UNIVERSITY OF ECONOMICS, PRAGUE
Faculty of Informatics and Statistics

Score-driven models for Value at Risk and Expected Shortfall

MASTER THESIS

Study programme: Quantitative Methods in Economics
Field of study: Econometrics and Operational Research

Author: Bc. Kateřina Nováková
Supervisor: Ing. Petra Tomanová, MSc

Prague, 2019

Declaration of Independence

I hereby declare that I am the sole author of the thesis entitled *Score-driven models for Value at Risk and Expected Shortfall*. I duly marked out all quotations. The used literature and sources are stated in the attached list of references.

November 25, 2019 in Prague ....

Student's name

Acknowledgements

I would like to express my sincere gratitude to my supervisor Ing. Petra Tomanová, MSc from University of Economics, Prague for the guidance which helped me in writing this thesis.

Abstrakt

Cílem této práce je odhad cenové volatility a srovnání odhadů kvantitativních ukazatelů řízení rizika jako jsou „hodnota v riziku“ (angl. Value at Risk či VaR) a „podmíněná hodnota v riziku“ (angl. Expected Shortfall či ES) pro čtyři světové cenové indexy pomocí modelů podmíněné heteroskedasticity (angl. models of conditional heteroskedasticity). Aplikovány jsou Generalized Autoregressive Score (GAS) modely, které oproti Generalized Autoregressive Conditional Heteroskedasticity (GARCH) modelům dokáží popsat podmíněné pravděpodobnostní rozdělení časové řady komplexněji a vedou k adekvátnějším odhadům. Odhadnuté hodnoty VaR a ES jsou zpětně testovány (angl. backtesting) pomocí tzv. rolling windows, které provádí jednokrokové odhady časové řady pomocí odhadnutých parametrů modelu. Ty mohou být odhadovány v každém kroku nebo každý k-tý krok. Hlavním cílem této práce je zkoumání dopadu délky kroku, při kterém se znovu odhadují parametry GAS modelu při použití rolling windows a následné porovnání odhadnutých ukazatelů rizika VaR a ES pomocí testu dynamických kvantilů a ztrátových funkcí. Výsledky ukázaly, že pro GAS model se Studentovým t rozdělením nemá délka kroku výrazný vliv, zatímco GAS model s Gaussovým rozdělením a delším krokem odhaduje VaR a ES výrazně zápornější v období cenových šoků.

Klíčová slova: Volatilita, GAS model, Value at Risk, Expected Shortfall, rolling window

Abstract

The aim of this thesis is volatility estimation and estimates' comparison of financial risk measures, which are specially Value at Risk (VaR) and Expected Shortfall (ES) for four world market price indices. Models of conditional heteroskedasticity are utilized. Generalized Autoregressive Score (GAS) models are applied since they are able to describe the probability density of observations in a more complex way than Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models. Furthermore, they lead to more adequate estimates. VaR and ES estimates are backtested using rolling windows, which calculate one-step ahead predictions based on estimated parameters of GAS model. Parameters can be estimated in each step or in every k-step. The main contribution of this thesis is to study the impact of the length of the refit step of parameters for rolling windows in GAS models followed by comparison of estimated values of VaR and ES using Dynamic quantile (DQ) test and calculating the loss functions. The results showed that the length of the refit step does not significantly influence estimates of VaR and ES for GAS models with the Student's t distribution. However, it underestimates values of VaR and ES significantly in the periods of price shocks for GAS models with the Gaussian distribution.

Keywords: Volatility, GAS model, Value at Risk, Expected Shortfall, rolling window

Contents

Introduction

In finance, volatility of time series might measures the dispersion for a given market index. Despite impossibility of direct observation, it can be estimated as a simple standard deviation of the price returns. When the volatility is high, the level of risk increases. Financial data evinces volatility clustering meaning that periods of the same level of volatility group together and create clusters. It turned out that standard deviation is rather simple measure, which does not capture the volatility clustering and hence, it is not sufficient. As a consequence, Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models were developed. Moreover, the financial data often violates the normality assumption which led to introducing new models of conditional heteroskedasticity. Generalized Autoregressive Score (GAS) model is a general framework for time series modeling.

Many studies have been written about the historical volatility and its properties. The desire to capture the dynamic behavior of the time series' processes led to models with time-varying parameters. These models were firstly categorized by Cox et al. (1981) into observation-driven models and parameter-driven models. Observation-driven models let the time-varying parameters be functions of the lagged variables. The parameters are perfectly predictable given the past information. In parameter-driven models the parameters are stochastic processes and they are not perfectly predictable. Examples of the observation-driven models are GARCH models of Engle (1982) or GAS models by Creal et al. (2008).

The contribution of GAS models is in their ability to describe the probability density of observations in a more complex way. They are based on a score which is the first derivative of the logarithm of the conditional observation density with respect to the vector of time-varying parameters and it can be scaled applying a scaling function. The scaling function is represented by the Fisher information matrix to the power of an additional parameter $\gamma > 0$. The key features of GAS models can be simple estimation, the straightforward evaluation of likelihood and the generality of models.

Risk measures typically quantify the level of financial risk. Acerbi et al. (2002) defined four axioms of coherence to distinguish 'good' risk measures. The most applied measure Value at Risk (VaR) turned out not to be coherent since it violates one of the axioms, specifically the sub–additivity. VaR is a threshold of possible losses and it is indifferent to the size of the losses beyond. Therefore, Rockafellar et al. (2002) introduced a new risk measure Expected Shortfall (ES) which fulfills all the axioms, hence it is coherent risk measure. Acerbi et al. (2002) also mentioned that only fulfillment of all the axioms can lead to correct results and violating some of them can lead to wrong or even paradoxical conclusions. ES represents the expected loss in the 100α % worst cases which means that this measure takes into account sizes of the worst losses and leads to better quantification of the possible risk.

The model performance can be measured by value of loss function. Quantile Loss (QL) function is very frequently applied for quantile regressions in the VaR evaluation but ES does not have individual loss function. However, Fissler et al. (2016) showed that VaR and ES are jointly elicitable and the corresponding function is FZ loss (FZL) function.

This thesis examines price returns of four market indices corresponding to U.S. stock market (DJIA, S&P 500), market in Great Britain (FTSE 100) and Japanese market (TOPIX) over three different time periods (1984 − 1995 including the Black Friday, 2000 − 2010 including the financial crisis and the resent years 2010 − 2019). The aim is to compare estimated volatility using GARCH and GAS models and the corresponding values of VaR followed by testing equality of the static parameters in GAS models over time. Backtesting of VaR and ES with GAS models assuming different conditional distribution of observations is presented thereafter. The main contribution of this thesis is in studying the impact of the different length of refit step of parameters in the rolling windows for GAS models with a different conditional distributions of observations. Models are assesses by comparing the individual loss functions (QL function for VaR and FZL function for VaR and ES jointly) which lead to the choice of the best performing model.

The outline of the thesis is as follows. At first, Section 1 reviews the problem of volatility in financial time series and the models of conditional heteroskedasticity followed by describing the framework of GAS models and risk measures along with the approaches to model comparison. Section 2 introduces the R packages *GAS* and *RUGARCH*, Section 3 describes used data and Section 4 presents the application to financial return data along with the discussed results.

1. Methodology

Firstly, the volatility of time series is defined in Section 1.1, followed by the overview of the models of conditional heteroskedasticity in Section 1.2. There is a short summary of ARCH and GARCH models and their relation to GAS models which are described in detail in Section 1.3 starting with the model specification and followed by the discussion about suitable probability distributions and maximum likelihood estimation. The difference among parametric, semi-parametric and non-parametric models is mentioned thereafter in Section 1.4.

Section 1.5 covers the theory about the risk measures and their properties. Several methods of model comparison including information criteria and backtesting VaR and ES using the loss functions are discussed in Section 1.6. And the last Section 1.7 describes the basic methods of time series' analysis, for example the Shapiro–Wilk normality test or the Augmented Dickey–Fuller test of the stationarity.

The first two sections are based on (Satchell et al., 2011) and (Hušek et al., 2003), if not cited differently.

1.1 Volatility of Time Series

Substantial growth in financial econometrics led to estimating the level of the risk along with the analysis and prediction of returns. Models of conditional heteroskedasticity were developed as a consequence. The volatility of time series, also called conditional heteroskedasticity, is not directly observable. However, it is possible to estimate it. In finance, the volatility might measure fluctuations of some assets or its return. It represents the period of high variability or increasing variance and it can be seen as a level of risk. Furthermore, volatility modeling enables to compute Value at Risk (VaR) or Expected Shortfall (ES) which are the standards and two leading risk measures used in finance (Ardia et al., 2018).

Historical and implied volatility are distinguished in financial econometrics. Historical volatility is the volatility experienced by the underlying stock and it is represented by the sample standard deviation over a certain time interval. It is used for short-time prediction or for comparing two stocks with each other. On the contrary, Canina et al. (1993) state that implied volatility can be interpreted as informationally superior to historical volatility and it is used to forecast future volatility. An example can be the Black-Scholes model for the setting of implied volatility. The model assumes that the price for the stock follows a logarithmic diffusion process with constant instantaneous mean and volatility and it is possible to derive the conditional standard deviation which represents the implied volatility of the observed asset. However, in practice some assumption can be violated. The most common assumption is that the logarithmic price is not normally distributed.

A very common feature which can be observed in financial time series is volatility clustering. Volatility can be low in some periods and high in others but periods of low volatility (low risk) concentrate like periods of high volatility (high risk) and create clusters. This property leads to the violation of the homoskedasticity assumption because the variance of the residuals varies over time. Moreover, the maximum likelihood estimates would be biased. Consequently, models with unconditional variance can lead to wrong and biased results (Kaufman, 2013). Instead, Engle (1982) proposed conditional variance models (or models of conditional heteroskedasticity), which take into account the second conditional moment, i.e. variability and allow it to change over time. They are able to capture the changing conditions of market uncertainty which is also the reason why their application is so wide (Arlt et al., 2003).

Volatility is usually stationary and it develops smoothly and continuously. Therefore, sudden shocks are very rare and changes have some fixed range, which they usually do not exceed, they do not diverge to infinity. However, the behavior of volatility returns differs when prices suddenly rise and when they decrease. Positive and negative shocks are not symmetric for conditional variance and lead to asymmetric distributions of the returns. The asymmetry of the behavior is called the leverage effect.

1.2 Models of Conditional Heteroskedasticity

Models of conditional heteroskedasticity can be categorized into observation-driven models and parameter-driven models (Cox et al., 1981). These two classes are described in detail in Section 1.3. Another possibility is to distinguish the models according to the approach in the variance modeling. Variance can be modeled using an exact stochastic function or a stochastic equation. ARCH and GARCH models are representatives of the first group, models of stochastic volatility are representatives of the second group. These models usually express behavior or development of the conditional variance of the error term. Let's denote y_t as the price of the market index at time t. Then r_t is the logarithmic return of some financial asset at time t computed as the first differences of logarithmic values of y_t. Conditional mean and variance for a given F_{t-1} can be written as

$$r_t = \ln y_t - \ln y_{t-1}, \tag{1.1}$$

$$\mu_t = E(r_t|F_{t-1}), \tag{1.2}$$

$$\sigma_t^2 = var(r_t|F_{t-1}) = E\left[(r_t - \mu_t)^2 |F_{t-1}\right], \tag{1.3}$$

where F_{t-1} is a set of relevant information at time $t-1$, which can contain linear equations of the past returns. Therefore, it is assumed that r_t is possible to be captured by a stationary ARMA(p,q) model.

ARCH models

Autoregressive conditional heteroskedasticity (ARCH) models are used for financial time series when the variance depends on its past values and this dependence is possible to express by autoregression. These models were proposed Robert F. Engle in 1982 and they became commonly used. ARCH(1) was first applied to describe and estimate the inflation variance in Great Britain (Engle, 1982). Let's assume a simple linear regression model:

$$y_t = \beta_0 + \boldsymbol{x_t'}\boldsymbol{\beta} + u_t, \tag{1.4}$$

where y_t is the dependent variable, $\boldsymbol{x_t}$ is $k \times 1$ vector of regressors and $\boldsymbol{\beta}$ is $k \times 1$ vector of parameters. The random error term u_t has independent normal distribution with zero mean and constant finite variance σ^2. ARCH allows the variance to be time-varying and it is possible to express the dependence σ_t^2 on the past values in the form of

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2, \tag{1.5}$$

where $\alpha_0 \geq 0$ and $\alpha_1 \geq 0$ to ensure that the conditional variance σ_t^2 is non-negative.

Equation (1.5) is the fundamental ARCH(1) model and it is also called the equation of conditional variance or volatility. The expression (1) means that the conditional variance depends on one lagged value of the random error term u_{t-1} only. Both equations formulate a complete ARCH(1) model. It can be assumed that in the case of a significant shock in time $t-1$ (i.e. u_{t-1} is large) there will be a significant shock at time t with a substantial probability (i.e. u_t will be large, too).

However, the conditional variance can depend on more than one lagged value. That is why it is possible to extend ARCH for q lagged value. ARCH(q) is used when it is expected that the volatility of time series changes slower that in ARCH(1). These models are not commonly used but the dependence σ_t^2 on past values can be generally written as

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \ldots + \alpha_q u_{t-q}^2. \tag{1.6}$$

There are several drawbacks to this approach. ARCH models assume that negative and positive shocks have the same impact on the volatility. But financial practice has shown that this is not true and the assets react differently to both shocks. Another problem is their limitation in expressing higher kurtosis then kurtosis of the Gaussian distribution or the fact that the probability of estimating a negative model parameter is higher with a higher number of parameters. Parameter negativity violates the requirement of non-negativeness of all estimated parameters. These and some more limitations caused the formation of generalized version of ARCH known as GARCH models.

GARCH models

Generalized autoregressive conditional heteroskedasticity (GARCH) models were independently introduced by Bollerslev (1986) and Taylor (1986) and they are the most applied models of all volatility models. The most valued benefits are their abilities to describe volatility clustering and heavy tails or significant kurtosis of time series. Another advantage can be effective estimation of parameters in case of a long memory process.

The fundamental model is GARCH(1,1) which is the simplest generalized ARCH model and also the most utilized one. The difference between ARCH(1) and GARCH(1,1) is its extension in the lagged value of the conditional variance and its form is

$$\sigma_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \tag{1.7}$$

The model for variance σ_t^2 can be rewritten if the lagged value is substituted and it actually became a function of all lagged error terms:

$$\sigma_t^2 = \frac{\alpha_0}{1-\beta_1} + \alpha_1 (u_{t-1}^2 + \beta_1 u_{t-2}^2 + \beta_1^2 u_{t-3}^2 + \beta_1^3 u_{t-4}^2 + \dots). \tag{1.8}$$

Non-negative conditional variance is assured by validity of $\alpha_0 \geq 0$ and the identity conditions of $0 < \beta_1 \leq 1$ are assured by $\alpha_1 > 0$. The big benefit of GARCH(1,1) is that it is possible to replace ARCH(∞) with only three parameter, which will be estimated. And moreover, it has much more degrees of freedom, which are substantial in case of small sample size. If the model is rewritten by adding u_t^2 on both sides and transferring σ_t^2 to the right side of the equation, ARMA(1,1) model for u_t^2 can be obtained:

$$u_t^2 = \alpha_0 + (\alpha_1 + \beta_1) u_{t-1}^2 + v_t - \beta_1 v_{t-1}, \tag{1.9}$$

where $v_t = u_t^2 - \sigma_t^2$, lagged value u_{t-1}^2 is the AR constituent and v_t, v_{t-1} are the MA constituents. The condition of stationarity can be written as $\alpha_1 + \beta_1 < 1$ and if it holds, the unconditional mean of u_t^2 or the conditional variance of u_t is equal to

$$\sigma^2 = \frac{\alpha_0}{1-(\alpha_1+\beta_1)}, \tag{1.10}$$

which means that the unconditional variance of u_t is homoskedastic and the value of $(\alpha_1 + \beta_1)$ close to one refer to the significant persistence in volatility.

A general model GARCH(p,q) exits but it is rarely used in practice since GARCH(1,1) is usually sufficient to express volatility clustering.

Model parameters can be estimated using the maximum likelihood method. Maximum likelihood estimates can be obtained by maximization of the logarithm of the likelihood function. However, a common problem with the financial time series is the normality assumption. It can be sometimes assumed that the random error term follows Student's t distribution and then the parameters can be estimated by maximization of the logarithmic maximum likelihood corresponding to that distribution. If the full model is correctly specified, the estimated parameters are still consistent and asymptotically normal.

1.3 GAS Models

Generalized autoregressive score (GAS) models were introduced by Creal et al. (2008). They are also known as Dynamic Conditional Score (DCS) proposed by Harvey (2013) or score-driven models (Ardia et al., 2019). These models belong to the class of observation-driven models using scaled score of the likelihood function as the driving mechanism. The possibility to set some parameters time-varying is necessary for capturing the dynamic behavior of time series (Creal et al., 2013).

Cox et al. (1981) mention two categories of time-varying models: observation-driven models and parameter-driven models. Observation-driven models have the parameters dependent on the lagged values of past observations and exogenous variables. They are stochastic but it is possible to predict them given past information. Moreover, it simplifies likelihood evaluation which is straightforward. Typical representatives are GARCH models, Autoregressive Conditional Duration (ACD) or Autoregressive Conditional Intensity models (ACI) and others. In parameter-driven models, the parameters evolve according to some stochastic process. However, these parameters are not perfectly observable due to their dependence on the given past and present observations. The estimation is more demanding because the associated likelihood functions are not available in closed-form. Typical examples can be the Stochastic Volatility (SV) or Stochastic Intensity models (Creal et al., 2008).

The main difficulty for observation-driven models is the choice of a linking function between the past observations and parameters. Creal et al. (2013) suggest to use a scaled score function of the model density at time t. This function should be an effective choice and it should also be applicable to a wide variety of non-Gaussian and non-linear models. If the scaling is chosen appropriately, GARCH, ACD and ACI can be obtained as a special case.

A huge benefit of GAS models is their ability to take advantage of the complex density structure rather than only consider means and higher moments. Moreover, the likelihood evaluation is straightforward as well as for GARCH models.

Basic Model Specification

A general class of observation-driven time-varying parameter models is formulated in this section. This section is based on Ardia et al. (2019) and the same notation is adopted.

