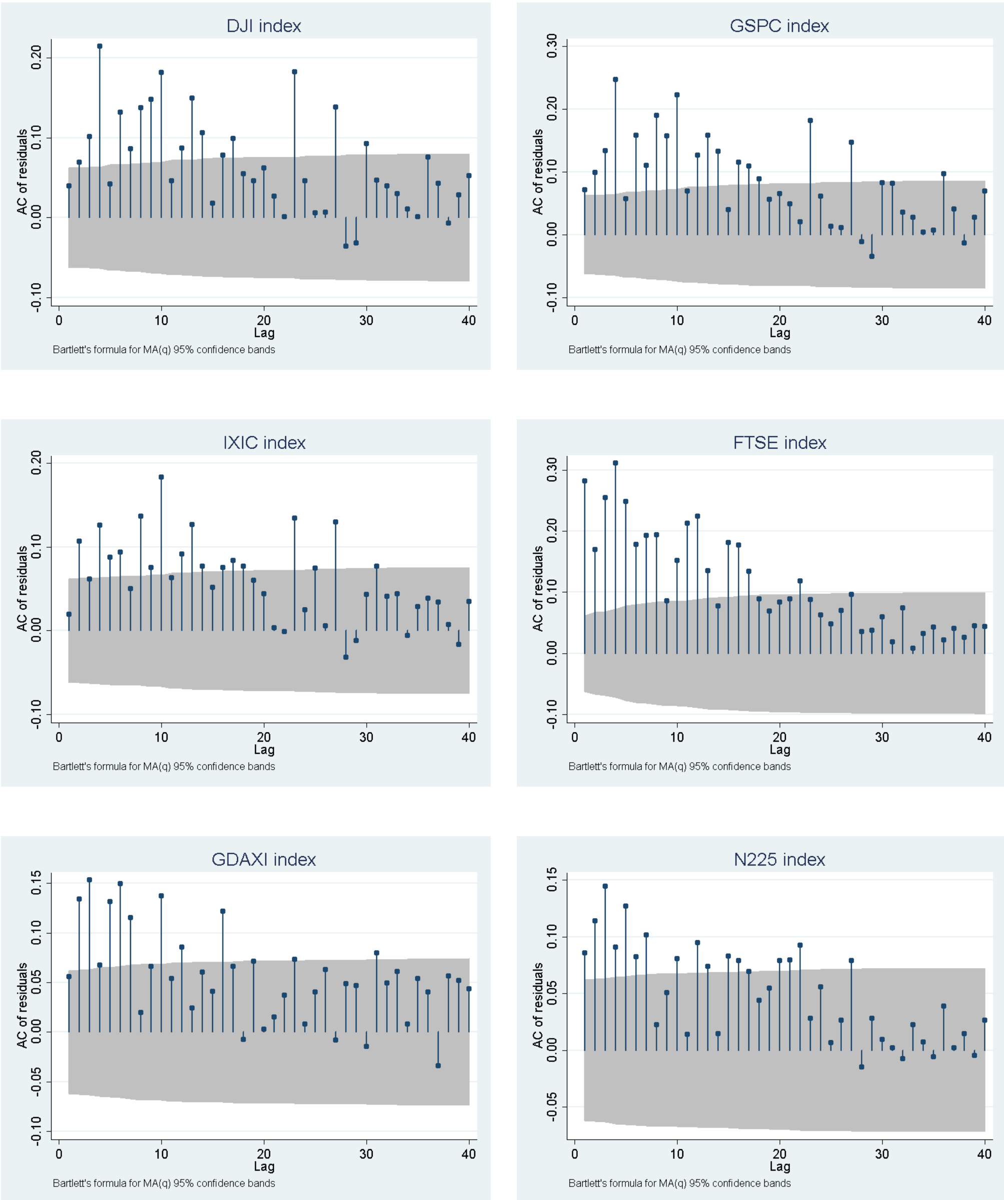

Figure B.4: In–sample autocorrelations of residuals

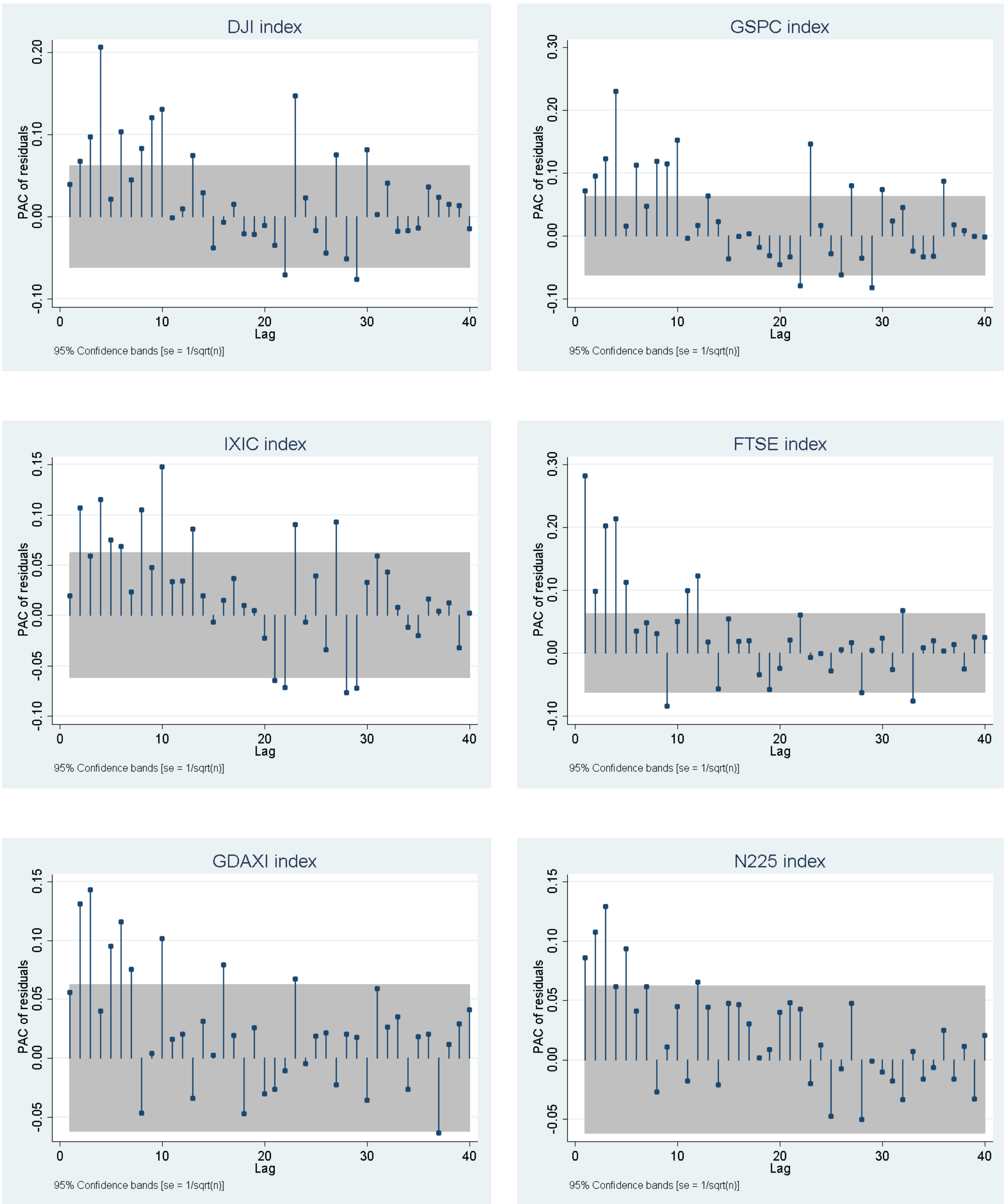

Figure B.5: In–sample partial autocorrelations of residuals

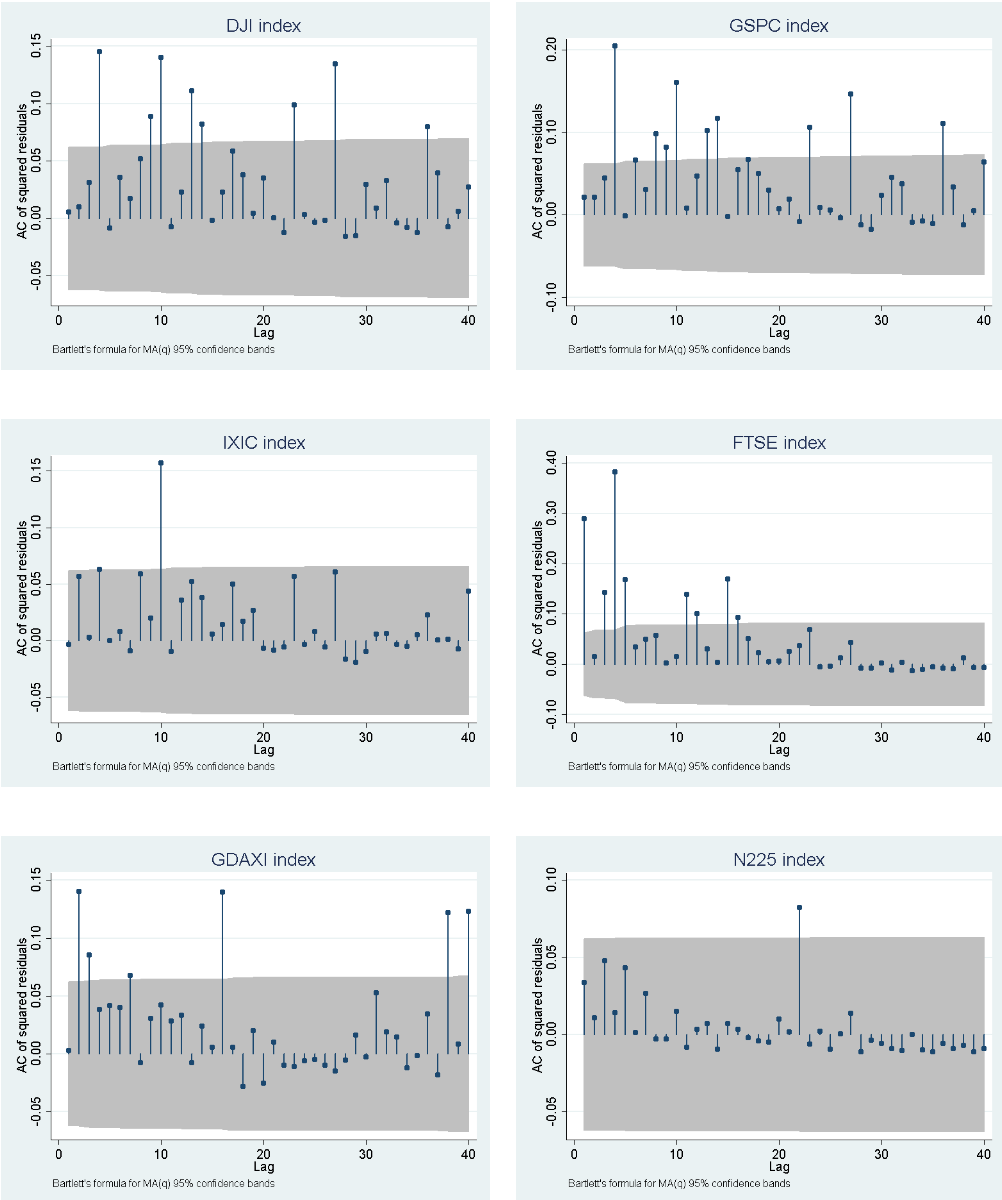

Figure B.6: In–sample autocorrelations of squared residuals

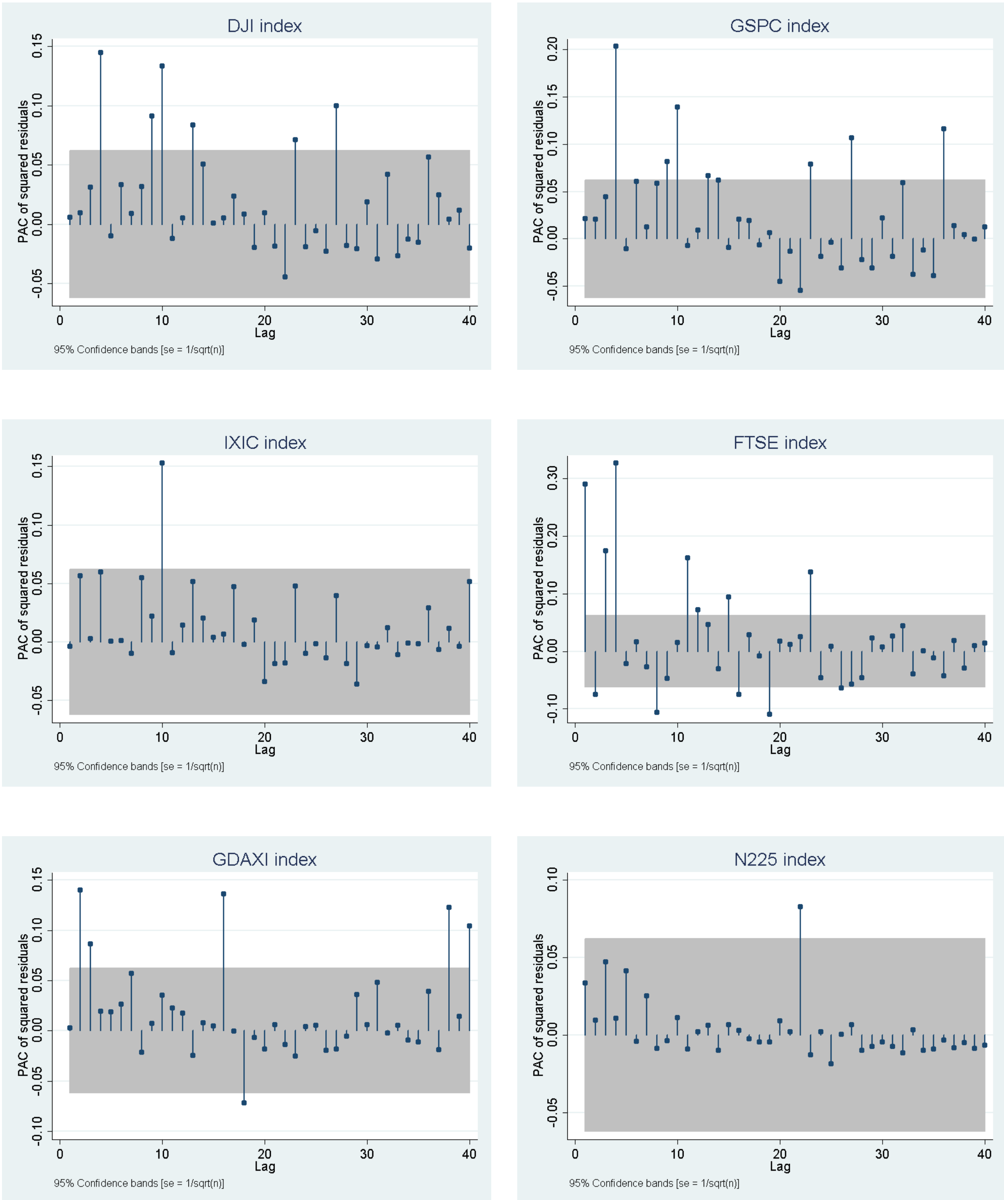

Figure B.7: In–sample partial autocorrelations of squared residuals

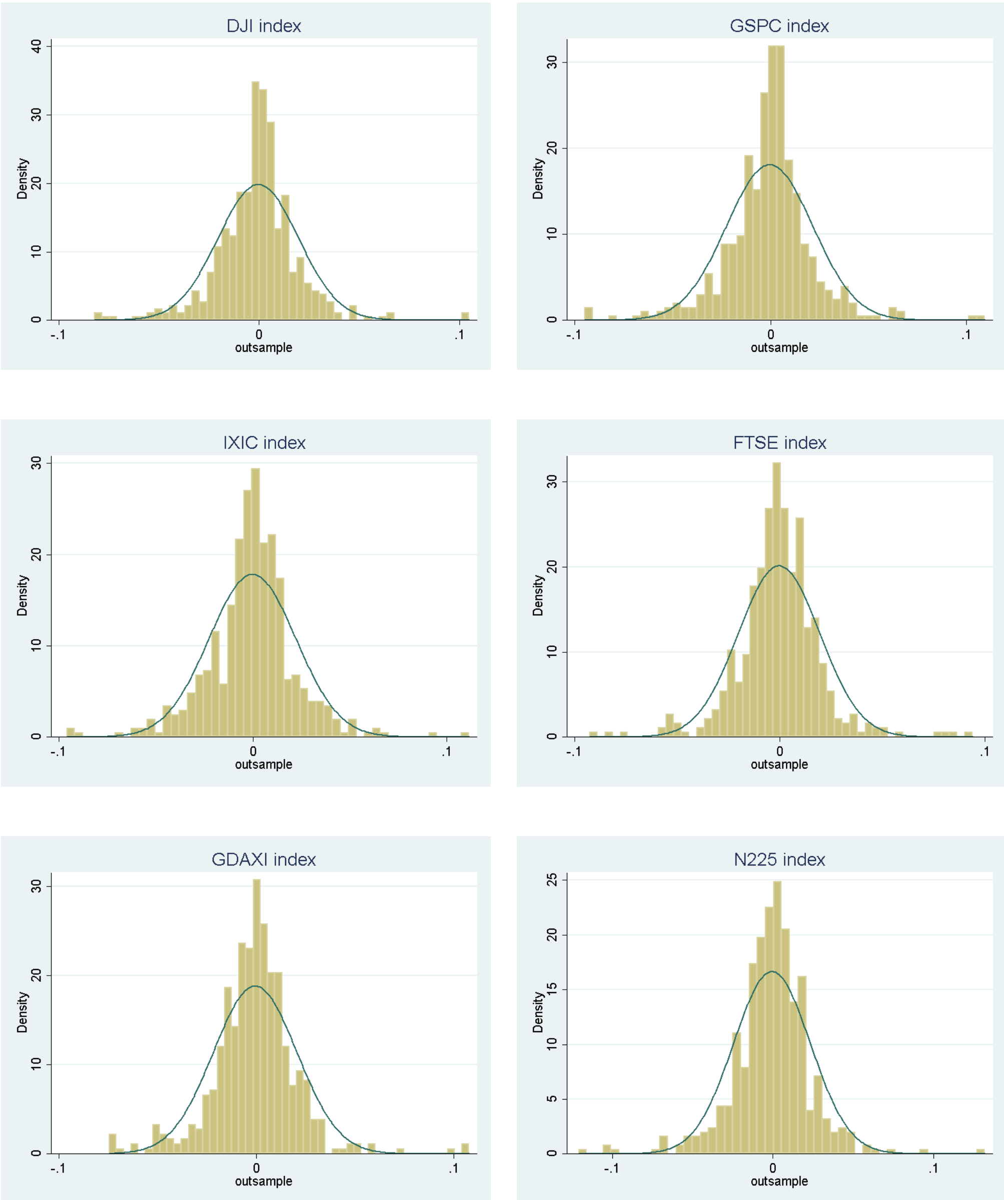

Figure B.8: Out–of–sample histogram plots of the log–returns

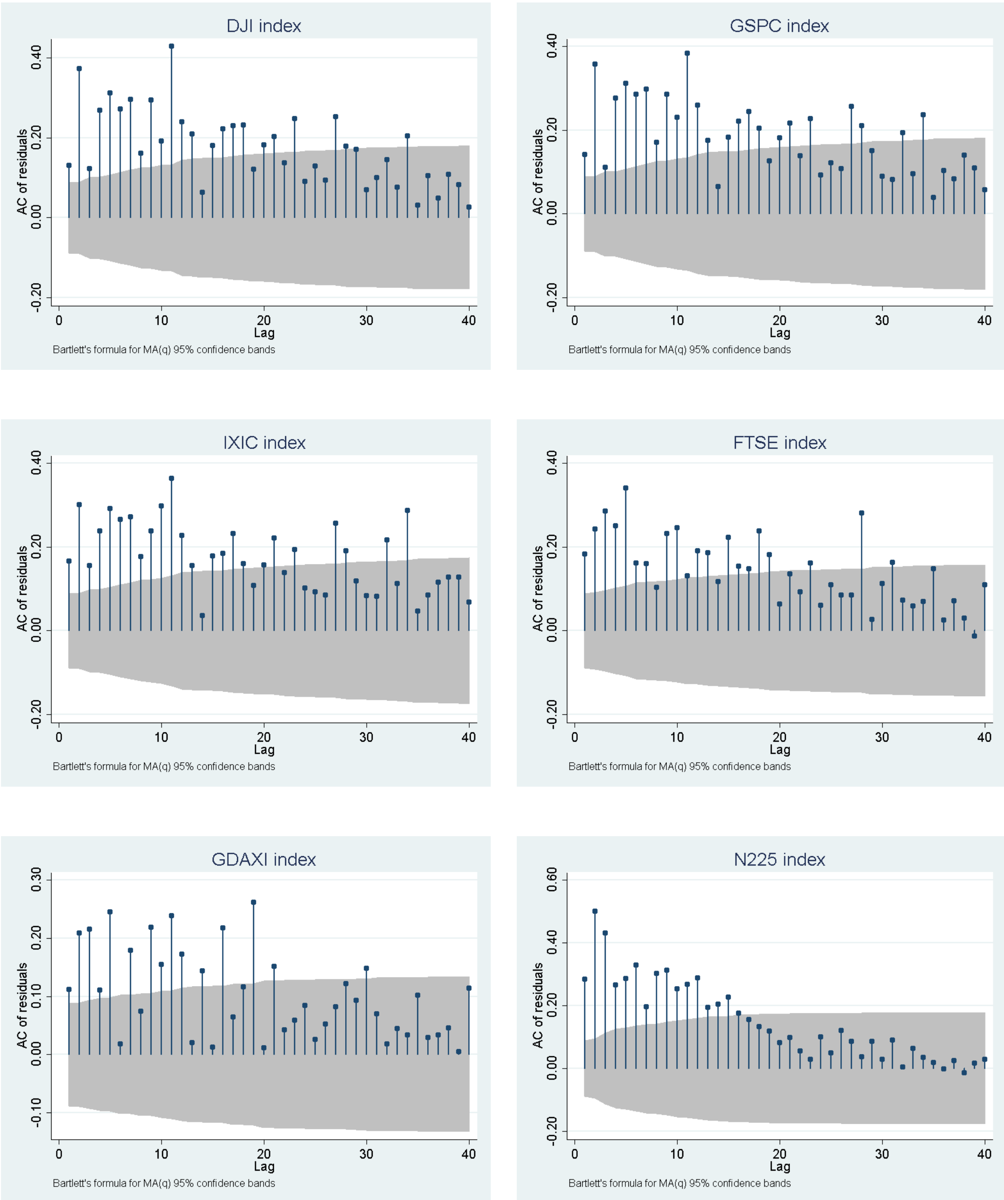

Figure B.9: Out–of–sample autocorrelations of residuals

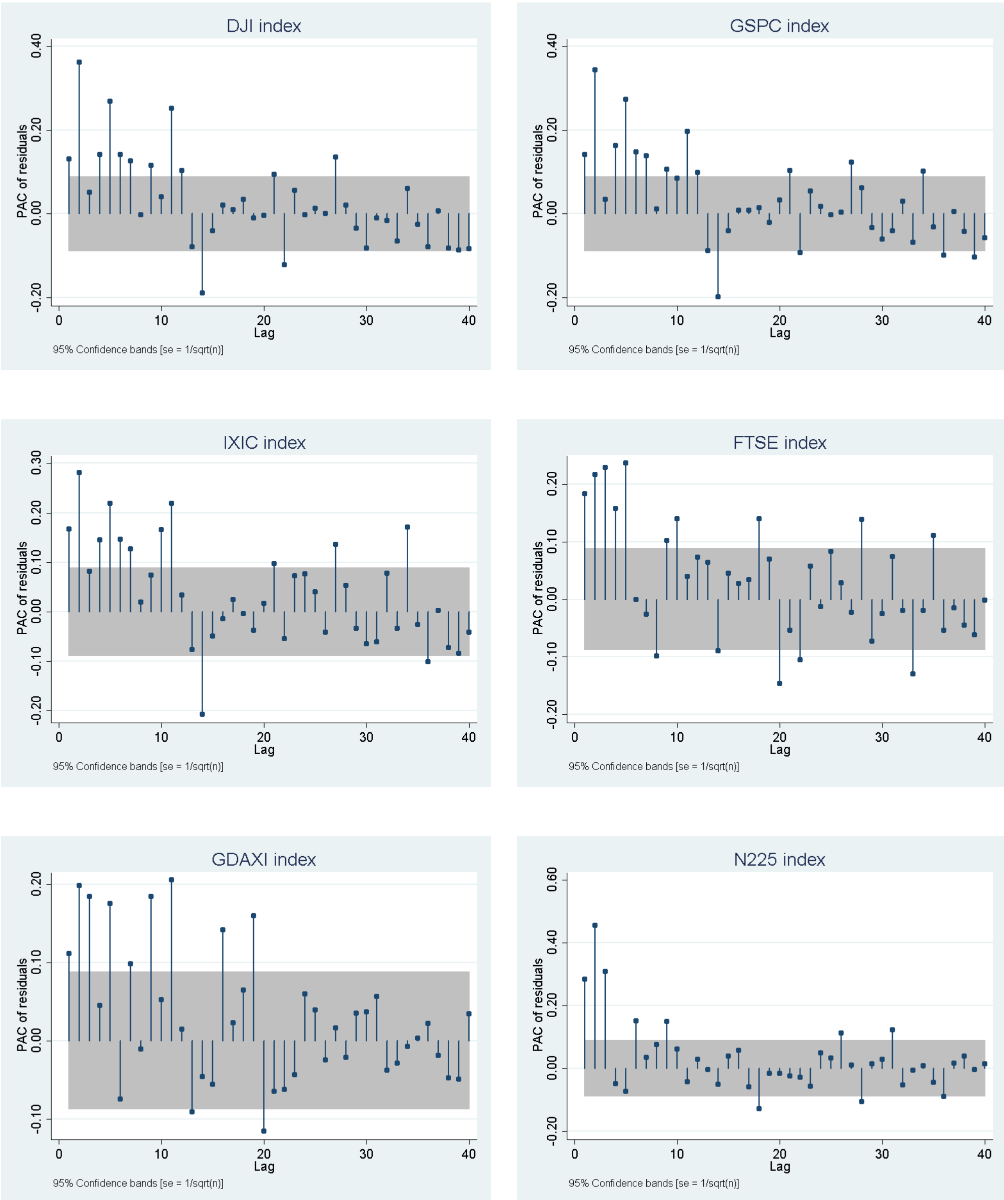

Figure B.10: Out–of–sample partial autocorrelations of residuals

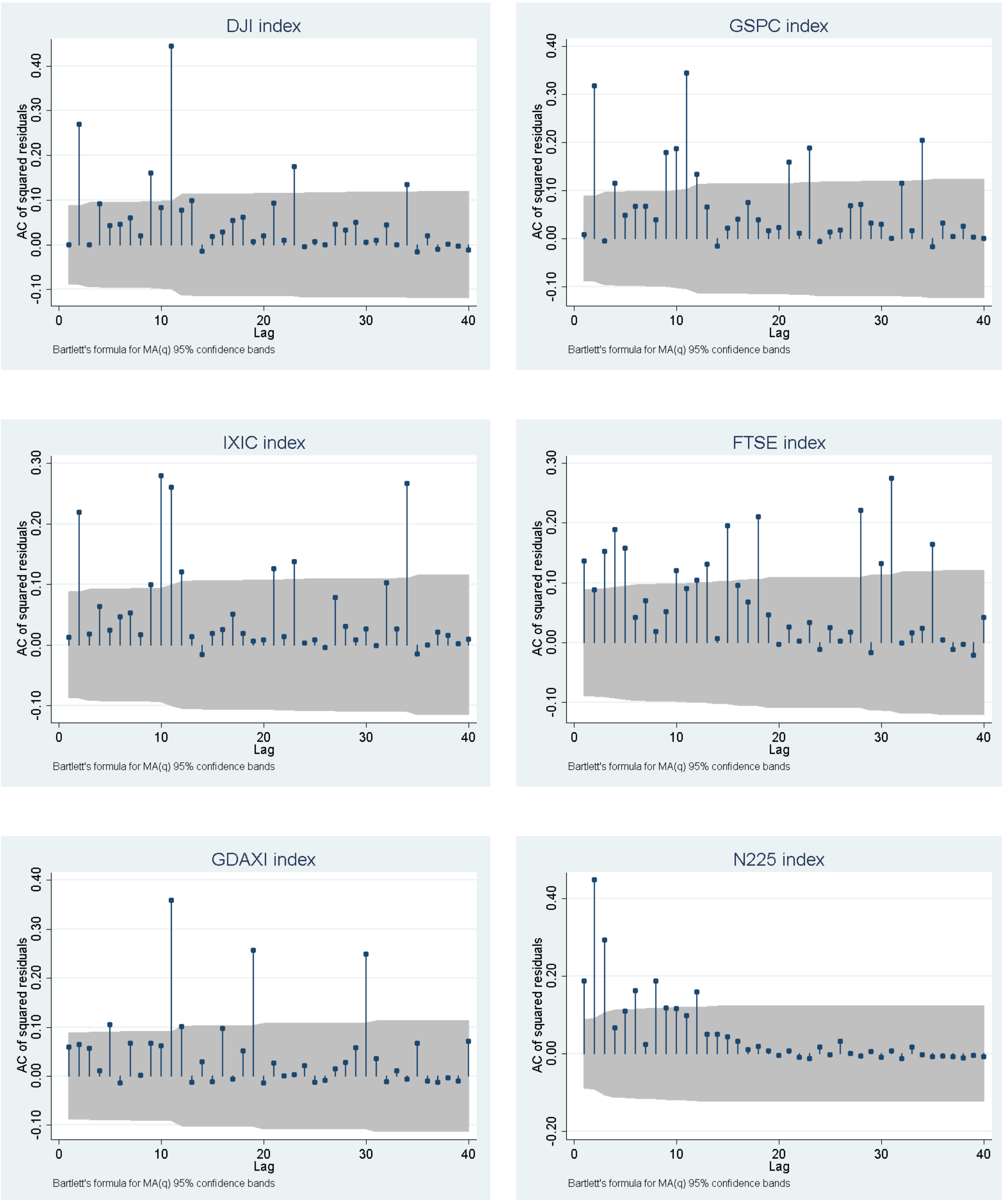

Figure B.11: Out–of–sample autocorrelations of squared residuals

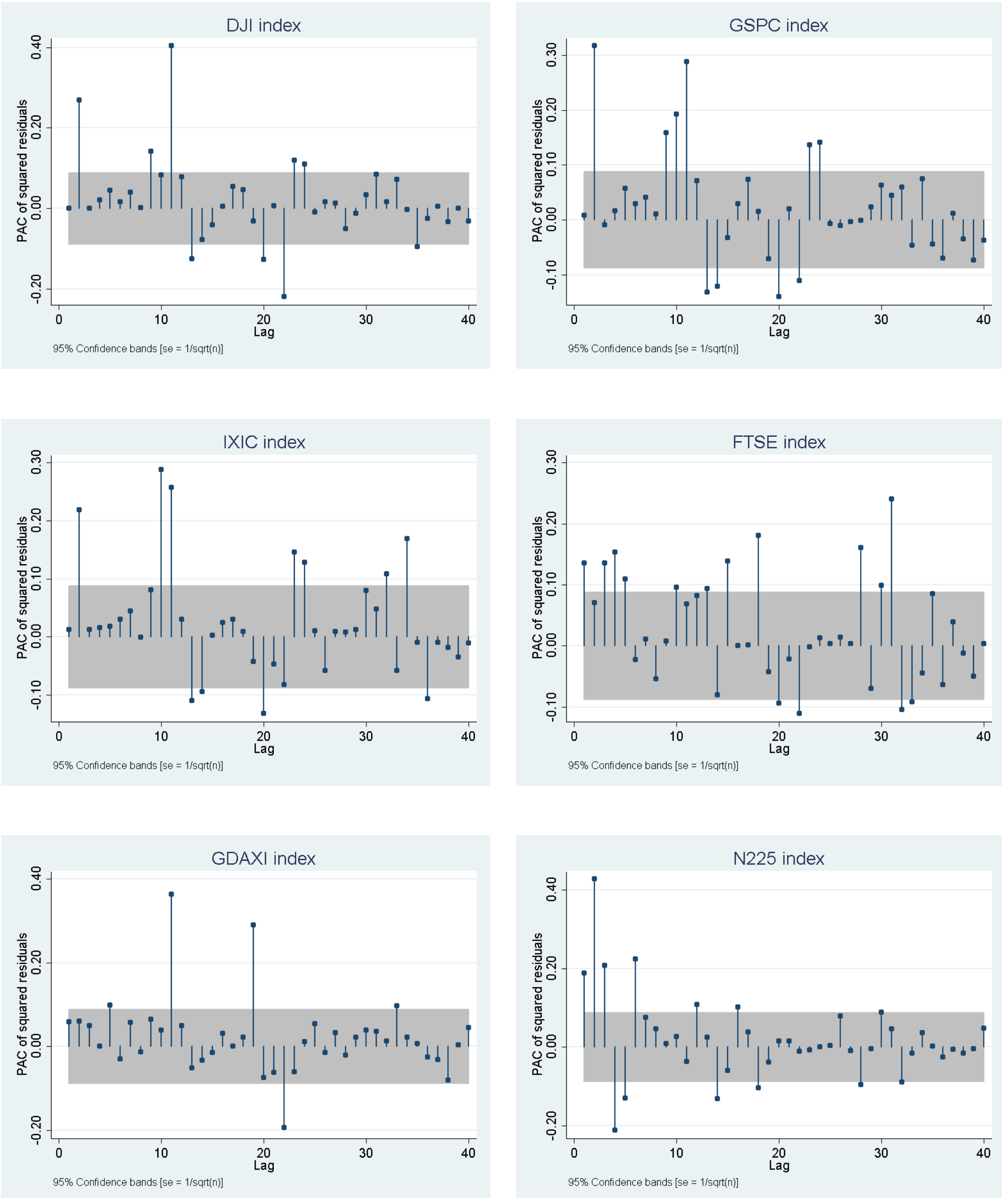

Figure B.12: Out–of–sample partial autocorrelations of squared residuals

Appendix C

In–sample estimates

```