Let $\boldsymbol{y}_t$ be an N-dimensional random vector of the dependent variables at time t and $\boldsymbol{\theta}_t$ be a vector of time-varying parameters. Then $\boldsymbol{y_t}$ follows conditional observation density p

$$\boldsymbol{y}_t|\boldsymbol{y}_{1:t-1} \sim p(\boldsymbol{y}_t;\boldsymbol{\theta}_t) \tag{1.11}$$

for $t = 1,\ldots,T$, where $\boldsymbol{y}_{1:t-1}$ is a matrix which contains the past values of $\boldsymbol{y_t}$ up to time $t-1$. The vector of time-varying parameters $\boldsymbol{\theta}_t$ only depends on $\boldsymbol{y}_{1:t-1}$ and a set of static additional parameters $\boldsymbol{\xi}$ which can be written as $\boldsymbol{\theta}_t \equiv \boldsymbol{\theta}(\boldsymbol{y}_{1:t-1},\boldsymbol{\xi})$ for all t.

The main characteristic of GAS models is the formulation of the the mechanism for updating the time-varying parameter $\boldsymbol{\theta}_t$:

$$\boldsymbol{\theta}_{t+1} \equiv \boldsymbol{\kappa} + \boldsymbol{A}\boldsymbol{s}_t + \boldsymbol{B}\boldsymbol{\theta}_t, \tag{1.12}$$

where $\boldsymbol{\kappa}$ is a vector of constants, $\boldsymbol{A}$ is a diagonal matrix of coefficients having proper dimensions and $\boldsymbol{B}$ is a diagonal matrix of autoregressive parameters having the same dimension as $\boldsymbol{A}$. $\boldsymbol{\kappa}$, $\boldsymbol{A}$ and $\boldsymbol{B}$ are collected in the set $\boldsymbol{\xi}$. $\boldsymbol{s}_t$ is the scaled score, which depends on the past observations and the time-varying parameters:

$$\boldsymbol{s}_t \equiv \boldsymbol{S}_t(\boldsymbol{\theta}_t)\boldsymbol{\nabla}_t(\boldsymbol{y}_t,\boldsymbol{\theta}_t), \tag{1.13}$$

where $\boldsymbol{S}_t$ is the scaling function and $\boldsymbol{\nabla}_t$ is the score, which can be calculated as a first derivative of the logarithm of the conditional observation density with respect to $\boldsymbol{\theta}_t$ for $\boldsymbol{\theta}_t$ at time t. It can be formally written as

$$\boldsymbol{\nabla}_t(\boldsymbol{y}_t,\boldsymbol{\theta}_t) \equiv \frac{\partial \log p(\boldsymbol{y}_t;\boldsymbol{\theta}_t)}{\partial \boldsymbol{\theta}_t}. \tag{1.14}$$

The scaling matrix $\boldsymbol{S}_t$ is a positive definite matrix. The variance of the score $\boldsymbol{\nabla}_t$ is defined to be the information matrix $\mathcal{I}_t$ and Creal et al. (2013) suggested to set the scaling matrix using the Fisher information matrix. Additional parameter $\gamma > 0$ is added:

$$\boldsymbol{S}_t(\boldsymbol{\theta}_t) \equiv \mathcal{I}_t(\boldsymbol{\theta}_t)^{-\gamma}, \tag{1.15}$$

with

$$\mathcal{I}_t(\boldsymbol{\theta}_t) \equiv \mathsf{E}_{t-1}\Big[\boldsymbol{\nabla}_t(\boldsymbol{y}_t,\boldsymbol{\theta}_t)\boldsymbol{\nabla}_t(\boldsymbol{y}_t,\boldsymbol{\theta}_t)'\Big]. \tag{1.16}$$

The expectation is taken with respect to the conditional distribution of $\boldsymbol{y}_t$ given $\boldsymbol{y}_{1:t-1}$. There are three most common sets of the parameter γ which lead to different GAS models. The parameter is fixed and obviously, it scales the conditional score $\boldsymbol{\nabla}_t$. That is why $\boldsymbol{s}_t$ can be also called the scaled score. When $\gamma = 0$, the scaling matrix $\boldsymbol{S}_t$ represents the identity matrix of appropriate size and therefore, there is no scaling. When $\gamma = 1$, the scaling matrix equals to the inverse information matrix. An finally, when $\gamma = \frac{1}{2}$, the conditional score is multiplied by the the square root of the information matrix $\mathcal{I}_t$ and the additional moment condition $\mathsf{VAR}_{t-1}[\boldsymbol{s}_t] = \boldsymbol{I}$ can be easily derived.

The vector $\boldsymbol{\kappa}$ can be interpreted as the level of the process, the matrix $\boldsymbol{B}$ controls for the persistence of the process and the matrix $\boldsymbol{A}$ can be comprehended as the step of the update. It basically controls the impact of scaled score on the updated vector of the time-varying parameters. And therefore, $\boldsymbol{s}_t$ indicates the direction in which the vector of time–varying parameters $\boldsymbol{\theta}_t$ is updated.

In practice, the space of the vector of time-varying parameters $\boldsymbol{\theta}_t$ is often restricted. However, the vector in the model specification in Equation (1.12) has a linear specification and therefore, it is unbounded. For example, the vector needs to be positive for the Student's t distribution. The standard solution is using the linking function $\Lambda(\cdot)$ which maps $\boldsymbol{\theta}_t$ and ensure the linear dynamic specification:

$$\boldsymbol{\theta}_{t+1} \equiv \Lambda(\tilde{\boldsymbol{\theta}}_{t+1}), \tag{1.17}$$

$$\tilde{\boldsymbol{\theta}}_{t+1} \equiv \boldsymbol{\kappa} + \boldsymbol{A}\tilde{\boldsymbol{s}}_t + \boldsymbol{B}\tilde{\boldsymbol{\theta}}_t. \tag{1.18}$$

The Jacobian matrix of the linking function $\Lambda(\cdot)$ which is compounded of all the first partial derivatives can be calculated:

$$\mathcal{J}(\tilde{\boldsymbol{\theta}}_t) \equiv \frac{\partial\Lambda(\tilde{\boldsymbol{\theta}}_t)}{\partial\tilde{\boldsymbol{\theta}}_t} \tag{1.19}$$

and it is used for evaluating the new information matrix $\tilde{\mathcal{I}}_t(\tilde{\boldsymbol{\theta}}_t)$ and the score as well:

$$\widetilde{\boldsymbol{\nabla}}_t(\boldsymbol{y}_t,\tilde{\boldsymbol{\theta}}_t) = \mathcal{J}(\tilde{\boldsymbol{\theta}}_t)'\boldsymbol{\nabla}_t(\boldsymbol{y}_t,\boldsymbol{\theta}_t), \tag{1.20}$$

$$\tilde{\mathcal{I}}_t(\tilde{\boldsymbol{\theta}}_t) = \mathcal{J}(\tilde{\boldsymbol{\theta}}_t)'\mathcal{I}_t(\boldsymbol{\theta}_t)\mathcal{J}(\tilde{\boldsymbol{\theta}}_t). \tag{1.21}$$

Using a proper linking (mapping) function and the associated Jacobian matrix can transform almost any non-linear constraint. Then the vector of static parameters $\boldsymbol{\xi} \equiv (\boldsymbol{\kappa},\boldsymbol{A},\boldsymbol{B})$ can be estimated by maximizing the log-likelihood function.

Probability distribution specification

Cont (2001) pointed out that the financial returns follow some stylized statistical properties. Especially, the distribution is (usually left) skewed and fat tailed which is caused by uneven shocks to conditional variance. Therefore, the fat tailed distribution is more suitable than the Gaussian distribution. Furthermore, the presence of volatility clustering leads to the time-varying variance. The expected value of returns supposes to be zero by nature and therefore, it does not make sense to set the location parameter time-varying. In case of estimating the model with time-varying location, the estimated coefficient should not result to be statistically significant. Time-varying skewness and shape can lead to more suitable distribution but not necessarily. These properties justify testing only few common distributions in this thesis: Student's t, skew-Student's t and Gaussian as a benchmark. More about the setting of time-varying parameters is discussed later in Section 4.2.

The score function for each distribution differs. The notation is adopted from Ardia et al. (2018). Let's assume that r_t is a logarithmic return at time t, it follows conditional skew-Student's t distribution ($r_t|\mathcal{I}_{t-1} \sim \mathcal{SKST}$) with four parameters: location, scale, skewness and shape and only scale (volatility) is set to be time-varying parameter:

$$r_t|\mathcal{I}_{t-1} \sim \mathcal{SKST}(r_t; \mu,\sigma_t,\zeta,\nu), \tag{1.22}$$

where location $\mu \in \mathbb{R}$, scale $\sigma_t > 0$, skewness $\zeta > 0$ and shape $\nu > 2$. To ensure the positiveness of the volatility parameter, the vector of time-varying parameters is set to $\boldsymbol{\theta}_t \equiv \theta_t \equiv \log \sigma_t$ and the vector of static parameters $\boldsymbol{\psi} \equiv (\mu,\zeta,\nu)$. The corresponding score s_t can be obtained from the logarithmic density function f evaluated in r_t:

$$\log f_{\mathcal{SKST}}(r_t; \mu,\sigma_t,\zeta,\nu) = \log g + \log k + c - \log \sigma_t - \frac{\nu+1}{2} \log \left[1 + \frac{\left[(\frac{r_t-\mu}{\sigma_t})k + m \right]^2}{(\nu-2)(\zeta_t^*)^2} \right], \tag{1.23}$$

where

$$m \equiv \mu_1 \Big(\zeta - \frac{1}{\zeta} \Big) \tag{1.24}$$

$$k \equiv \sqrt{(1-\mu_1^2)\Big(\zeta^2 + \frac{1}{\zeta^2}\Big) + 2\mu_1^2 - 1} \tag{1.25}$$

$$g \equiv \frac{2}{\zeta + \frac{1}{\zeta}} \tag{1.26}$$

$$c \equiv \frac{1}{2}[-\log(\nu - 2) - \log \pi] + \log \Gamma\Big(\frac{\nu + 1}{2}\Big) - \log \Gamma\Big(\frac{\nu}{2}\Big), \tag{1.27}$$

with

$$\mu_1 \equiv \frac{2\sqrt{\nu - 2}}{(\nu - 1)} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})\Gamma(\frac{1}{2})}, \tag{1.28}$$

$$\zeta_t^* \equiv \zeta^{I\{z_t \geq 0\} - I\{z_t \leq 0\}}, \tag{1.29}$$

$$z_t \equiv \Big(\frac{r_t - \mu}{\sigma_t}\Big)k + m \tag{1.30}$$

where $I\{\cdot\}$ is the indicator function and Γ is the gamma function. Then the partial derivative of the log density function is taken with respect to θ_t and the resulting score s_t for the skew-Student's t is given by

$$s_t \equiv \frac{\partial \log f_{\mathcal{SKST}}}{\partial \theta_t} = \Big(\frac{z_t(\nu + 1)(z_t - m)}{(\zeta_t^*)^2(\nu - 2) + z_t^2} - 1\Big). \tag{1.31}$$

The Student's t score can be derived easily. Let's assume that $\zeta = 1$ since Student's t has one parameter (skewness) less, then the whole expression in m in Equation (1.24) became $m = 0$ and the score is:

$$s_t = \Big(\frac{z_t^2(\nu + 1)}{(\nu - 2) + z_t^2} - 1\Big). \tag{1.32}$$

Going even further to the Gaussian distribution which has only two parameters (location and scale) and no shape, when $\nu \to \infty$, the score for the Gaussian distribution resulted in:

$$s_t = z_t^2 - 1. \tag{1.33}$$

GARCH as a special case of GAS models

Creal et al. (2013) show that an appropriate choice of the scaling matrix $\boldsymbol{S}_t$ can lead to already known observation-driven models. A typical example is an equality of GAS(1,1) and standard GARCH(1,1).

Let a basic model $y_t = \sigma_t \varepsilon_t$ have Gaussian disturbance ε_t with zero mean and unit variance and time-varying standard deviation σ_t. If the scaling matrix $\boldsymbol{S}_t$ is equal to the inverse information matrix $\boldsymbol{S}_t(\boldsymbol{\theta}_t) = \mathcal{I}_t(\boldsymbol{\theta}_t)^{-1}$ at time t and the vector of time-varying parameters is equal to the time-varying variance $\boldsymbol{\theta}_t = \sigma_t^2$, the GAS(1,1) can be reduced to

$$\sigma_{t+1}^2 = \kappa + A(y_{t+1}^2 - \sigma_t^2) + B\sigma_t^2, \tag{1.34}$$

where κ, A and B are parameters to be estimated. The reduction arose very intuitively. In Equation (1.14) the score is expressed as a derivative of the logarithm of the probability density function. For the Gaussian distribution the logarithm is expressed as

$$\log p(y_t; \sigma_t^2) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log \sigma_t^2 - \frac{1}{2}\frac{y_t^2}{\sigma_t^2}. \tag{1.35}$$

Then, the first and second derivatives of this expression with respect to σ_t^2 lead to

$$\begin{aligned}
\frac{\partial \log p(y_t; \sigma_t^2)}{\partial \sigma_t^2} &= -\frac{1}{2}(\sigma_t^2)^{-1} + \frac{y_t^2}{2}(\sigma_t^2)^{-2} \\
&= \frac{y_t^2 - \sigma_t^2}{2(\sigma_t^2)^2}, && (1.36) \\
\frac{\partial^2 \log p(y_t; \sigma_t^2)}{\partial (\sigma_t^2)^2} &= \frac{1}{2}(\sigma_t^2)^{-2} + \frac{y_t^2}{2}(-2)(\sigma_t^2)^{-3} \\
&= \frac{\sigma_t^2 - 2y_t^2}{2(\sigma_t^2)^3}. && (1.37)
\end{aligned}$$

The variance of the score is defined to be an information matrix which is the second derivative of the score and also the expected value of the second moment. If the scaling matrix is equal to an inversion of the information matrix, the expression can be step by step modified to:

$$\begin{aligned}
S_t = \mathcal{I}_{t-1}^{-1} &= -\mathbb{E}_{t-1}\left[\frac{\partial^2 \log p(y_t; \sigma_t^2)}{\partial (\sigma_t^2)^2}\right]^{-1} \\
&= -\mathbb{E}_{t-1}\left[\frac{\sigma_t^2 - 2y_t^2}{2(\sigma_t^2)^3}\right]^{-1} \\
&= -\left[\frac{\sigma_t^2 - 2\sigma_t^2}{2(\sigma_t^2)^3}\right]^{-1} \\
&= -\left[\frac{-\sigma_t^2}{2(\sigma_t^2)^3}\right] \\
&= 2(\sigma_t^2)^2. && (1.38)
\end{aligned}$$

If the resulting expressions in Equation (1.37) and Equation (1.38) are substituted to the updating mechanism of the GAS model and the static parameters are reparametrized, the standard GARCH(1,1) model is obtained:

$$
\begin{aligned}
\sigma_{t+1}^2 &= \kappa + AS_t(\sigma_t^2)\frac{\partial \log p(y_t;\sigma_t^2)}{\partial \sigma_t^2} + B\sigma_t^2 \\
&= \kappa + A2\sigma_t^4\frac{y_t^2-\sigma_t^2}{2\sigma_t^4} + B\sigma_t^2 \\
&= \kappa + A(y_t^2-\sigma_t^2) + B\sigma_t^2 && (1.39)\\
&= \kappa^* + A^* y_t^2 + B^*\sigma_t^2, && (1.40)
\end{aligned}
$$

where can be noticed that $\kappa^* = \kappa$, $A^* = A$ and $B^* = B - A$. These coefficients are unknown and need to be estimated with some conditions to fulfill stationarity. Equation (1.39) corresponds to GAS model in Equation (1.34) and Equation (1.40) corresponds to GARCH model. QED.

Moreover, Creal et al. (2013) also show that if the distribution of u_t follows the Student's t with the unit variance and ν degrees of freedom, the model for GAS(1,1) changes to

$$
\boldsymbol{\theta}_{t+1} = \boldsymbol{\kappa} + \boldsymbol{A}\Big(1+3\nu^{-1}\Big)\left(\frac{(1+\nu^{-1})}{(1-2\nu^{-1})(1+\nu^{-1}\boldsymbol{y}_{t+1}^2/(1-2\nu^{-1})\boldsymbol{\theta}_t)}\boldsymbol{y}_{t+1}^2 - \boldsymbol{\theta}_t\right) + \boldsymbol{B}\boldsymbol{\theta}_t. \quad (1.41)
$$

Notice that if $\nu^{-1} = 0$, the equation reduces to the form in Equation (1.34) and the distribution is again a Gaussian. A different impact of both updating mechanisms and an important contribution of GAS models can be seen with an example of the fat-tailed Student's t distribution and Equation (1.41) for the GARCH model. If an observation y_t gains an extreme value as long as ν is finite, the variance increases moderately comparing to GARCH models.

Maximum likelihood estimation

As Ardia et al. (2018) or Creal et al. (2008) state, the vector of time-varying parameters $\boldsymbol{\theta}_t$ is perfectly predictable given the past information and the static parameter vector $\boldsymbol{\xi}$. Moreover, the log-likelihood function is relatively easily evaluated. That is considered as a very useful property of the observation-driven models which also simplifies the application of this approach. The maximization problem for a sample of T realizations of $\boldsymbol{y}_t$ can be formulated as

$$
\tilde{\boldsymbol{\xi}} \equiv \arg\max_{\boldsymbol{\xi}} \boldsymbol{\ell}(\boldsymbol{\xi}; \boldsymbol{y}_{1:T}) \quad (1.42)
$$

where $\boldsymbol{y}_{1:T}$ contains all the past observations from time 1 till T and the log-likelihood function has a form of

$$\boldsymbol{\ell}(\boldsymbol{\xi}; \boldsymbol{y}_{1:T}) \equiv \log p(\boldsymbol{y}_1; \boldsymbol{\theta}_1) + \sum_{t=2}^{T} \log p(\boldsymbol{y}_t; \boldsymbol{\theta}_t) \tag{1.43}$$

and $\boldsymbol{\theta}_1 \equiv (\boldsymbol{I} - \boldsymbol{B})^{-1}\boldsymbol{\kappa}$. Moreover, for $t > 1$, $\boldsymbol{\theta}_t \equiv \boldsymbol{\theta}(\boldsymbol{y}_{1:t-1},\boldsymbol{\xi})$.