ARCH family regression -- ARMA disturbances

Sample: 1 - 1000                                    Number of obs   =       1000
Distribution: GED                                   Wald chi2(3)    =   77054.16
Log likelihood =   3598.18                          Prob > chi2     =     0.0000

------------------------------------------------------------------------------
             |                 OPG
 log_returns |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
log_returns  |
       _cons |  -.0000605    .000219    -0.28   0.782    -.0004897    .0003686
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   .9836262   .0281988    34.88   0.000     .9283576    1.038895
         L2. |   .0233818    .027909     0.84   0.402    -.0313188    .0780825
             |
          ma |
         L1. |  -1.006093   .0042451  -237.00   0.000    -1.014414   -.9977731
-------------+----------------------------------------------------------------
ARCH         |
      abarch |
         L1. |   .0234357   .0450452     0.52   0.603    -.0648512    .1117227
         L2. |   .0209444   .0487933     0.43   0.668    -.0746888    .1165776
             |
      atarch |
         L1. |  -.2451851   .0413642    -5.93   0.000    -.3262574   -.1641128
         L2. |   .1722595   .0459106     3.75   0.000     .0822764    .2622426
             |
     sdgarch |
         L1. |   1.526676   .1945302     7.85   0.000     1.145404    1.907949
         L2. |  -.5472296   .1880843    -2.91   0.004    -.9158681   -.1785912
             |
       _cons |   .0001014   .0000547     1.85   0.064    -5.76e-06    .0002087
-------------+----------------------------------------------------------------
    /lnshape |   .4191231   .0610303     6.87   0.000     .2995059    .5387402
-------------+----------------------------------------------------------------
       shape |   1.520627   .0928043                      1.349192    1.713846
------------------------------------------------------------------------------
```

Table C.1: Top AIC value in–sample model for the DJI index

```
ARCH family regression -- ARMA disturbances

Sample: 1 - 1000                                   Number of obs   =      1000
Distribution: t                                    Wald chi2(2)    =     58.89
Log likelihood =  3562.491                         Prob > chi2     =    0.0000

------------------------------------------------------------------------------
             |                 OPG
 log_returns |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
log_returns  |
       _cons |    .000488   .0001656     2.95   0.003     .0001634    .0008125
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |   .7153106   .1556689     4.60   0.000     .4102052    1.020416
             |
          ma |
         L1. |  -.7711534   .1446424    -5.33   0.000    -1.054647   -.4876594
-------------+----------------------------------------------------------------
ARCH         |
       earch |
         L1. |   -.273966   .0431201    -6.35   0.000    -.3584799   -.1894522
         L2. |   .2254824   .0395164     5.71   0.000     .1480317    .3029331
             |
     earch_a |
         L1. |  -.2938445   .0777547    -3.78   0.000     -.446241   -.1414479
         L2. |   .3231884   .0804692     4.02   0.000     .1654717    .4809051
             |
      egarch |
         L1. |   1.530684    .127009    12.05   0.000     1.281751    1.779617
         L2. |  -.5414155   .1238022    -4.37   0.000    -.7840633   -.2987677
             |
       _cons |  -.1073776   .0461632    -2.33   0.020    -.1978559   -.0168993
-------------+----------------------------------------------------------------
     /lndfm2 |    1.70273   .2904605     5.86   0.000     1.133438    2.272022
-------------+----------------------------------------------------------------
          df |   7.488912   1.594312                      5.106317    11.69899
------------------------------------------------------------------------------
```