1.4 Parametric, Semi-parametric and Non-parametric Methods

Methods can be also divided into parametric, semi-parametric and non-parametric approaches (Huang et al., 2010), (Gill et al., 1989). The models described above and applied in this thesis are called parametric. Parametric models demand to completely specify some conditional probability distribution of returns, which is also their disadvantage. Initial assumption of a specific distribution increases the possibility of misspecification. Nonetheless, it simplifies the evaluation and leads to a straightforward calculation of one-step ahead predictions of VaR (Gao et al., 2016). Manganelli et al. (2004) mentioned three sources of possible misspecification. First problem can appear during the specification of the variance equation, the second can arise from assuming a particular distribution, which is then used for the likelihood optimization and the last one is related to the standardized residuals, which may not be independent and identically distributed.

On the other hand, a big advantage of the semi-parametric models is their elimination of assuming specific conditional distribution. These models have elements of both parametric and non-parametric approach and since these models still contain some static parameters, the number of estimated parameters increases (Patton et al., 2019). However, the main disadvantage is in the specification of quantile behavior over time. An example can be CAViaR models (Huang et al., 2010).

The last group called non-parametric models does not require to assume any probability distribution in advance. It is determined from the data applying historical simulations and the parameters are estimated a priori. The quantiles are then counted empirically based on rolling windows. The price returns are sorted in ascending order and the desired quantile is represented by the α % value (Manganelli et al., 2004). An example of these models can be histogram, kernel density estimation (A. Harvey et al., 2012), non-parametric regression and others.

1.5 Value at Risk and Expected Shortfall

A crucial piece of information about the returns is estimating the likelihood of a possible loss. Financial mathematicians call these tools risk measures and they use them to evaluate the risks that a financial institution goes through. One of the most favorite and the oldest is Value at Risk (VaR) shortly after followed by Expected Shortfall (ES). These constitute the two leading risk measures (Ardia et al., 2018).

Value at Risk became used by investment companies in the late 1990's and it supposes to measure the risk of market assets. This method was introduced by J.P.Morgan, a founder of an American bank J.P.Morgan & Co., at the risk management conference in 1993. Due to significant success, the whole methodology was published titled *Risk Metrics* (Morgan et al., 1996). VaR measures the largest expected portfolio loss over a particular time horizon at a given probability level assuming normal market conditions (Linsmeier et al., 2000). It can also be comprehended as an estimate of the largest loss that could occur with 100α % probability based on the already known losses of a certain period of time. A probability level is also referred to a risk level and it is typically set to one of the following: $\alpha \in \{0.01,0.05\}$ (Ardia et al., 2018).

Despite being widely used by all banks and regulators, VaR was refused by science since it did not fulfill one of the axioms of coherence (Acerbi et al., 2002). These axioms were created to define what is a good risk measure. However, they seem to be too strict and for a while, none of the measures did manage to fulfill all of them. But thinking about the impact of breaking some of the axioms, it became evident that only meeting all of them could lead to a good quality risk measure and breaking some of them to a wrong or paradoxical results.

VaR is simply a threshold of possible losses and it is indifferent to the size of losses beyond (Acerbi et al., 2002). It violates the second axiom of sub-additivity. The coherence axioms were promoted by Acerbi et al. (2002) and following the paper, if a function $\rho : V \to \mathbb{R}$ is

1. *monotonous:* $X \in V,\ \ X \geq O \implies \rho(X) \leq 0,$
2. *sub-additive:* $X,Y,X+Y \in V \implies \rho(X+Y) \leq \rho(X) + \rho(Y),$
3. *positively homogeneous:* $X \in V,h > 0,hX \in V \implies \rho(hX) = h\rho(X),$
4. *translation invariant:* $X \in V,a \in \mathbb{R} \implies \rho(X+a) = \rho(X) - a,$

then it can be called a risk measure where V is a set of real-valued random variables, X and Y are possible portfolios and h and a are constants. The second axiom expresses that a sum of portfolios should lead to maximally the same risk as a portfolio made of these sub-portfolios. A situation with strict inequality in this axiom is known as a risk diversification. The point is that having a portfolio made of more than one financial assets diversifies the risk of a loss and therefore reduces it. In the figurative sense, if there is a bank made of several branches and the analytics estimate the risk of each using a risk measure, then they should be confident that the overall risk is the same or less. If it is more, it apparently violates the second axiom. Therefore, only fulfillment of all axioms leads to the correct results.

Four years later, a new risk measure was introduced by Rockafellar et al. (2002) initially called Conditional Value at Risk, nowadays more known as Expected Shortfall (ES) and it was proven to be a coherent indicator (Delbaen, 2002). ES measures the expected loss in the 100α % worst cases where usually $\alpha \in \{0.01, 0.05\}$ and therefore, it takes into consideration the shape of the distribution tail (Acerbi et al., 2002). But in the same time it does not consider only the worst case that can occur but the average of the worst cases. ES is a very simple concept and moreover, it answers the natural questions about the portfolio risk and the effort of banks in switching to the new concept does not require a whole new computational procedure. In the literature, ES is sometimes also called Average Value at Risk (AVaR) or Expected Tail Loss (ETL).

Formally, the risk measure estimation requires firstly an accurate estimate of the conditional distribution of the future returns (Ardia et al., 2018). Then, the VaR at time t for a risk level α can be written

$$VaR_t(\alpha) \equiv F^{-1}(\alpha; \boldsymbol{\theta}_t, \boldsymbol{\xi}), \tag{1.44}$$

where F^{-1} denotes the inverse of the continuous cumulative density function which is actually the quantile function, $\boldsymbol{\theta}_t$ is a vector of time-varying parameters and $\boldsymbol{\xi}$ is a vector of additional static parameters. As it seems to be intuitive, the VaR is simply the 100α % quantile of the return distribution at time t. The ES is then the average of the worst cases formally written as

$$ES_t(\alpha) \equiv \frac{1}{\alpha} \int_{-\infty}^{VaR_t} z dF(z, \boldsymbol{\theta}_t, \boldsymbol{\xi}). \tag{1.45}$$

There are some interesting properties. For example, as α decreases, the ES increases. Also the ES is always greater than or equal to the corresponding VaR (meaning at the same risk level α).

1.6 Model Comparison

An essential part of the return modeling is the comparison of the estimated models and selecting the best according to some criteria. The most common approach is to use information criteria and two major representatives are Akaike information criterion and Bayesian information criterion. These can be used in the first step of the model comparison. An interesting part can be comparing the estimates of the static parameters and see whether they change in different time periods or if they stay statistically same.

An advanced approach is to evaluate the accuracy of VaR and ES predictions. It requires to use rolling windows to perform backtesting and evaluating the Dynamic Quantile (DQ) test of the correct coverage. Then the average loss functions are computed to compare the estimates of VaR and ES. Description of these approaches follows.

Information criteria

Akaike criterion (AIC) assesses the relative quality of the model (Wagenmakers et al., 2004). AIC for a model h is given by

$$AIC_h \equiv -2\ell_h + 2k, \tag{1.46}$$

where ℓ is the value of the log-likelihood function and k is the total number of parameters in the model. The best model is the one having the minimum value of AIC among all other models.

Burnham et al. (2004) presented how to compare more models using the AIC criterion and especially how to order them since the individual AIC values are not directly interpretable. They are effected by sample sizes and contain arbitrary constants. Therefore, the values are rescaled for each model h by

$$\Delta_h = AIC_h - AIC_{min}, \tag{1.47}$$

where AIC_{min} is the lowest AIC value among the comparing models and therefore, $\Delta_h = 0$ for the best model. The value of Δ_h can be interpreted as the information loss experienced in the case of using the model h rather than the best model. Such an interpretation is intuitive, meaningful and the comparison is straightforward and quick.

Values with the differences $\Delta_h \leq 10$ can be seen as indifferent and the models are considered to be equally good. On the other hand, values with the differences $\Delta_h \geq 10$ can lead to a significant difference between the models and the model h is seen as the worse model.

Bayesian information criterion (BIC) was introduced in Schwarz et al. (1978) and it is closely related to AIC. The value of likelihood is possible to be increased by adding more parameters to the model. Therefore, AIC could choose an overfitted model to be the best option more easily than BIC, which is proposed to penalize models with more parameters. The BIC for a model h is defined:

$$BIC_h \equiv -2\ell_h + \log(N)k, \tag{1.48}$$

where N is the number of observations. For the time series it holds that $N = T$, where T is the length of the time series. As for AIC, the best model is the one having the minimum value of BIC among all other models.

Comparison of the estimated coefficients

Evaluated models can be also compared from a different point of view. The estimates of the static parameters for each index can be compared over different time periods. This could lead to an idea whether these parameters actually change in time or stay constant. The values of pairs of coefficients can be tested using Z-test in a formulation provided by Paternoster et al. (1998). For a pair of coefficients from two different models, the formula for the test statistic is given by

$$Z = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\sqrt{SE(\hat{\beta}_1)^2 + \hat{\beta}_2)^2}}, \tag{1.49}$$

where $\hat{\beta}_1$ and $\hat{\beta}_2$ are the values of estimated coefficients for model 1 and model 2, respectively and $SE()$ are the corresponding standard errors. Under the null hypothesis the coefficients are equal and the test is based on the Gaussian distribution.

Backtesting VaR and ES

Backtesting verifies the precision of VaR and ES predictions. It refers to a procedure which uses historical data for testing the adequacy of the predictive model (P. Christoffersen, 2010) and controls the correctness of the unconditional and conditional left-tail of the returns' distribution (Ardia et al., 2018). The sample of the length T is divided into two parts: in-sample of the length m and out-of-sample of the length $T-m$. Firstly, the model is estimated with the in-sample to forecast the observation $m+1$ (i.e. one-step ahead forecast). Then the original observation $m+1$ is included into the sample and the model is estimated again to make another one-step ahead forecast for the observation $m+2$ and so on.

This approach is called rolling and expanding windows and it is described for example in Vasconcelos (2017). According to whether the observations are kept in the sample or dropped in each iteration to keep the sample of the same size, the windows are called expanding or rolling, respectively. The rolling windows are preferred due to their better fit in both small and large samples in case that both windows seem to be identically accurate.

It can be also set how often the model coefficients should be re-estimated meaning that the model does not have to be estimated in every iteration but for example, every fifth iteration. The estimated model is then used for four predictions before it is estimated again. Note that for a longer prediction step it is necessary to use simulations. The output of the rolling windows are predicted values of the length $T - m$ and they are used to calculate the values of VaR and ES.

Correct unconditional (UC) and conditional (CC) coverage is tested after obtaining the predictions for VaR and ES. The difference between UC and CC is that UC evaluates the left-tail of the unconditional distribution of returns while CC focuses on the conditional density function of the returns. Afterwards, the series of VaR exceedance is obtained:

$$d_t \equiv I\{r_r < VaR_t(\alpha)\}. \tag{1.50}$$

This series is also called the hitting series and if the model is correctly conditionally covered, d_t should be independently distributed.

Dynamic quantile (DQ) test of Engle and Manganelli (2004) is based on the joint hypothesis that firstly, the hitting series are independently distributed and secondly, that the expected proportion of exceedance is equal to the risk level α, i.e. $\mathsf{E}[d_t] = \alpha$. DQ test is then basically the traditional Wald test of the joint null hypothesis that all model parameters are equal to zero in the following linear regression:

$$Hit_t^{\alpha} = \delta_0 + \sum_{l=1}^{L} \delta_l Hit_{t-1}^{\alpha} + \delta_{L+1} VaR_{t-1}(\alpha) + u_t, \tag{1.51}$$

where Hit is a de-meaned process which the implementation of DQ test involves and it is defined as $Hit_t^{\alpha} \equiv d_t - \alpha$. If the model is correctly specified, Hit_t^{α} has zero mean and is serially uncorrelated. Therefore, the null hypothesis of DQ test can be interpreted as the correct unconditional and conditional coverage and not rejecting the null hypothesis is desired. The Wald test statistic has asymptotic Chi-square distribution with $L + 2$ degrees of freedom. Ardia et al. (2018) refer that the standard choice is to set $L = 4$ lags.

Loss functions

The final step is to choose the best performing model by ordering them according to the values of the corresponding loss functions for quantile prediction or count their ratios (Ardia et al., 2018). Quantile Loss (QL) function is very frequently applied for quantile regressions in the VaR evaluation and Fissler and Ziegel (FZL) loss function for the joint VaR and ES evaluation. The QL function for time t at risk level α is defined:

$$QL_t(\alpha) \equiv (\alpha - d_t)(r_t - VaR_t(\alpha)). \tag{1.52}$$

Ardia et al. (2018) point out that QL is an asymmetric loss function which penalizes more heavily the observations that exceeded the VaR. Quantile losses are averaged over the forecasting periods and the preferred model is the one with the lower average value. Moreover, two models (A and B) can be compared by looking at the ratio of associated average QLs. If

$$\frac{QL_A}{QL_B} < 1, \tag{1.53}$$

then model A is better than model B. Furthermore, VaR can be estimated by minimizing the average QL for a static conditional distribution over the sample. It is not possible to do the same for the second risk measure ES since there is no appropriate loss function, which is elicitable. However, Fissler et al. (2016) show that VaR and ES are jointly elicitable using FZ loss function, which is named according to the first letters of their names (Fissler and Ziegel). The loss function for v_t and e_t which correspond to VaR and ES for the minimum sample average is given by

$$\begin{aligned} FZ(r_t,v_t,e_t,\alpha,G_1,G_2) \equiv (d_t-\alpha)\left(G_1(v_t)-G_1(r_t)+\frac{1}{\alpha}G_2(e_t)v_t\right)- \\ -G_2(e_t)\left(\frac{1}{\alpha}d_t r_t - e_t\right)-G_2(e_t). \end{aligned} \tag{1.54}$$

G_1 is a weakly increasing function, G_2 is strictly positive and strictly increasing function and $G_2' = G_2$. Let's assume that VaR and ES are strictly negative and the generated loss differences are homogeneous of degree zero iff $G_1(x) = 0$ and $G_2(x) = -\frac{1}{x}$. Then the associated joint loss function FZL for time t at risk level α is formulated as

$$FZL_t^\alpha \equiv \frac{1}{\alpha ES_t^\alpha} d_t(r_t - VaR_t^\alpha) + \frac{VaR^\alpha}{ES_t^\alpha} + \log(-ES_t^\alpha) - 1 \tag{1.55}$$

for the case when $ES_t^\alpha \leq VaR_t^\alpha < 0$. FZLs are also averaged over the forecasting period and the preferred models are models with lower average values. Furthermore, the ratios of FZLs can be evaluated for two models (A and B) and if

$$\frac{FZL_A}{FZL_B} < 1, \tag{1.56}$$

then model A is better than model B considering both risk measures VaR and ES.

1.7 Stock Indices Analysis

Arlt et al. (2003) mentioned that financial time series and the returns especially have some characteristic properties, which are needed to test. One of the oldest models capturing the behavior of the asset prices is the model of martingale. It assumes that all the past prices of the assets are known and the best information for the prediction of tomorrow's price is the price today. Therefore, the martingale is given by

$$P_t = P_{t-1} + a_t, \tag{1.57}$$

where P_t is the price at time t and a_t is the martingale difference. But comparing it to the random walk, a_t does not follow the white noise. If it is assumed that a_t follows the Gaussian distribution with the zero mean and constant variance, there is an imperfection since the prices can not be negative. That is the reason why the analytics model returns and not the prices themselves. The returns of assets can be both positive and negative and if they follow log-Normal distribution then the logarithm of the returns follows the Gaussian distribution. The logarithmic returns multiplied by 100 are then written as

$$r_t = (\ln P_t - \ln P_{t-1}) \times 100. \tag{1.58}$$

The basic assumption about the financial time series is that the logarithmic returns have the Gaussian distribution with a constant mean value and constant variance. However, this does not hold for the market assets. These series are rather leptokurtic and skewed (Bernardi et al., 2015).

From now on, the term 'returns' continue to be used meaning the logarithmic returns multiplied by 100. The normality of the series can be tested using Shapiro-Wilk normality test. This test became preferred since it evinces good power properties. The sample size N must be less than 5000 but greater than 3 and it detects the deviation from normality by looking at the skewness and the kurtosis (Razali et al., 2011). For the ordered random sample with values y_i where $i = 1, \ldots, N$, the test statistic is formulated

$$W = \frac{(\sum_{i=1}^{N} a_i y_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}, \tag{1.59}$$

where $\bar{y}$ is the sample mean and the coefficients a_i are given by

$$\boldsymbol{a} = (a_1, \ldots, a_N) = \frac{(\boldsymbol{m}'\boldsymbol{V}^{-1})}{(\boldsymbol{m}'\boldsymbol{V}^{-1}\boldsymbol{V}^{-1}\boldsymbol{m})^{1/2}}. \tag{1.60}$$

The vector $\boldsymbol{m}$ contains the expected values of the order statistics of independent and identically distributed random variables sampled from the standard Gaussian distribution and $\boldsymbol{V}$ is the covariance matrix of those order statistics (Razali et al., 2011). The statistic value W lies between zero and one and the values closer to zero rather lead to the rejection of normality. Note that for the time series holds that $N = T$ where T is the length of the time series.

Another characteristic is stationarity, which means that the statistical properties that generate the stochastic process do not change over time. This is important due to its simplification of the analysis and the possibility to predict the series. Arlt et al. (2003) point out that stationarity is not a property of the realized time series but a property of the stochastic process.

Stationarity can be strong or weak. Strong stationarity means that the probability behavior of the random stochastic process is time invariant. If the mean of the process is constant, the variance is finite and constant and the covariance depends only on the difference between two times, then the process is called weakly stationary. In practice, only weak stationarity is used because it is quite easy to estimate the first two moments and the term stationarity is commonly understood as the weak version.