Table C.2: Top AIC value in–sample model for the GSPC index

```
ARCH family regression -- MA disturbances

Sample: 1 - 1000                                   Number of obs   =      1000
Distribution: t                                    Wald chi2(2)    =      1.69
Log likelihood =  3273.931                         Prob > chi2     =    0.4286

------------------------------------------------------------------------------
             |                 OPG
 log_returns |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
log_returns  |
       _cons |   .0003171   .0002731     1.16   0.246    -.0002182    .0008524
-------------+----------------------------------------------------------------
ARMA         |
          ma |
         L1. |  -.0117053   .0288429    -0.41   0.685    -.0682364    .0448258
         L2. |  -.0350109   .0288777    -1.21   0.225      -.09161    .0215883
-------------+----------------------------------------------------------------
ARCH         |
       earch |
         L1. |  -.2041116   .0432195    -4.72   0.000    -.2888203   -.1194028
         L2. |   .1679943   .0399851     4.20   0.000     .0896249    .2463636
             |
     earch_a |
         L1. |   -.214968    .078054    -2.75   0.006     -.367951    -.061985
         L2. |   .2547634   .0830723     3.07   0.002     .0919447    .4175821
             |
      egarch |
         L1. |   1.532697   .1374248    11.15   0.000     1.263349    1.802044
         L2. |  -.5474957   .1332015    -4.11   0.000    -.8085659   -.2864255
             |
       _cons |  -.1387147   .0621232    -2.23   0.026     -.260474   -.0169555
-------------+----------------------------------------------------------------
     /lndfm2 |   2.937893    .489181     6.01   0.000     1.979116     3.89667
-------------+----------------------------------------------------------------
          df |   20.87603   9.233797                      9.236343    51.23822
------------------------------------------------------------------------------
```