A commonly applied test is the Augmented Dickey-Fuller (ADF) test which is an extended version of the Dickey-Fuller test (Qiu, 2015). The difference is in handling more complicated time series and larger input data. These tests equal for the lag order $p = 2$. ADF test is applied to the model

$$\Delta y_t = \mu + \beta t + \rho y_{t-1} + \beta_1 \Delta y_{t-1} + \ldots + \beta_{p-1} \Delta y_{T-p+1} + u_t, \tag{1.61}$$

where Δ denotes the first order differences of y_t, μ is a constant term, β is the coefficient for the time trend, p is the lag order and T is the length of the series. The ADF statistic is then given by

$$ADF = \frac{\hat{\rho}}{SE(\hat{\rho})}, \tag{1.62}$$

where $\hat{\rho}$ is the estimated coefficient and SE is the corresponding standard error. ADF statistic is a negative number and more negative values stand for stronger rejection of the null hypothesis. The alternative hypothesis is the stationarity.

2. GAS and RUGARCH Packages in R

One of the very common programming languages for statistical data analysis is the R language. It was created by Ross Ihaka and Robert Gentleman at the University of Auckland in New Zealand (Ihaka et al., 1996) and it continues to be developed by R Core Team. It first appeared in 1993. R is a free software and it runs on the base of user-created packages which extends the statistical techniques and graphical tools.

Beside others, two main packages were used in this thesis for modeling the price returns. The first one is *GAS*, which was published recently in 2016 and it can be used for both univariate and multivariate GAS models (Catania et al., 2019). The second one is *RUGARCH*, which was created in 2011 and it contains functions that help with univariate GARCH models (Ghalanos et al., 2019).

In this Section, the functions and properties of both packages will be shortly described following the R documentation (Catania et al., 2019), (Ghalanos et al., 2019) or (Ardia et al., 2018).

2.1 Package *GAS*

This package was designed to cover estimation, simulation and forecasting GAS models and contains numerous functions. Basically, functions for univariate modeling has a prefix Uni and for multivariate Multi. They return an object with a prefix u for univariate and m for multivariate approach. These functions are GASFit for estimating the vector of static parameters, GASFor for one-step ahead forecast (prediction) of the time series, GASRoll for one-step ahead rolling forecast and GASSim for simulations.

UniGASSpec :

Firstly, the specification of the model must be set. There is a special function UniGASSpec which returns an object uGASSpec with the model setting. This object enters the fitting function afterwards. There are three arguments including the type of the distribution (Dist), scaling type (ScalingType) and the list of time-varying parameters (GASPar).

Dist is a character indicating the conditional distribution. By default (Dist = norm), i.e. the Gaussian distribution. Available distributions together with their properties can be displayed running command DistInfo() and the setting for used distributions are shown in Table 2.1.

ScalingType is a character representing the scaling mechanism for the scaled score. There are three most common sets, which were described in Section 1.3 and which are related to the additional parameter γ formulated in Equation 1.15. Therefore, ScalingType = Identity is used in case of $\gamma = 0$, Inv is for $\gamma = 1$ and InvSqrt is for $\gamma = \frac{1}{2}$. Some distributions do not support all the scaling types. The list of available ones is attached to Table 2.1. The default value is ScalingType = Identity, that is no scaling occurs.

GASPar is a list of all parameters containing information about which parameters are set to be time-varying. Maximum of four parameters can be taken into account according to the type of distribution as can be seen in Table 2.1. Parameters are location, scale, skewness and shape referring to location parameter, scale parameter, parameter controlling the skewness and shape parameter. However, it is recommended to check their statistical meaning in DistInfo() or the literature. For example, parameter shape refers to the degrees of freedom for Student's t distribution or parameter location represents the usual intensity rate parameter for Exponential distribution. The default values are set to GASPar = list(location = FALSE, scale = TRUE, skewness = FALSE, shape = FALSE).

It is necessary to assign a name to the specification settings, GASSpec is often used. Then it can be displayed by calling the object GASSpec and see what was set.

Distribution	Label	Parameters	#	Scaling type
Gaussian	norm	location, scale	2	Identity, Inv, InvSqrt
Skew-Gaussian	snorm	location, scale, skewness	3	Identity
Student's t	std	location, scale, shape	3	Identity, Inv, InvSqrt
Skew-Student's t	sstd	location, scale, skewness, shape	4	Identity

Table 2.1: Statistical distribution and their available settings in R

UniGASFit :

After the specification setting is ready, the maximum likelihood estimation can be run. The function is called UniGASFit and it returns an object uGASFit with several lists. Total of four arguments enter the function but only two are required: GASSpec and data.

GASSpec is an object of the class uGASSpec which was created using the function UniGASSpec. Argument data is a numeric vector containing the time series. Objects of class ts, xts and zoo are also possible to be used.

fn.optimizer is by default equal to fn.optim and it is a function where the optimizer is set. BFGS is the optimization method for fn.optim. It is an iterative method and very famous for solving unconstrained non-linear problems (Dai, 2002). If this argument wants to be changed, a new function needs to be created (see the R documentation for more details (Catania et al., 2019)). Starting values for the optimizer are chosen in two steps. Firstly, the static version is estimated with $\boldsymbol{A} = \mathbf{0}$ and $\boldsymbol{B} = \mathbf{0}$ and on this basis, the initial value for the intercept is set. Secondly, a grid search for the coefficients in $\boldsymbol{A}$ and $\boldsymbol{B}$ is executed.

Compute.SE is a logical argument signifying whether the asymptotic standard errors should be computed. By default Compute.SE = TRUE.

The output can be seen by calling the assigned object and it is divided into six parts: 1) summary of the model specification, 2) estimated coefficients along with the standard errors and significants, 3) values of the unconditional parameters meaning the value of the time-varying parameters, 4) AIC and BIC, number of estimated parameters and the evaluated log-likelihood at its optimum, 5) convergence and 6) the computation time.

The estimated coefficients called kappa correspond to the elements of vector $\boldsymbol{\kappa}$ and coefficients a, resp. b represent the diagonal elements of matrix $\boldsymbol{A}$, resp. $\boldsymbol{B}$. If the parameter is not set to be time-varying, corresponding coefficients a and b are not estimated neither displayed.

Several methods of extracting objects of interest are also implemented in the package. These are coef for the estimated coefficients, getMoments for the moments, plot for the basic graphs and others. The argument of these methods is object of class uGASFit.

UniGASRoll :

One-step ahead forecast is a pivotal element in the time series analysis. There is UniGASFor function in the package which predicts the future values. However, forecasting is not a goal of this thesis and this function is not used.

Nevertheless, another approach based on forecast are expanding and rolling windows. It is described in Section 1.6 in detail. The corresponding function in R is called UniGASRoll and take at least eight arguments. It return an object uGASRoll.

Argument data is simply a numeric vector of the whole time series, GASSpec is an object of the class uGASSpec created using UniGASSpec.

ForecastLength is an integer stated to the length of the out-of-sample period. By default ForecastLength = 500.

Nstart is an integer setting the length of the in-sample before the first prediction is performed. This argument is ignored when ForecastLength is supplied. Therefore, the default value is NULL.

RefitEvery is an integer which ensures how often the model coefficient are re-estimated. By default RefitEvery = 23.

RefitWindow is a character signifying the type of the windows. Possible choices are recursive for the expanding windows approach or moving for the rolling windows.

Compute.SE is a logical value, where can be set whether the asymptotic standard errors should be computed. By default Compute.SE = TRUE.

Beside other additional arguments connected to UniGASFit, *PARALLEL* package can be used to speed up the computations. Firstly, a cluster object must be created. And then, it is assigned within the argument cluster.

It is possible to extract the summarized info about the procedure, the whole history of estimated coefficients and associated moments. In case that the model is assigned to be called Roll then these information can be get by calling one of follow: Roll@Info, Roll@Forecast$Coef, Roll@Forecast$Moments.

BacktestVaR :

GAS package also provides some common tests. The first one is called BacktestDensity, which evaluates the average Negative Log Score (NLS) and weighted Continuous Ranked Probability Score (wCRPS). More about these in Gneiting et al. (2011). However, they are not used in this thesis.

Another one is BacktestVaR and this one is used for testing the correct model specification for Value at Risk (VaR). The function evaluates six tests or quantities (the shortcut used in the output in R and the references are mentioned in the parenthesis behind):

- The unconditional coverage test of Kupiec (LRuc) (Kupiec, 1995),
- The conditional coverage test of Christoffesen (LRcc) (P. F. Christoffersen, 1998),
- The Dynamic Quantile test of Engle and Manganelli (DQ) (Engle; Manganelli, 2004),
- Mean and maximum absolute deviation between the observations and the quantiles (AD),
- Average quantile loss and quantile loss series (Loss),
- Actual over Expected exceedance ratio (AE).

2.2 Package *RUGARCH*

Initially, GARCH models were fully covered in the package *RGARCH*. Later, it was split into two packages for univariate (*RUGARCH*) and multivariate models (*RMGARCH*). Package *RUGARCH* can be used for estimating, forecasting or simulations and many others. An additional ability is to run pure ARFIMA models with constant variance. These functions contain in their name arfima. The functionality is similar to *GAS* package and the package does not guarantee the convergence of the estimated model.

ugarchspec :

Firstly, the specification of the model must be set. There is a special function ugarchspec which returns an object uGARCHspec with the model setting. This object enters the fitting function afterwards. There are many arguments that can be set including the variance model, mean model, distribution of the model and others. There are three most utilized: variance.model, mean.model and distribution.model. It is necessary to assign a name to the specification settings to be able to call it for the fitting function.

Argument variance.model specifies the variance model. It is a list with four elements. First one model sets the type of the GARCH model. Valid models are sGARCH for the standard GARCH model and variance models fGARCH, eGARCH, gjrGARCH, apARCH, iGARCH and csGARCH. Second one garchOrder adjusts the GARCH order (p,q). Third element submodel is included when model = fGARCH and set the sub-model. Valid options are GARCH, TGARCH, AVGARCH, NGARCH, NAGARCH, APARCH, GJRGARCH and ALLGARCH. The last element external.regressors is a matrix which contains the external regressors which should be included in the variance equation. The whole argument is set to variance.model = list(model = sGARCH, garchOrder = c(1, 1), submodel = NULL, external.regressors = NULL) by default.

Argument mean.model is related to the specification of a mean model and it is a list. The first element armaOrder specifies the autoregressive (AR) and moving average (MA) orders. The second element include.mean is a logical argument. If include.mean = TRUE, the mean is included in the model. The third element archm is a logical argument and can be used when the ARCH volatility is required to be included in the mean model. By default archm = FALSE. The fourth element archpow can gain values $\{1; 2\}$. By default archpow = 1 which means that standard deviations will be used in the mean model. If it equals to 2, the variance will be used instead. The fifth element arfima is a logical argument and by default arfima = FALSE states for no fractional differencing in the ARMA model. The sixth element external.regressor is again a matrix which contains the external regressors which should be included in the mean equation and the last element archex is an integer specifying whether the last external regressors should be multiplied by the conditional standard deviation.

The whole argument is set to mean.model = list(armaOrder = c(1, 1), include.mean = TRUE, archm = FALSE, archpow = 1, arfima = FALSE, external.regressors = NULL, archex = FALSE).

The argument distribution.model requires a character, which describes the observation conditional density. Possible choices are norm for the Gaussian distribution, snorm for the skew-Gaussian distribution, std for the Student's t, sstd for the skew-Student's t distribution and some others (more in R documentation (Ghalanos et al., 2019)).

The remaining arguments, which can be set are for example start.pars, which accepts a list of starting parameters for the optimization calculation or fixed.pars is a list of parameters, which should be fixed during the calculation. However, both are not required.

ugarchfit :

After the specification setting is ready, the fitting procedure can be applied. The corresponding function is called ugarchfit and it returns an object of class uGARCHfit. Two arguments are required spec and data, others are set by default but they can be changed by the user.

The argument spec is an object of class uGARCHfit created using the function ugarchspec and argument data can be a numeric vector, matrix, data frame, zoo, xts, ts or irts. Other arguments help to for example change the optimization function.

The package *RUGARCH* provides numerous functions but no more functions from this package are used in this work and therefore, they are not described.

3. Data Description

This chapter covers the description of the used data along with a short history about each price index that is used. Afterwards, a rationale for the chosen time periods is provided followed by statistical analysis of the time series' corresponding to each time period.

The datasets were downloaded from Thomson Reuters Datastream (Datastream, 2019).

3.1 Financial Background

Stock market indices help to describe stock market behavior and compare the returns of investments. Each index is composed of variously selected companies, whose stock prices are used for the index calculation, which can be done in varied ways. There are many different ones, only four of them were chosen: DJIA, S&P 500, FTSE 100 and TOPIX. First two mentioned are related to the U.S. stock market, FTSE 100 assesses the market in Great Britain and TOPIX covers the Japanese market. These four are representatives of the major world stock market indices.

Open, high, low and last positions are available in the data for each day. Last (closing) position seems to be most commonly used, for example in the analysis of (Bernardi et al., 2015) or (Catania et al., 2019). Therefore, this position and its logarithmic returns multiplied by 100 are used in the thesis.

Stock index DJIA

Dow Jones Industrial Average (DJIA) is the second U.S. oldest market index after the Dow Jones Transportation Average. It is named after American journalist Charles Dow and statistician Edward Jones. DJIA was first calculated on May 26, 1896 and it is computed as a price index. The index is composed of 30 largest companies on stock exchanges in the United States. Corresponding weights are assigned to each company according to their stock prices. This causes that even large companies have small impact for their numerous share splits in the past. However, the company selection is in motion of a committee of experts led by Wall Street Journal and they choose by subjective parameters, for example American economy representation ability. Therefore, index changes are very rare. Market capitalization is not taken into account. Nevertheless, it is said to be a good representative of the U.S. Stock Market (Svoboda, 2008).

Stock index S&P 500

Standard & Poor's 500 Index begun to be published in 1943 using an acronym S&P 500 and it is commonly considered to be one of the best representative of the U.S. economy. S&P 500 is a capitalization-weighted index and it is made up of 500 companies listed on stock exchanges in the United States. It is a net price index, therefore paid dividends are not included into calculations. Individual shares are selected according to the market value and total trading volume. Analogously to DJIA, the selection of companies also is in motion of a committee of experts (Svoboda, 2008).

Stock index FTSE 100

The Financial Times Stock Exchange 100 Index (also called FTSE 100) is one of the most common indicators at the stock markets in Great Britain. It started to be measured on January 1, 1984 and its initial value was set to 1000 points. This index is composed of one hundred biggest companies within the meaning of market capitalization. The base is variable and new companies can replace the old ones when they fulfill a particular limit securing the full market value. FTSE 100 is followed by FTSE 250, which contains next 250 biggest companies (Ftse., 2003), (Svoboda, 2008).

Stock index TOPIX

Tokyo Stock Price Index is the second most important Japanese stock market index behind Nikkei 255. While Nikkei 225 works with the level of current stock prices only, TOPIX sets weights to companies according to their market capitalization. The index is composed of more than 1600 companies. The number changes because TOPIX selects companies on the basis of the free float criterion and only freely traded shares are taken into account. The index includes all shares of the first business segment of the Tokyo Stock Exchange which is much more than Nikkei 225. However, some argues that TOPIX is a better Japanese benchmark and it is a matter of time when TOPIX substitutes Nikkei 225 worldwide. The only drawback is its relatively short history (Svoboda, 2008).

The whole closing price history of all four indices in in Figure 3.1 where the horizontal axis represents the original values of the price index (no returns or logarithms). Both DJIA and S&P 500 go far to the past, while FTSE 100 and TOPIX started around 1980. However, the far historical values show stable prices until approximately 1980. Then, it begun to change rapidly. On the contrary, the plots for all four returns in Figure 3.2 evince slightly different conclusions. The volatility is higher in the far history due to the Great Depression of the 1930s followed by a significant reduction for DJIA and S&P 500. The remaining two index returns seems to maintain higher volatility. However, the Black Friday in 1987 cased larger drop for DJIA and S&P 500.

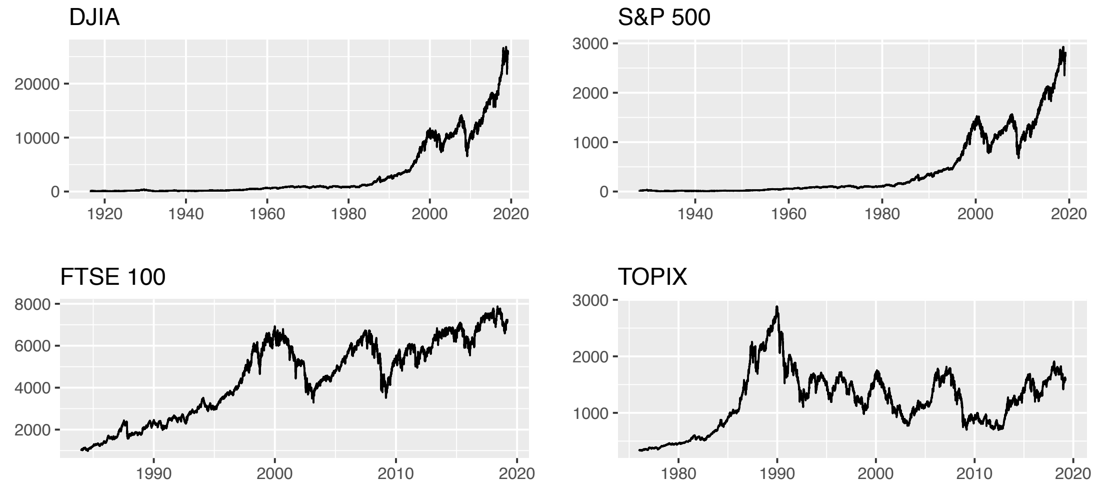

Figure 3.1: Historical closing prices for selected stock market indices

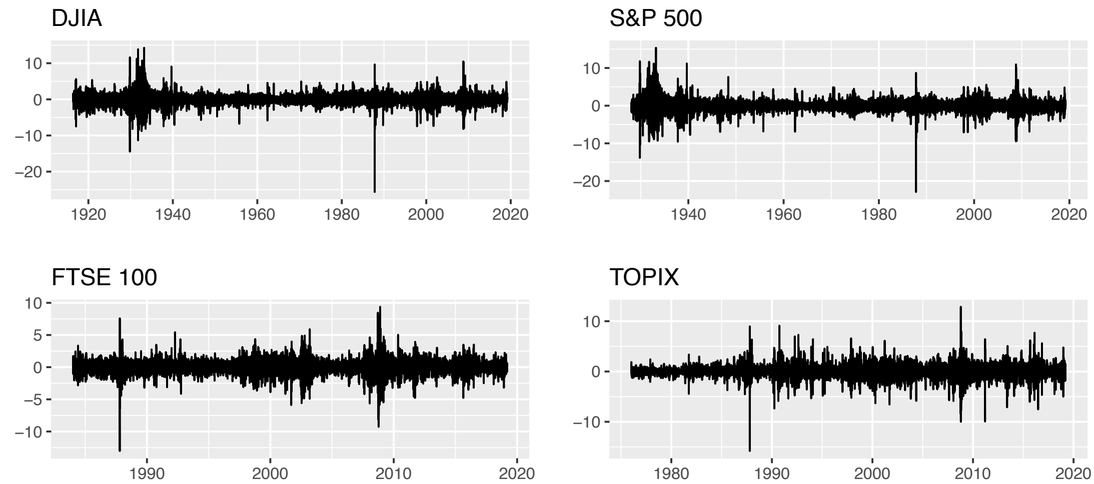

Figure 3.2: Logarithmic returns for selected stock market indices

3.2 Chosen Time Periods

Three different time periods are taken into account for the best demonstration of the differences between distributions and modeling approaches. The first period covers Black Monday, second the global financial crisis in 2007 – 2008 and third represents the last present years. Each of the chosen periods evinces different properties. The return series will face diverse shapes, which will examine the models and will lead to an overview of each model and distribution reaction.

Black Monday is dated on October 19, 1987. The stock market collapsed due to a spectacular fall of 22.61 % in the DJIA and it is the largest drop in the history of the stock market indices. The reason is the declaration of Treasury Secretary James Baker on the weekend. On October 21, 1987, the DJIA rose back thanks to the additional decline of interest rates by 10.15 % (Charles et al., 2014). Sudden significant one-day jumps influence the volatility of distributions which are sensitive to extreme values. An example of such a distribution is the Gaussian distribution, an example of the opposite one is the Student's t distribution with heavier tails. Therefore, it will be interesting to compare the results and see how both distribution behave under GAS and GARCH modeling. These properties and more for both distributions were described earlier in Section 1.3.

The first period takes approximately ten years from January 2, 1985 to December 30, 1994 and it covers 2529 days. Detailed returns are presented in Figure 3.3, where the drop is very well noticeable. In addition, the indices DJIA and S&P 500 seem to be more stable than FTSE 100 or especially TOPIX, which seems to contend with larger volatility.

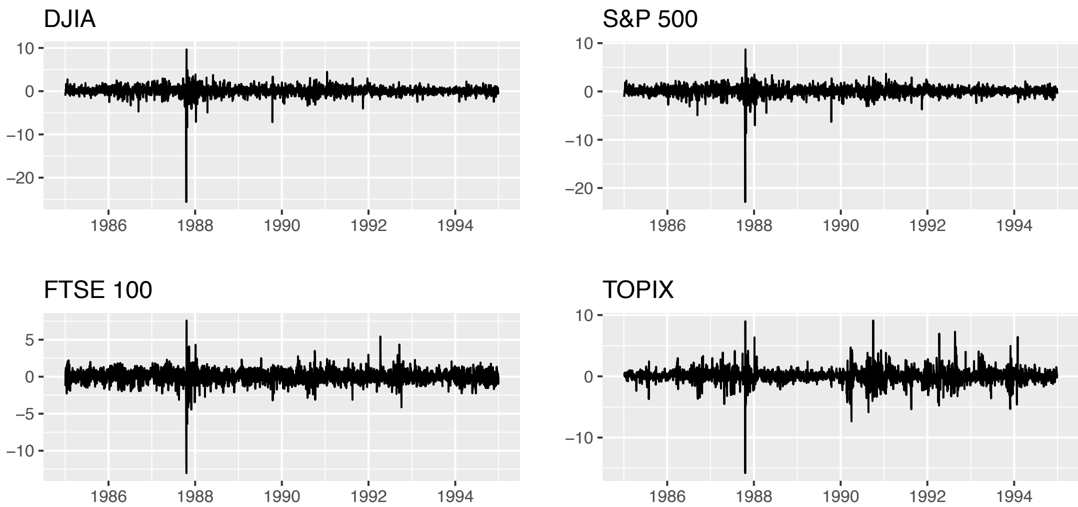

Figure 3.3: Stock market indices returns in 1985 – 1994

The global financial crisis brought about a collapse of international trade for around two years. It says to be the most serious since the Great Depression of the 1930s. The crisis started in the United States due to a bursting housing bubble and the growth of subprime mortgage defaults (Helleiner, 2011). This period differs in the length of the drop. In the previous case, the change was sudden and took only couple of days while the fall in this case took almost two years. Therefore, the variance rises for all distributions, it is not considered as an extraordinary situation.

The second period takes almost ten years from Januray 3, 2000 to December 31, 2010 and it covers 2767 days. Detailed returns can be found in Figure 3.4. The period starts with higher volatility already for all indices followed by a very stable period until the financial crisis, where the volatility significantly increases for almost two years from 2007 to 2008 with an evident peak in the middle.

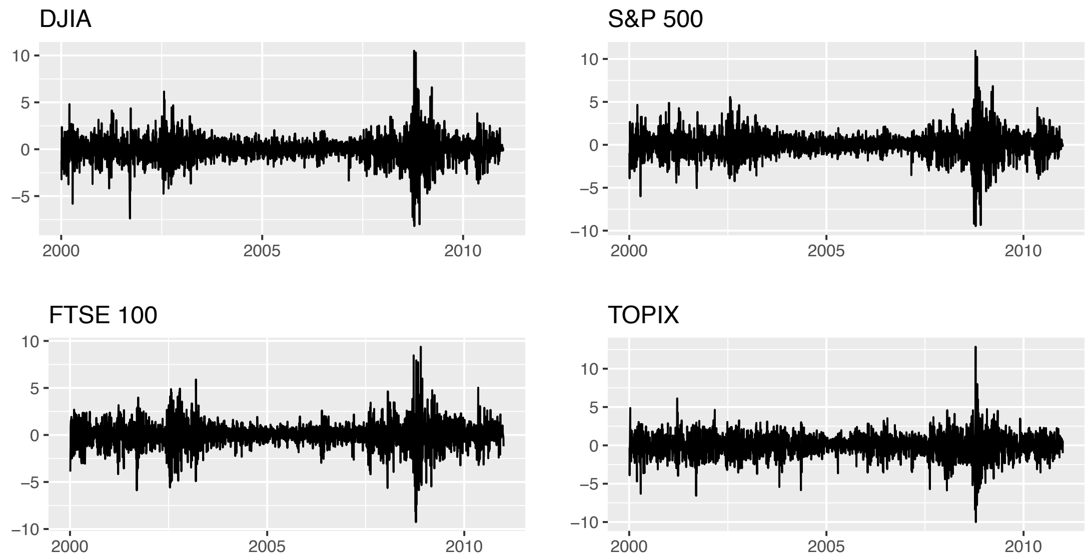

Figure 3.4: Stock market indices returns in 2000 – 2010

The last present years covers a period from January 4, 2000 to March 15, 2019 (i.e. the date when the data were downloaded). There are several cases, where a sudden drop can be observed. Despite of much slighter change, some of the sensitive distributions again treat it as a significant change in variance.

The last period contains 2315 days and the returns can be found in Figure 3.5. Comparing this figure with the previous two it could seem to lead to much larger volatility. However, it is important to notice that the scale is half. Therefore, the fluctuations look larger.

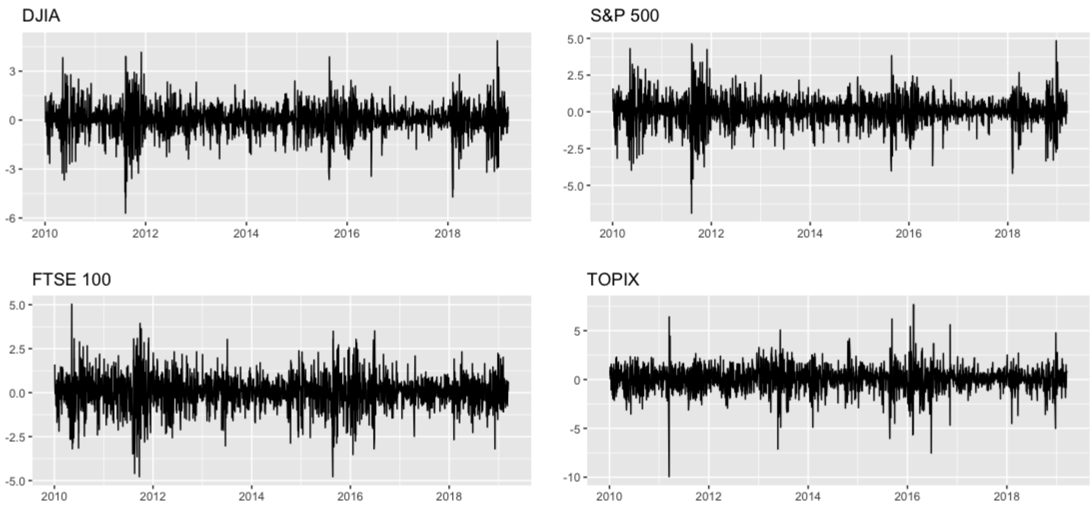

Figure 3.5: Stock market indices returns in 2010 – 2019

3.3 Time Series Analysis

Each period and index is analyzed by looking at the basic statistics, which are minimum, maximum, mean, median, first quantile and third quantile to get an idea about the data. There is the same pattern in all cases. The mean value is almost zero (slightly higher) which fulfill the assumption about the zero mean values of the series. Looking at the quantiles the first one is always lower than the third one in absolute term, which follow the next assumption about skewed distributions. The data are obviously significantly left-skewed. And finally, the maximum value is lower than the minimum value in absolute term, which is caused by the behavior of the returns. The positive ones do not behave the same as the negative ones.

1985 – 1994						
Index	Min	1st qu.	Median	Mean	3rd qu.	Max
DJIA	-25.632	-0.409	0.059	0.046	0.515	9.666
S&P 500	-22.900	-0.369	0.056	0.040	0.491	8.709
FTSE 100	-13.029	-0.505	0.062	0.036	0.634	7.597
TOPIX	-15.810	-0.503	0.028	0.022	0.558	9.116
2000 – 2010						
Index	Min	1st qu.	Median	Mean	3rd qu.	Max
DJIA	-8.201	-0.579	0.042	0.000	0.589	10.508
S&P 500	-9.470	-0.629	0.051	0.006	0.621	10.957
FTSE 100	-9.266	-0.625	0.036	0.006	0.659	9.384
TOPIX	-10.007	-0.749	0.016	0.024	0.822	12.865
2010 – 2019						
Index	Min	1st qu.	Median	Mean	3rd qu.	Max
DJIA	-5.706	-0.329	0.056	0.039	0.484	4.864
S&P 500	-6.896	-0.336	0.056	0.040	0.502	4.840
FTSE 100	-4.779	-0.468	0.041	0.013	0.530	5.032
TOPIX	-9.952	-0.611	0.061	0.025	0.711	7.715

Table 3.1: Fundamental statistics of the returns' series for each period

Normality test

Earlier in Section 1.7 basic characteristics of the time series are described along with corresponding tests. The first one was normality tested using the Shapiro-Wilk normality test which is often violated for the price returns. Its distribution is usually fat tailed and skewed. The results of this test are in Table 3.2 for each index and all three time periods. It can be seen that tests for all time series reject the normality very strongly, the values of Shapiro-Wilk statistic are close to one and p-values are almost zero. The lowest is $W = 0.764$ for DJIA in the first period but it is still very strongly rejected. These results are not surprising.

	1985 − 1994		**2000 − 2010**		**2010+**	
Index	W	p-value	W	p-value	W	p-value
DJIA	0.764	0.000	0.920	0.000	0.944	0.000
S&P 500	0.784	0.000	0.917	0.000	0.937	0.000
FTSE 100	0.893	0.000	0.931	0.000	0.969	0.000
TOPIX	0.879	0.000	0.952	0.000	0.943	0.000

Table 3.2: Shapiro-Wilk normality test

Stationary test

The second presented test is Augmented Dickey-Fuller test which tests the stationarity. This property is important for volatility modeling, however it should follow from the characteristics of the return time series that this property is fulfilled. Running the tests for each index in each time period confirmed that stationarity is fulfilled for all cases. The p-values are lower than 0.01 which leads to rejecting null hypothesis about unit roots, i.e. about non-stationary process. The lag is determined by the test itself and it was set to $p = 13$.

	1985 − 1994		**2000 − 2010**		**2010+**	
Index	DF	p-value	DF	p-value	DF	p-value
DJIA	-13.387	<0.01	-14.288	<0.01	-14.028	<0.01
S&P 500	-13.572	<0.01	-14.122	<0.01	-14.151	<0.01
FTSE 100	-11.789	<0.01	-14.157	<0.01	-14.016	<0.01
TOPIX	-12.853	<0.01	-14.325	<0.01	-13.235	<0.01

Table 3.3: Augmented Dickey-Fuller test

4. Empirical Study

Firstly in this section, it is explained why only some distributions are considered for the modeling along with the setting of the time-varying parameters and choosing the scaling type. The volatility estimation for different distributions follows along with the associated estimates of VaR and ES. The GAS models are compared using two information criteria AIC and BIC. The equality of the estimated coefficients is tested over the three time periods.

The last part is concerned with the backtesting. Three different refit steps in rolling windows are utilized. VaR and ES are then estimated and the GAS models are evaluated using the dynamic quantile test, the quantile loss function and the FZ loss function resulting in selecting the most suitable GAS model for the return series.

4.1 Choosing the Conditional Probability Distributions

GAS models provide a wide range of possible conditional distributions. However, since the price returns are often fat tailed and skewed, the most common distributions are Student's t and skew-Student's t as it was described in Section 1.3. This property is often demonstrated by comparing the results with the Gaussian distribution which is symmetric and very sensitive to extreme values and changes in variance.

4.2 Choosing the Time-varying Parameters

All evaluated distributions have the same parameters location and scale. From the characteristics of the returns it does not make sense to allow location to be time-varying. In Section 3.3, there was shown that the series has zero mean. Therefore, location will not be consider as a time-varying parameter.

Another parameter is scale (or variance), which is directly connected to the volatility. Volatility is estimated by the sample standard deviation mentioned in Section 1.1. Therefore, the square root of the scale is used to estimate it and visualized it. Setting scale to vary over time follows the properties of price returns.

The third parameter is skewness, which measures the asymmetry of the distribution. Gaussian and Student's t distributions are not skewed since they are symmetric distributions. Skew-Student's t is a representative of skewed distributions. It was tested whether the parameter of the skewness is significant or there is no point including it in the further models. GAS estimations of skew-Student's t for two different indices in two different periods were obtained, specifically returns of S&P 500 in the period of the financial crisis and first 500 observations of DJIA index. In both cases the estimated parameters related to the skewness resulted not to be statistically significant. Therefore, the skewness parameter is not considered to be time-varying from now on in the apporaching model settings. The values of estimated parameters along with the standard errors, t statistics and p-values can be found in Appendix in Tables A.1 and A.2.

The last parameter is shape and it describes the shape of the distribution. As it was mentioned earlier, the return series are often leptokurtic, which means that the kurtosis is higher than for the Gaussian distribution (kurtosis of the Gaussian is three) and also the distribution has fatter tails. An example of a leptokurtic distribution is Student's t. However, it was tested whether this parameter changes over time or not considering both Student's t and skew-Student's t distribution.

GAS estimations for different time periods and indices were obtained, specifically it was DJIA from 1985 till 1994 and S&P 500 for the period from 2010 till 2019 for Student's t and DJIA and S&P 500 both from 2010 till 2019 for skew-Student's t. In all the cases the estimated coefficients corresponding to the parameter of kurtosis resulted not to be statistically significant. It can be assumed that this parameter does not change over time for indices assumed in this thesis and the parameter is not considered to be time-varying from now on in any of the following model. The values of the estimated parameters along with the standard errors, t statistics and p-values can be found in Appendix in Tables A.3, A.4, A.5 and A.6.

4.3 Scaling Functions

There are three scaling functions, which scale the conditional score ∇_t. If the scaling matrix S_t represents the identity matrix, the scaling function is called 'identity' meaning that there is no scaling. If the scaling matrix is equal to the inverse information matrix, the scaling function is called 'inverse' and the last type is 'squared inversion,' where the scaling matrix is the square root of the information matrix.

However, the scaling does not have to produce significant differences. Two scaling functions were tested against each other, identity and inverse, over two periods 1916 – 1919 and 2012 – 2013. Only the scale parameter was set to be time-varying. The result of the GAS estimations can be seen in Figure 4.1. The red line represents the square root of estimated second moment (i.e. standard deviation) considering the identity scaling and the green line considering the inverse scaling. Both are plotted against the original return values of the price index DJIA, which is black colored. It seems that there is basically no difference. The lines almost blended into one. Therefore, it can be concluded that in these cases the scaling functions of the price indices return does not influence the estimated volatility. Henceforward, only the identity scaling function is taken into account in the following modeling.

Two models did not manage to converge. Both with the skew-Student's t distribution, specifically for the FTSE 100 in the period 1985 – 1994 and for TOPIX in the period of the financial crisis. Therefore, they are not considered from now on due to incorrect results.