Table C.3: Top AIC value in–sample model for the IXIC index

```
ARCH family regression -- ARMA disturbances

Sample: 1 - 1000                                    Number of obs   =      1000
Distribution: t                                     Wald chi2(4)    =  5.11e+06
Log likelihood =  3634.532                          Prob > chi2     =    0.0000

------------------------------------------------------------------------------
             |                 OPG
 log_returns |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
log_returns  |
       _cons |    .000187   .0010007     0.19   0.852    -.0017743    .0021483
-------------+----------------------------------------------------------------
ARMA         |
          ar |
         L1. |  -.9838961   .0005555 -1771.18   0.000    -.9849849   -.9828073
         L2. |    .012365   .0012759     9.69   0.000     .0098643    .0148657
             |
          ma |
         L1. |   .9297054   .0007888  1178.67   0.000     .9281594    .9312513
         L2. |  -.0704104   .0003881  -181.44   0.000    -.0711709   -.0696498
-------------+----------------------------------------------------------------
ARCH         |
       earch |
         L1. |  -.1310023   .0010512  -124.62   0.000    -.1330626    -.128942
             |
     earch_a |
         L1. |   .1063809   .0060935    17.46   0.000     .0944378    .1183239
             |
      egarch |
         L1. |   .9711669   1.72e-06  5.6e+05   0.000     .9711635    .9711702
             |
       _cons |   -.284319    .020903   -13.60   0.000    -.3252882   -.2433499
-------------+----------------------------------------------------------------
     /lndfm2 |   31.14093   .0229207  1358.64   0.000       31.096    31.18585
-------------+----------------------------------------------------------------
          df |   3.34e+13   7.67e+11                      3.20e+13    3.50e+13
------------------------------------------------------------------------------
```

Table C.4: Top AIC value in–sample model for the FTSE index

ARCH family regression -- ARMA disturbances

Sample: 1 - 1000	Number of obs	=	1000
Distribution: GED	Wald chi2(4)	=	1902.05
Log likelihood = 3329.387	Prob > chi2	=	0.0000

log_returns	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
log_returns						
_cons	.0008273	.0002614	3.16	0.002	.000315	.0013397
ARMA						
ar						
L1.	.3179623	.0405725	7.84	0.000	.2384417	.3974829
L2.	-.9180841	.039051	-23.51	0.000	-.9946226	-.8415455
ma						
L1.	-.3307553	.0298507	-11.08	0.000	-.3892617	-.2722489
L2.	.9528917	.0290591	32.79	0.000	.8959369	1.009847
ARCH						
abarch						
L1.	.1263833	.0356596	3.54	0.000	.0564917	.1962749
L2.	-.106246	.0364772	-2.91	0.004	-.1777401	-.034752
atarch						
L1.	-.2456664	.042911	-5.73	0.000	-.3297705	-.1615624
L2.	.2262409	.0402146	5.63	0.000	.1474217	.3050601
sdgarch						
L1.	1.751816	.0806066	21.73	0.000	1.59383	1.909802
L2.	-.766888	.0750138	-10.22	0.000	-.9139123	-.6198636
_cons	.0000625	.0000398	1.57	0.117	-.0000156	.0001406
/lnshape	.3692524	.0691173	5.34	0.000	.2337849	.5047199
shape	1.446653	.0999888			1.263373	1.656521

Table C.5: Top AIC value in–sample model for the GDAXI index

```
ARCH family regression -- MA disturbances

Sample: 1 - 1000                                   Number of obs   =       1000
Distribution: t                                    Wald chi2(1)    =       0.03
Log likelihood =  17898.67                         Prob > chi2     =     0.8593

------------------------------------------------------------------------------
             |                 OPG
 log_returns |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
log_returns  |
       _cons |   .0002007    .003066     0.07   0.948    -.0058085      .00621
-------------+----------------------------------------------------------------
ARMA         |
          ma |
         L1. |   .0337188   .1902067     0.18   0.859    -.3390795    .4065171
-------------+----------------------------------------------------------------
ARCH         |
       earch |
         L1. |  -.2693833   .3097676    -0.87   0.385    -.8765166      .33775
         L2. |   .0410691   .1085756     0.38   0.705    -.1717351    .2538733
             |
     earch_a |
         L1. |  -.0281678   .0007323   -38.47   0.000    -.0296029   -.0267326
         L2. |   .2168371   .2267851     0.96   0.339    -.2276537    .6613278
             |
      egarch |
         L1. |   .8597736   .0533203    16.12   0.000     .7552677    .9642795
             |
       _cons |  -1.282041   .6599377    -1.94   0.052    -2.575495    .0114131
-------------+----------------------------------------------------------------
     /lndfm2 |   36.41557   8.72e-08  4.2e+08   0.000     36.41557    36.41557
-------------+----------------------------------------------------------------
          df |   6.53e+15   5.70e+08                      6.53e+15    6.53e+15
------------------------------------------------------------------------------
```

Table C.6: Top AIC value in–sample model for the N225 index