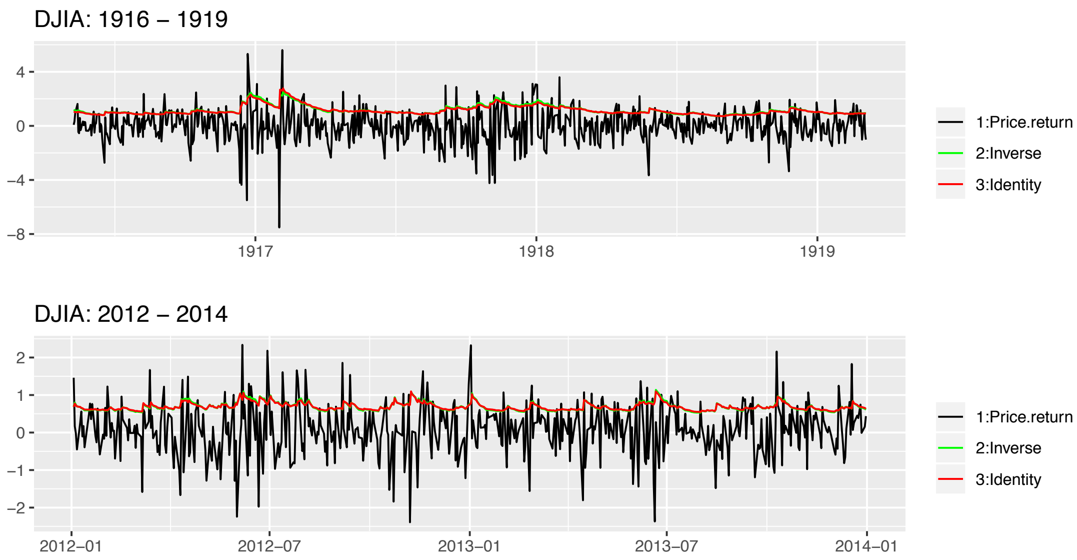

Figure 4.1: Comparing two different types of scaling (identity and inverse)

4.4 GAS Models

Reducing the number of considered distributions, time-varying parameters and scaling functions helps to simplify and speed up the analysis. To remind the final setting for GAS models, Section 4.2 set the time-varying parameter to be only the scale, Section 4.3 explained why the scaling function is set to identity and Section 4.1 chose three comparing distributions: Student's t, skew-Student's t and Gaussian.

In this way, GAS models were estimated for each price index over each time period and the values of AIC and BIC were evaluated. Three models in each period for each index were compared. The values of AIC are the lowest for the model with the conditional Student's t distribution. Only in two cases the AIC is lower for skew-Student's t but the difference is negligible since the difference Δ_h corresponding to Equation (1.47) is less than 10. The second information criterion BIC is the lowest for models with the Student's t distribution in all cases. The summary of the AIC, Δ_h and BIC values for the period of 1985 – 1994 are in Tables 4.1 and 4.2. The values for the remaining periods are in Appendix in Tables A.7 and A.8.

	DJIA			**S&P 500**		
Distribution	AIC	Δ	BIC	AIC	Δ	BIC
Student's t	6207	0	6236	5997	0	6026
skew-Student's t	6209	1.90	6244	5999	1.87	6034
Gaussian	6746	539.45	6769	6415	418.06	6438

Table 4.1: The values of AIC, Δ_h and BIC in the period 1985 – 1994 for DJIA and S&P 500

	FTSE 100			TOPIX		
Distribution	AIC	Δ	BIC	AIC	Δ	BIC
Student's t	6379	0	6408	6635	1.06	6664
skew-Student's t	6512	132.62	6547	6634	0	6669
Gaussian	6542	162.98	6565	7079	444.92	7102

Table 4.2: The values of AIC, Δ_h and BIC in the period 1985 – 1994 for FTSE 100 and TOPIX

Square roots of scales estimated over the whole period can lead to better comprehension in the differences between both considered distributions. They are plotted against each other in Figures 4.2, where two plots can be found. The first one represents the index S&P 500 over the period 1985 – 1994 and the second one the index TOPIX over the same period. The green line corresponds to the Student's t distribution and the red line to the Gaussian. This period is characterized by few sudden big drops or growths where the model with the Gaussian distribution significantly increase the variance while the leptokurtic Student's t assesses it as the rare extreme value from the tail and the variance does not increase so dramatically. This figure leads to a belief in better GAS model with conditional Student's t distribution since the variance is not influenced by drops so heavily.

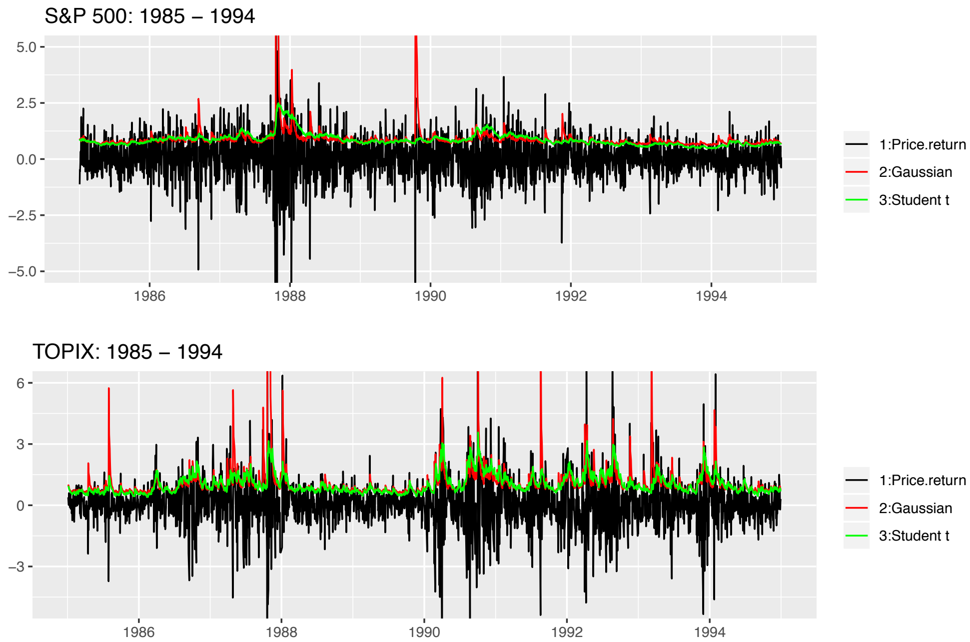

Figure 4.2: Comparing the estimates of volatility for models with Student's t and Gaussian distribution for two indices in period 1985 – 1994

Value at Risk and Expected Shortfall

A common interest is in the values of VaR and ES. A very simple indicator is the average value of VaR or ES over the time horizon. Arithmetic mean is sensitive to extreme values and therefore, significantly different mean values could point out to one or more extreme values. These were counted in Table 4.3 for GAS models for the time period from 1985 to 1994 and the remaining values for the second and third period can be found in Appendix in Tables A.9 and A.10.

The lowest values out of all the cases do not belong to the GAS model with the Student's t distribution, but very often to the Gaussian due to the estimated spikes. However, the difference is not dramatical (except of the TOPIX in 1985 – 1994, where the Gaussian model boomed and increase the variance to an enormous value during Black Monday which caused extreme difference between these two distributions).

	DJIA		**S&P 500**		**FTSE 100**		**TOPIX**	
Distribution	$\overline{VaR}$	$\overline{ES}$	$\overline{VaR}$	$\overline{ES}$	$\overline{VaR}$	$\overline{ES}$	$\overline{VaR}$	$\overline{ES}$
Student's t	-2.31	-3.12	-2.2	-2.93	-2.11	-2.56	-2.66	-3.5
skew-Student's t	-2.33	-3.14	-2.21	-2.95	-2.34	-2.95	-2.57	-3.38
Gaussian	-3.21	-3.69	-2.29	-2.63	-2.07	-2.37	-8.22	-9.43

Table 4.3: Average VaR and ES for GAS models in the period 1985 – 1994

The reason can be intuitive by looking at Figure 4.3, which compares GAS models with two conditional distributions, specifically Student's t and Gaussian for two indices S&P 500 and TOPIX in the period 1985 – 1994. The red line stays to represent the estimated volatility, the blue line corresponds to the estimated values of VaR and the green line to ES. Estimated VaR for both indices using the GAS model with the Student's t distribution copies the fluctuations of the returns much more strictly than the GAS model with the Gaussian distribution. Therefore, it is evident that the average value of VaR can be lower for the second model.

Estimates of the ES show the differences between the distributions even better. Sudden shocks in the returns along with the jumps in the variance are accompanied by the extreme drops in the estimated values of ES. That is exactly the reason why the VaR is criticized. It does not take into account the tails of the distribution, which can rapidly influence the estimates. The ES seems to be low for most of the time but when there is an unexpected drop or boom, it jumps. Therefore, the GAS model with the Student's t distribution estimates the volatility better despite having lower average value of ES since it copies the behavior of returns without many extreme booms.

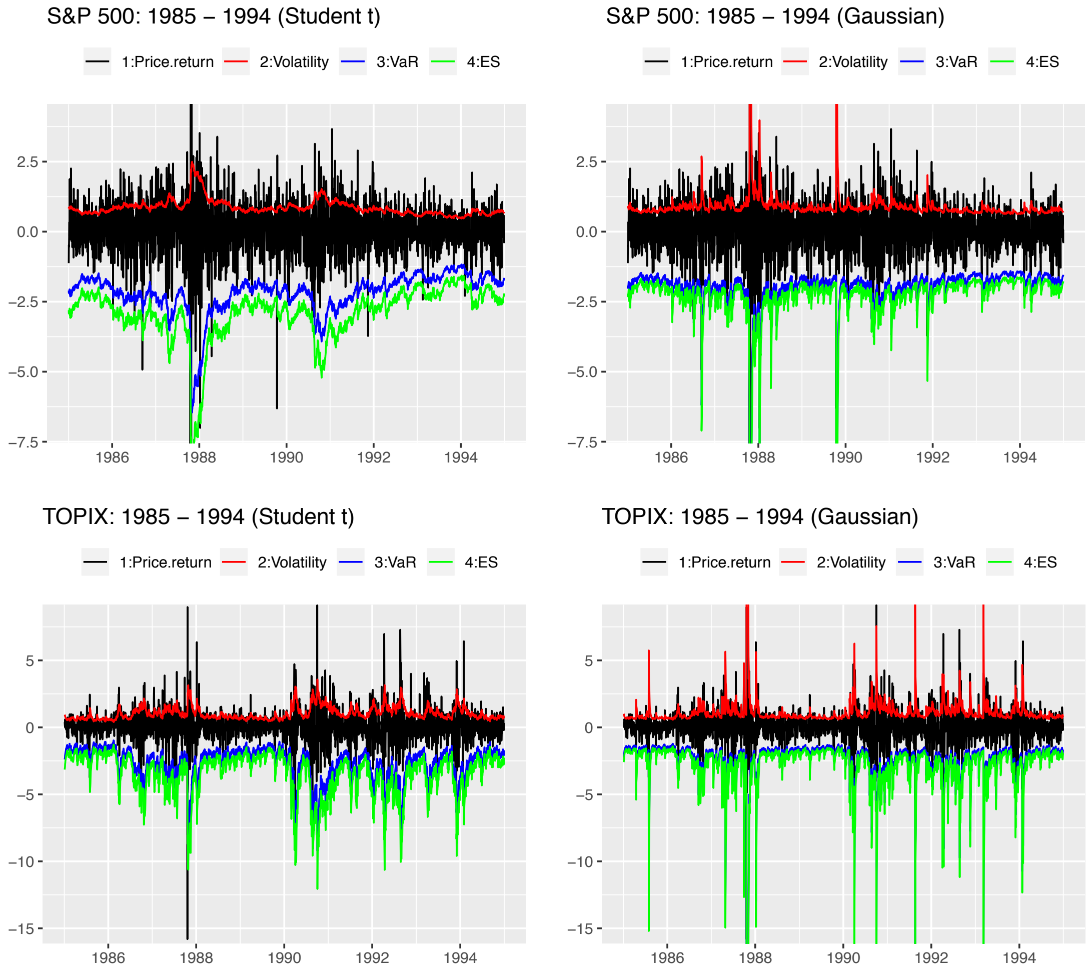

Figure 4.3: Comparing VaR and ES for GAS models with Student's t and Gaussian distribution for S&P 500 and TOPIX in the period 1985 – 1994

GAS coefficients comparison over time

Until now the models were compared by graphical analysis. The scale is set to vary over time but the static parameters do not. Once they are estimated, they stay constant over the corresponding time period. It could be interesting to test whether they also change over time. Since the GAS model is estimated for the same index with the same conditional distribution on three different time periods, it is possible. The Z-test given by Equation (1.49) is used at 5% confidence level.

Most of the coefficients, which are equal over time are elements of the vector $\boldsymbol{\kappa}$. An interesting discovery is an inequality of the elements of the matrix $\boldsymbol{A}$, which precede the scaled score and express the relationship to it. The equality holds only for indices DJIA and S&P 500. On the other hand almost all of the coefficients are equal for the indices FTSE 100 and TOPIX. It is also connected to the behavior of the returns. Figure 3.2 in Section 3.1

could suggest that the variance is more likely similar for the indices FTSE 100 and TOP-IX, while DJIA and S&P 500 evince more imbalance and unexpected changes. However, the first two named indices show continual larger volatility than the second two. Therefore, the equality of the coefficients is compensated by larger volatility. This characteristic also led to the differences between the GAS models with Gaussian and Student's t distribution in Figure 4.2.

Moreover, the values of the matrix $\boldsymbol{B}$, which expresses the relation to the past value of the vector $\boldsymbol{\theta}_t$ are statistically equal in almost all cases and the p-value is often very high. It can be concluded that the autoregressive term does not change in different time periods. Z-scores and p-values for DJIA for all three distributions are stated in Table 4.4. The remaining values are in Appendix in Tables A.11 and A.12.

Student's t						
	1985–1994/2000–2010		**1985-1994/2010+**		**2000–2010/2010+**	
Coeff	Z-score	p-value	Z-score	p-value	Z-score	p-value
κ_1	0.451	0.674	-0.819	0.206	-1.249	0.106
κ_2	1.007	0.843	-2.308	0.011	-2.839	0.002
κ_3	-2.067	0.019	-0.684	0.247	1.677	0.953
A_2	-1.660	0.049	-4.300	0.000	-3.132	0.001
B_2	0.165	0.566	2.596	0.995	2.555	0.995
skew-Student's t						
	1985–1994/2000–2010		**1985-1994/2010+**		**2000–2010/2010+**	
Coeff	Z-score	p-value	Z-score	p-value	Z-score	p-value
κ_1	0.419	0.662	-0.817	0.207	-1.249	0.106
κ_2	0.100	0.540	-1.880	0.030	-1.816	0.035
κ_3	-0.081	0.468	0.102	0.541	0.188	0.575
κ_4	-2.075	0.019	-0.682	0.248	1.677	0.953
A_2	-1.664	0.048	-4.277	0.000	-3.114	0.001
B_2	0.175	0.569	2.594	0.995	2.546	0.995
Gaussian						
	1985–1994/2000–2010		**1985-1994/2010+**		**2000–2010/2010+**	
Coeff	Z-score	p-value	Z-score	p-value	Z-score	p-value
κ_1	0.905	0.817	-0.096	0.462	-1.070	0.142
κ_2	1.848	0.968	-1.247	0.106	-3.134	0.001
A_2	0.555	0.711	-1.974	0.024	-2.624	0.004
B_2	-3.857	0.000	-1.040	0.149	3.369	0.999

Table 4.4: Z-test of coefficient equality for GAS models for index DJIA

4.5 Comparison of GAS and GARCH Models

GARCH models are so far one of the most applied out of all volatility models. Both GARCH and GAS models described in Section 1.2 and 1.3 belong to the class of observation-driven models. However, they differ in the updating mechanism. GAS models have the advantage of using the complex density structure rather than only means and higher moments. How much the estimated volatility differs may be dependent on the properties of the corresponding time series as well.

The first example covers TOPIX and S&P 500 in the period from 1985 to 1994 and the conditional distribution is assumed to be Student's t. In Figure 4.4 the blue line represents the estimated volatility from GARCH model and the red line refers to the estimated volatility from GAS model. The original returns are black colored.

TOPIX is characterized by more fluctuations, which both models have a problem to smooth out comparing to the S&P 500. It can be very well seen that GARCH model is even more sensitive than GAS model and the estimated volatility is larger, it exceeds the GAS in a significant part of the cases.

On the other hand, S&P 500 evinces more constant variance and both models managed to smooth out the series very well and similarly. However, the difference between GAS and GARCH models is significant for the extreme values and especially for Black Monday, where GARCH unexpectedly jumped away. So far, it seems that GARCH is sensitive to extreme values (meaning price shocks) since the dependency on the first lagged value is very strong.

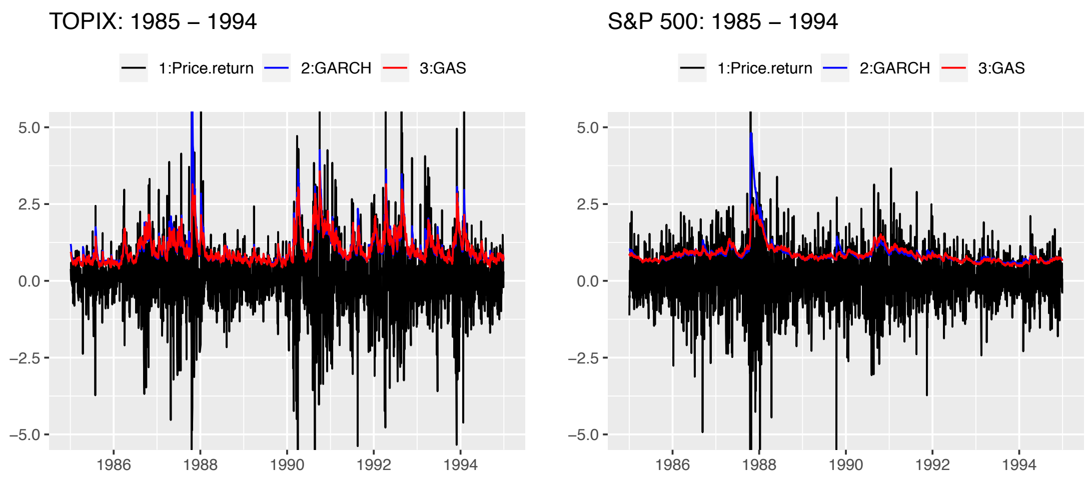

Figure 4.4: Comparing estimated volatility from GAS and GARCH models for TOPIX and S&P 500 considering conditional Student's t distribution in the period 1985 – 1994

The second example in Figure 4.5 represents estimated volatility for DJIA also assuming the Student's t distribution in the period 2010 – 2019. The lines denote the same as the previous figure. However, there is suddenly almost no difference between both models, the lines copy each other and the GARCH model does not seem to contend with higher variance. DJIA has different behavior in this period, it is more stable and the fluctuations are more balanced, no extreme shocks, which could be the reason why both model almost do not differ.

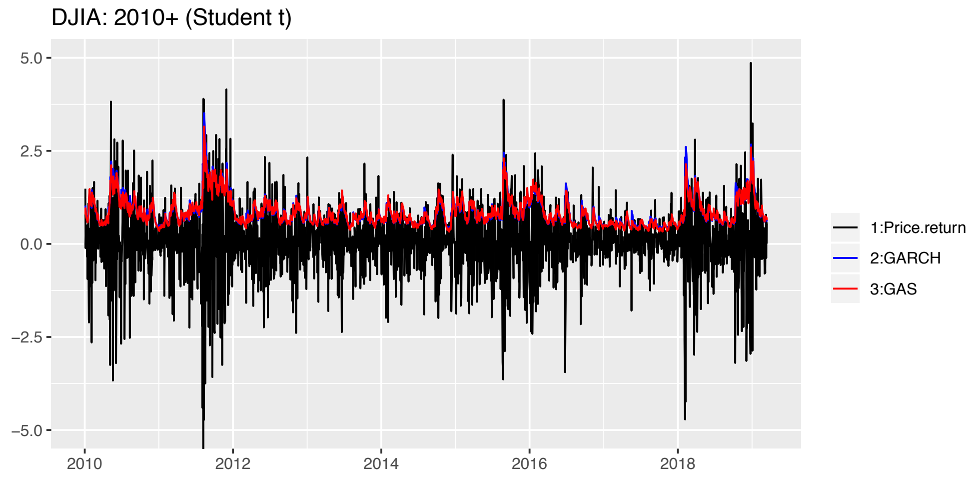

Figure 4.5: Comparing estimated volatility from GAS and GARCH models for DJIA assuming conditional Student's t distribution in the period 2010 − 2019

The evaluated 1 % VaR evinces a similar impacts for TOPIX and S&P 500. Both cases are plotted in Figure 4.6, where the VaR from GAS model is blue lined and the VaR from GARCH is green lined. In general, the VaR for TOPIX fluctuates much more than for S&P 500. However, extreme values, especially Black Monday, show a huge jump for both indices applying GARCH model. The same stands for the DJIA in Figure 4.7 (note that colors are reversed for better visibility). Since the GARCH model did not create such deviations or price shocks, there are no sudden jumps. Despite of no significant differences between both estimates, the GAS model estimates the values of VaR constantly higher than the GARCH model. This did not apply to the previous two examples. It shows that the behavior of returns influence the resulted estimated VaR and ES and it can not be generally said that GARCH estimated the values lower than the GAS model.

Calculated averages of the 1 % VaR are significantly different and higher for the GAS model for all the indices in the period from 1985 − 1994 and for 2010 − 2019, whilst they are lower or the same for the period including the financial crisis. It seems that each period has a different behavior but the whole market shaped up same and therefore, the GARCH and GAS managed to behave similarly for all the indices. To summarize it, for the first period from 1985 to 1994, GARCH model evinced a big jumps for the occasional extreme values. The second period, which covers the financial crisis seems to work with both models the same and the estimated VaR values are almost identical (Figure 4.8 was added to demonstrate that the values copy each other over the whole period). And for the last period 2010 − 2019, it holds that the estimated VaR from GAS model exceeded GARCH almost the whole time but the shape is nearly the same. Average VaR values for both models are in Table 4.5.

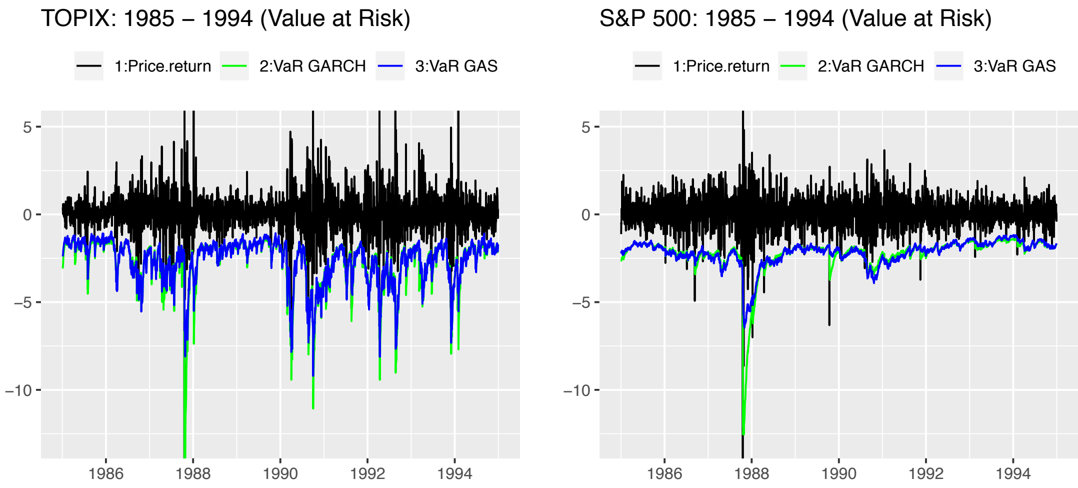

Figure 4.6: Comparing 1 % VaR for GAS and GARCH models for TOPIX and S&P 500 assuming conditional Student's t distribution in the period 1985 – 1994

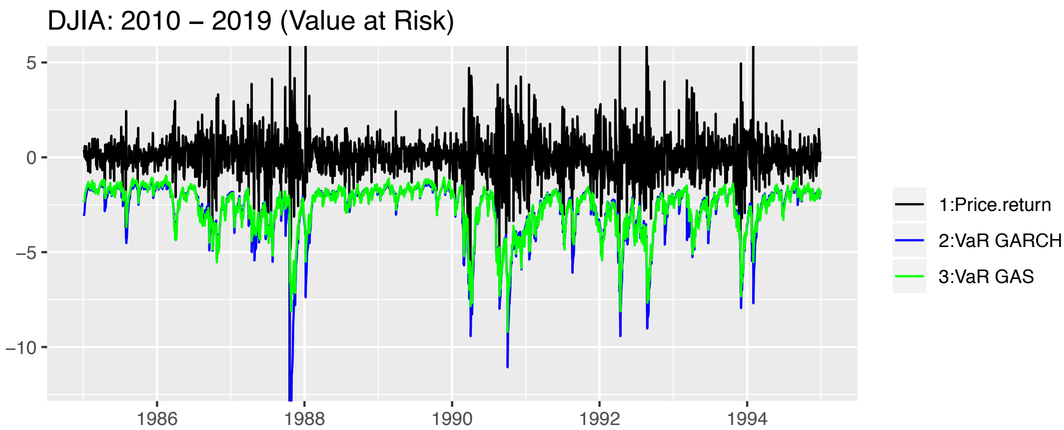

Figure 4.7: Comparing 1 % VaR for GAS and GARCH models for DJIA assuming conditional Student's t distribution in the period 2010 – 2019

	1985 – 1994		**2000 – 2010**		**2010 – 2019**	
Index/model	GAS $\overline{VaR}$	GARCH $\overline{VaR}$	GAS $\overline{VaR}$	GARCH $\overline{VaR}$	GAS $\overline{VaR}$	GARCH $\overline{VaR}$
DJIA	-2.31	-2.15	-2.82	-2.83	-2.11	-1.85
S&P 500	-2.20	-2.04	-2.98	-2.98	-2.21	-1.94
FTSE 100	-2.11	-2.05	-2.82	-2.82	-2.23	-2.04
TOPIX	-2.66	-2.48	-3.25	-3.27	-2.88	-2.68

Table 4.5: Average VaR for Student's t distribution for GARCH and GAS model

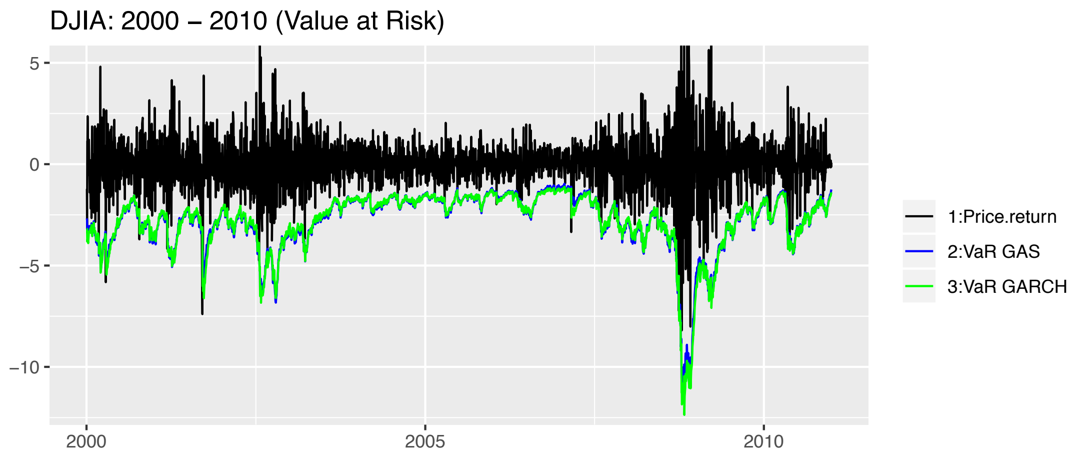

Figure 4.8: Comparing 1 % VaR for GAS and GARCH models for DJIA assuming conditional Student's t distribution in the period 2000 − 2010

4.6 Backtesting VaR and ES

Making predictions requires voluminous simulations. However, one-step ahead predictions can be used in backtesting, where the simulations are not necessary. The one-step ahead forecasts are applied in every step of the rolling windows but it depends on the length of the refit step how often the parameters are re-estimated. The output of the rolling window is a vector of predicted values, but they are made as one-step ahead forecast and therefore, there is no need for simulations. Eventually, the values of VaR and ES are evaluated and they can be compared with the original values from the GAS model. However, the crucial comparison is based on the dynamic quantile (DQ) test and the loss functions. These can be counted for each model and assess the quality of the model and the estimated VaR and ES values.

The sensitivity of the refit step for rolling windows

There are two parameters that the rolling windows required to be set. Firstly, the forecast length (or in other words the length of the out-of-sample) and the length of the refit step. Since the length of the time series of the returns is around 10 years, which is between two and three thousand-observations, the length of the out-of-sample was set to 1000, which represents approximately one third of the whole sample.

The length of the refit step can be from 1 to T where T is the length of the series. Obviously, the refit step equals to T does not make much sense for the dynamic models. But there is no rule how to set it. The most common seems to be the length 1 and 5 for daily data (Ardia et al., 2018). An interesting choice is the length 4 for quarterly data used in (Ardia et al., 2019), where the motivation is to refit parameters once per year.

Therefore, the sensitivity of the refit step was tested for the returns data and the values of the estimated VaR and ES using GAS models were compared. The considered lengths were 1, 5 and 30. The length of 1 was chosen since it allows to adapt to the changes every day. The length of 5 was chosen to cover the length of the work week (there is no price of the market indices during the weekends) and the length of 30 represents the average length of a month.

It emerged that the lengths of 1 and 5 evince almost identical results. The estimated values of VaR and ES basically copied each other so it seems that such a small difference does not cause a change for daily data. The difference between the lengths of 1 and 30 led to small changes. They were almost negligible for the Student's t but they changed a little during the financial crisis in the end of 2008. This situation is in Figure 4.9 for the FTSE 100 in the period from 2000 to 2010, where is also the comparison with the original estimated values of VaR (i.e. without the rolling windows). There are three plots that compare the values of VaR. In the first one, the black line represents the original values of VaR and the red line the values of VaR obtained after running the rolling windows with the refit step of 1. The second plot compares original VaR and VaR with the refit step of 30 and the last plot compares two values of VaR from the rolling windows. The black line represents the refit step equals to 1 and the red line to the length of 30.

The ES shows the same pattern as VaR. There are no significant dissimilarities, the lines only slightly differ in the end of 2008. By comparing it with the originally estimated values, the refitted ES tends to be slightly lower than the original ES in the first half of the out-of-sample period. This is also valid for the VaR, but the change is not that obvious. The situation with ES is in Figure 4.10. The notation is same as in the previous figure.

However, it seems that this does not hold for the conditional Gaussian distribution. In Figure 4.11 there is a comparison of DJIA in the period from 2010 to 2019. The biggest difference is between the original values of VaR and the re-estimated ones. The colors were reverted due to better visibility, therefore the original values now represent the red line and the refitted values the black line. For the third plot in this figure the black line corresponds to the refit step of the length 30 and the red line equals to the refit step of 1 (same is valid for ES in Figure 4.12). The rolling windows approach evinces a high sensitivity to the sudden changes in the original values of the price index and the estimated VaR drops rapidly. This holds for shocks more than for slight changes but it is valid for both. However, the longer refit step is even more sensitive. When the changes are gradual and not extreme, all three values of VaR are almost identical (but the original values of VaR are higher most of the time) and the refitted values of VaR jump significantly during the price shocks.

Looking at the values of ES in Figure 4.12, the differences are also significant between the refitted values and the original, but slightly less than for VaR. Refitted ES still tend to be much lower in the spikes but otherwise, they copy each other very well.

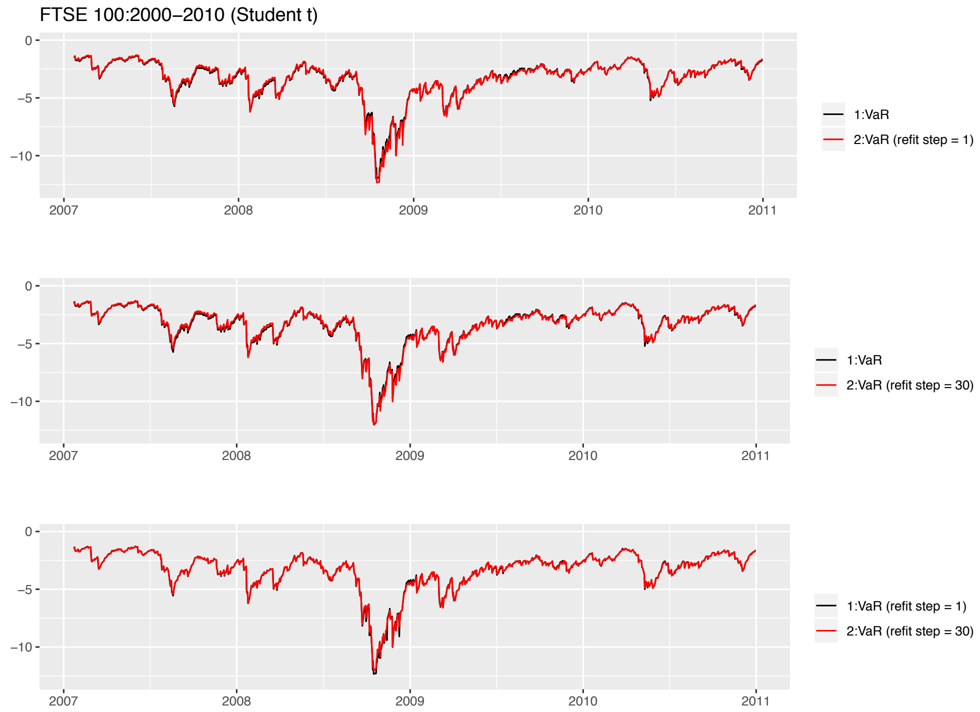

Figure 4.9: 1 % VaR for FTSE 100 with the Student's t distribution in 2000 – 2010

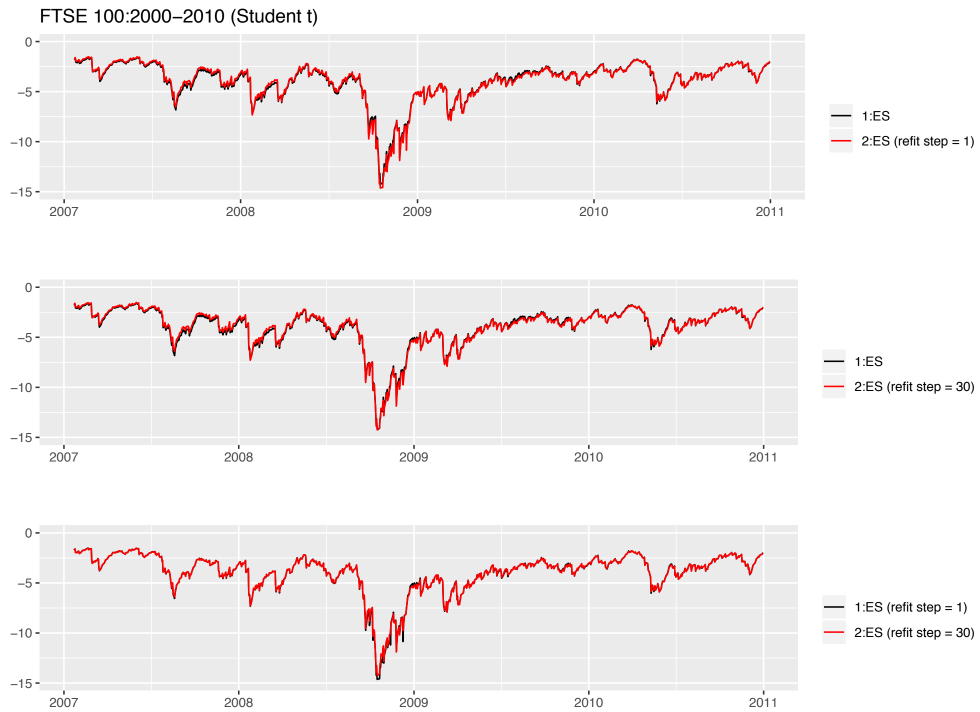

Figure 4.10: 1 % ES for FTSE 100 with the Student's t distribution in 2000 – 2010

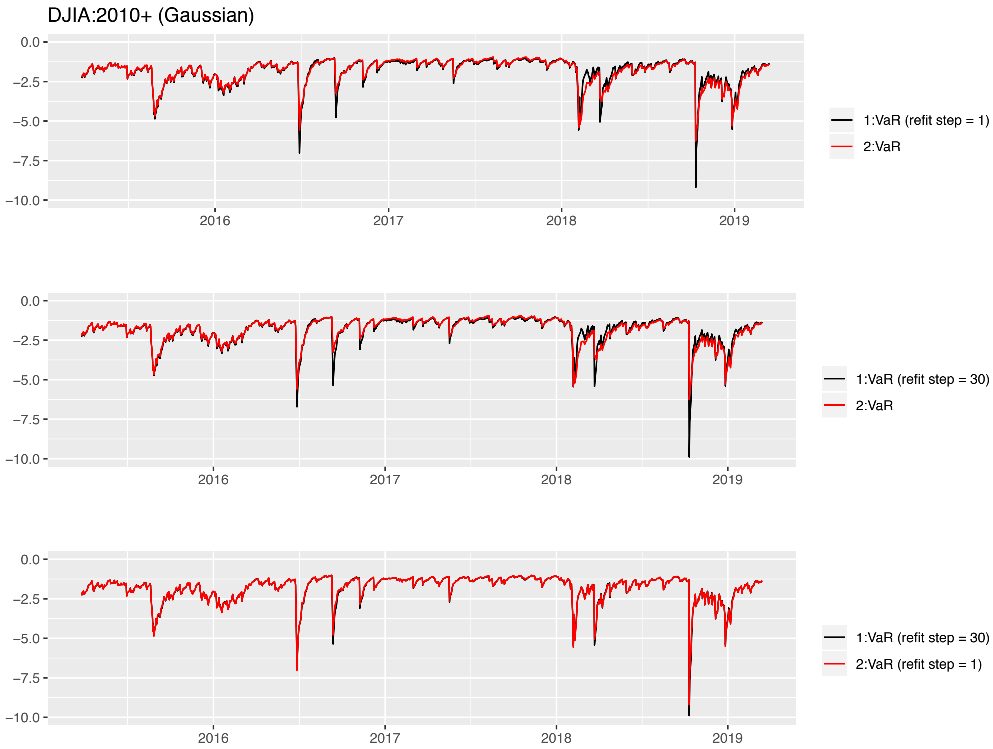

Figure 4.11: 1 % VaR for DJIA with the Gaussian distribution in 2000 – 2010

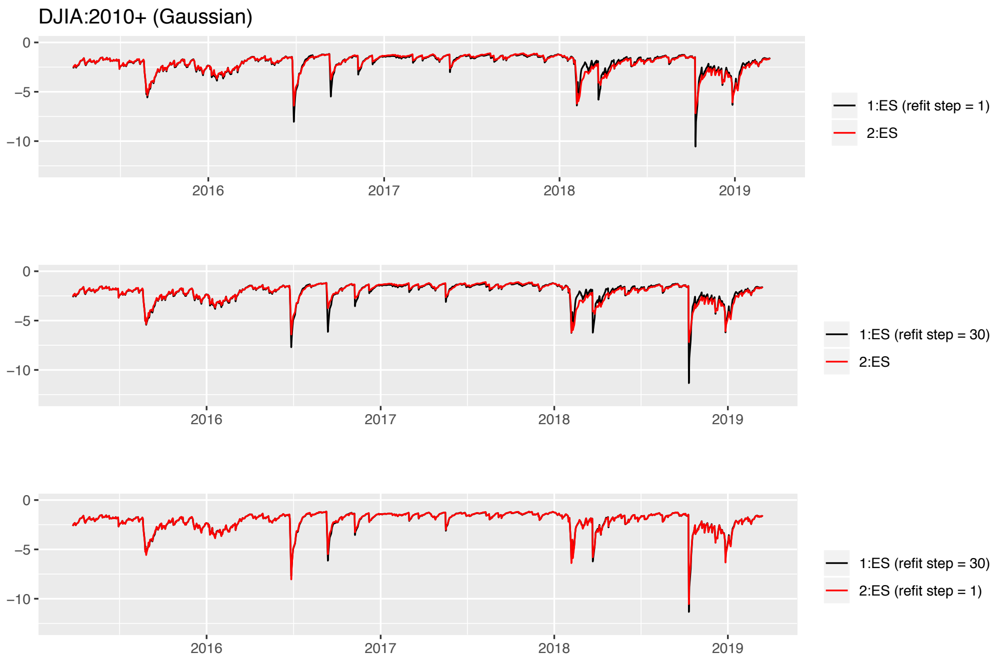

Figure 4.12: 1 % ES for DJIA with the Gaussian distribution in 2000 – 2010

It can be concluded that the length of the refit step does not influence the estimated VaR and ES for the Student's t distribution and moreover, it does not vary from the originally estimated values. However, there is a significant difference for the Gaussian distribution. The refitted values exceeds the original values dramatically and this holds for both risk measures. There is even a difference between the refitted values when the longer step tend to drop more than the shorter step.

Dynamic Quantile test

Dynamic Quantile (DQ) test tests the values of VaR obtained from the rolling window and uses them to count the Hit processes, which are formulated in Section 1.6 in Equation (1.51). Under the null hypothesis, the model has correct unconditional and conditional coverage, in other words the model is correctly specified. This test was evaluated for 1 % and 5 % coverage rate for each index in each considered period for all three conditional distributions for each setting of the refit step in GAS models.

The percentage of rejection of the coverage test changes over the periods and the indices, too. At 5 % confidence level, the GAS model is correctly specified for all indices in each period with the refit step of 1 looking at the results of 5 % DQ test. However, the 1 % DQ test does not support it that often. For example, the hypothesis cannot be rejected in the first period for S&P 500 for any refit step assuming Student's t or skew-Student's t distribution but it is rejected for the Gaussian distribution.

It can not be generally said that the DQ test rejects the correct GAS model specification for models with the Gaussian distribution rather than for models with the Student's t distribution or skew-Student's t distribution in cases covered in this thesis. The null hypothesis is more often rejected by all the models or none of them for a given refit step in given time period.

The percentage of rejection of the coverage test is lower for the Gaussian distribution in most of the cases. An example can be DJIA in the period from 2010 to 2019. All the p-values are significantly higher than 0.05 but the difference is in 0.1 and 0.2 points for the Gaussian distribution for the refit step of 5 and 30. The p-values for the refit step of 1 are almost equal for all three GAS models for 5 % DQ test. The same states for S&P 500 in this period.

On the other side, it was mentioned earlier in this Section that FTSE 100 does not evince significant changes in the variance in this period, it is quite large the whole time. This seems to be depicted in the results, where the percentage of rejection of the coverage test is higher for the GAS model with the Gaussian distribution and not for the Student's t or skew-Student's t. The corresponding p-values are mentioned in Table 4.6 for S&P 500 for the first period and for DJIA and FTSE 100 for the last period. The remaining values can be found in Appendix in Tables A.13 and A.14.

1985 – 1994:

S&P 500	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % DQ	1 % DQ	5 % DQ	1 % DQ	5 % DQ	1 % DQ
Student's t	0.075	0.072	0.367	0.001	0.393	0.001
skew-Student's t	0.076	0.056	0.250	0.062	0.281	0.061
Gaussian	0.081	0	0.021	0.000	0.019	0.000

2010 – 2019:

DJIA	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % DQ	1 % DQ	5 % DQ	1 % DQ	5 % DQ	1 % DQ
Student's t	0.099	0.000	0.412	0.000	0.397	0.000
skew-Student's t	0.099	0.000	0.412	0.000	0.393	0.000
Gaussian	0.100	0.000	0.244	0.000	0.275	0.000
FTSE 100	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % DQ	1 % DQ	5 % DQ	1 % DQ	5 % DQ	1 % DQ
Student's t	0.097	0.106	0.402	0.110	0.426	0.110
skew-Student's t	0.097	0.102	0.401	0.073	0.491	0.112
Gaussian	0.097	0.016	0.581	0.017	0.582	0.010

Table 4.6: P-values for 1 % and 5 % DQ test of the GAS model correct coverage for selected indices

Generally, it seems that the test is strongest for the GAS models with the refit step of 1 but it cannot be said that the test tends to reject the correct coverage for the models with longer refit step. It seems to be an individual property. However, it can be concluded that the 1 % DQ test is more strict that 5 % DQ test since it rejects the correct coverage in almost all cases.

Moreover, it could be tested how much the results change with a different in-sample and out-of-sample sizes since Dumitrescu et al. (2012) stated that the power of the test tends to increase when the size of the in-sample increases and decrease when the size of the out-of-sample increases. However, this is out of the scope of this thesis.

Quantile loss function

Quantile loss (QL) function evaluates the model performance and it penalizes the exceedances of VaR cases more heavily. A model with a lower value of the function is preferred over the selected models. Models can be assessed also by comparing a pair of models and calculating a ratio of the corresponding loss functions. This theory was formulated in Section 1.6 in Equations (1.52) and (1.53).

The averaged QL functions were calculated for 1 % and 5 % VaR and it turned out that the lowest values belong to the GAS models with the Student's t or skew-Student's t distribution, which confirmed their dominance against the GAS models with the Gaussian distribution. These two are very often almost identical, the difference is negligible and the gap between these and the Gaussian is sometimes less and sometimes more significant. However, there are three occurrences, where the lowest value corresponds to the GAS model with the Gaussian distribution. These cases are FTSE 100 in the first two periods and TOPIX in the second period. They are summarized in Table 4.7, where can be seen that the differences are very small.

It can be observed that the GAS model with the lowest values does not change with the refit step. The opposite happens only in two or three cases but for the rest of the testing field, if the model has the lowest value for the refit step of 1 then it will have the lowest for the refit steps of 5 and 30.

Furthermore, being the leading model according to the 5 % QL does not mean to be the leading model according to the 1 % QL. This happens actually very often and even the Gaussian distribution is included. For example for FTSE 100 in the second period, the lowest value of 5 % QL function has the model with the Gaussian distribution, but skew-Student's t for 1 % QL function.

1985 – 1994:

FTSE 100	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % QL	1 % QL	5 % QL	1 % QL	5 % QL	1 % QL
Student's t	0.085	0.026	0.085	0.026	0.085	0.026
Gaussian	0.086	0.026	0.087	0.026	0.086	0.025

2000 – 2010:

FTSE 100	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % QL	1 % QL	5 % QL	1 % QL	5 % QL	1 % QL
Student's t	0.168	0.050	0.167	0.049	0.167	0.050
skew-Student's t	0.167	0.049	0.167	0.049	0.168	0.049
Gaussian	0.166	0.050	0.166	0.050	0.166	0.050
TOPIX	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % QL	1 % QL	5 % QL	1 % QL	5 % QL	1 % QL
Student's t	0.175	0.049	0.175	0.049	0.175	0.049
Gaussian	0.174	0.049	0.174	0.049	0.174	0.049

Table 4.7: Results of 1 % and 5 % QL functions for selected indices

The ratios were calculated for each index in each period for each refit step and both 1 % and 5 % confidence level. Firstly, GAS models with the Student's t distribution and the Gaussian are compared and secondly, skew-Student's t and Gaussian. When the ratio exceeds the value of one, the model with the Gaussian distribution is better, while for values bellow one, the model with the Student's t or skew-Student's t distribution respectively is better. If the ratio is equal to 1, there is no difference between the models with the relation to the estimated VaR.

Calculated ratios for the period from 2010 to 2019 for all three refit steps can be found in Table 4.8. The GAS models with the Student's t and skew-Student's t seem to be better. There is only one case, where the Student's t is worse than the Gaussian, specifically FTSE 100 without the distinguishing in the length of the refit step. However, the ratio is equal to one which leads to the fact that there is no preferred or better model. For the remaining ratios, the GAS model with the Gaussian distribution is usually worse than the Student's t or the skew-Student's t. In few cases, the ratio is equal to one and in only small percentage of the cases the models with the Gaussian distribution are better (for example, FTSE 100 in the first period comparing Student's t and the Gaussian or DJIA in the first period comparing the skew-Student's t and the Gaussian). The values of the ratios in the first two periods can be found in the Appendix in Tables A.17 and A.18.

	refit step = 1		**refit step = 5**		**refit step = 30**	
DJIA	5 % QL ratio	1 % QL ratio	5 % QL ratio	1 % QL ratio	5 % QL ratio	1 % QL ratio
St/Gauss	0.99	0.98	0.99	0.99	0.98	0.98
skew-St/Gauss	0.99	0.99	0.99	0.99	0.98	0.98
S&P 500						
St/Gauss	0.99	0.99	0.99	0.99	0.98	0.98
skew-St/Gauss	0.99	0.99	0.99	0.99	0.98	0.99
FTSE 100						
St/Gauss	1.00	0.96	1.00	0.96	1.00	0.96
skew-St/Gauss	1.00	0.96	1.00	0.96	1.00	0.96
TOPIX						
St/Gauss	0.98	0.91	0.98	0.91	0.98	0.91
skew-St/Gauss	0.98	0.91	0.98	0.91	0.98	0.91

Table 4.8: Ratios of the 1 % and 5 % QL functions for the period from 2010 to 2019

FZ loss function

There is no individual loss function for ES. However, in Section 1.6 it was shown that VaR and ES are jointly elicitable and the loss function is FZL function. FZL function is computed for each point in time given the vector of VaR and ES estimates and it is averaged afterwards. The lower value is better. FZL was counted for each index in each period for all three refit steps for each considered conditional distribution at 1 % and 5 % risk level.

The results almost copy the previous loss function. Most often the lowest values of FZL have the GAS models with Student's t or skew-Student's t distribution and the difference between them is negligible. The gap is more significant between the Gaussian and Student's t or the Gaussian and the skew-Student's t. However, the difference is very small among all three models in the few cases, where the lowest value of average FZL belongs to the GAS model with the Gaussian distribution. These situations correspond to the ones, where the QL function was the lowest, specifically FTSE 100 in first two periods and TOPIX in the second period. The only difference is that it does not hold for all the refit steps of the given index. For FTSE 100 in the first period, the Gaussian model has the lowest value for 1 % risk level for the refit steps equal to 1 and 30, but for the refit step of 5, the lowest value has the model with the Student's t distribution. For the second period, the lowest value belongs to models with the Gaussian distribution only at 5 % risk level with the refit step of 30, while for the QL function it held for all three refit steps. The results for TOPIX copy the previous ones in each point.

Once again it seems that the length of the refit step does not influence which model has the lowest value of the average FZL function. There is only a few exceptions, but overall it stays the same. The difference is very often only between the Student's t and skew-Student's t distribution and it is very small. In Tables 4.9 and 4.10, there are mentioned the average values of FZL for FTSE 100 in the first two periods and for TOPIX in the second period. The remaining values can be found in Appendix in Tables A.19 and A.20.

1985 – 1994: FTSE 100	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % FZL	1 % FZL	5 % FZL	1 % FZL	5 % FZL	1 % FZL
Student's t	0.507	0.941	0.507	0.938	0.507	0.940
Gaussian	0.520	0.925	0.531	0.963	0.520	0.913
2000 – 2010: FTSE 100	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % FZL	1 % FZL	5 % FZL	1 % FZL	5 % FZL	1 % FZL
Student's t	1.175	1.615	1.171	1.605	1.173	1.614
skew-Student's t	1.170	1.599	1.169	1.595	1.172	1.597
Gaussian	1.172	1.670	1.170	1.664	1.171	1.664

Table 4.9: 1 % and 5 % average FZL functions for FSTE 100 in two different periods

2000 – 2010: **TOPIX**	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % FZL	1 % FZL	5 % FZL	1 % FZL	5 % FZL	1 % FZL
Student's t	1.216	1.571	1.216	1.572	1.215	1.570
Gaussian	1.214	1.601	1.214	1.600	1.213	1.596

Table 4.10: 1 % and 5 % average FZL functions for TOPIX in 2000 – 2010

And finally, the ratios of the average FZL function were computed for all three refit steps at both 1 % and 5 % risk level. The pairs are the same as in the previous example. Firstly, GAS models with the Student's t and the Gaussian distribution are compared and secondly, the skew-Student's t and the Gaussian. When the ratio exceeds the value of one, the GAS model with the Gaussian distribution is better considering both risk measures VaR and ES. If the ratio is equal to 1, there is no difference in the accuracy of the estimated VaR and ES.

GAS models with the Student's t and skew-Student's t distributions resulted to be better than GAS models with the Gaussian. There is only a few cases, where the model with the Gaussian is better and it is in less cases than QL ratios. An example can be DJIA in the first period, where the model with the Gaussian distribution was better than the skew-Student's t at 1% confidence level for all three refit steps. However, FZL ratios did not reach the value of one and therefore, the skewed distribution fits better. The GAS models were equally accurate for DJIA in the second period comparing the Student's t and the Gaussian at 5 % confidence level. But looking at the FZL ratios, the Student's t is better.

The last example represents the opposite situation. The suitability of the Gaussian distribution is even stronger for FTSE 100 in the first period at 1 % confidence level for the refit step of 30. In both cases the ratio exceeded one but for FZL the ratio is even slightly higher. In this particular case, the values change with the length of the refit step. If the refit step is equal to 1, the GAS model with the Gaussian outperformed the GAS model with the Student's t in both cases. If it is equal to 5, the GAS models were equal according to the QL ratio but the model with the Student's t distribution is better according to the FZL ratio.

To summarize it, GAS models with Student's t and skew-Student's t distributions outperform GAS model with the Gaussian distribution very significantly and the model with the Student's t is very similar to the skew-Student's t. The ratios of average FZL function for the last period for all three refit steps at both risk levels are mentioned in Table 4.11. The remaining values can be found in Appendix in Tables A.21 and A.21.

	refit step = 1		**refit step = 5**		**refit step = 30**	
DJIA	5 % FZL ratio	1 % FZL ratio	5 % FZL ratio	1 % FZL ratio	5 % FZL ratio	1 % FZL ratio
St/Gauss	0.97	0.93	0.97	0.93	0.96	0.92
skew-St/Gauss	0.97	0.93	0.97	0.94	0.96	0.92
S&P 500						
St/Gauss	0.99	0.94	0.99	0.95	0.98	0.94
skew-St/Gauss	0.99	0.94	0.99	0.95	0.98	0.94
FTSE 100						
St/Gauss	0.99	0.94	0.99	0.94	0.99	0.94
skew-St/Gauss	0.99	0.94	0.99	0.94	0.99	0.94
TOPIX						
St/Gauss	0.97	0.88	0.97	0.88	0.97	0.88
skew-St/Gauss	0.97	0.88	0.97	0.88	0.97	0.88

Table 4.11: 1 % and 5 % ratios of the average FZL function for the period from 2010 to 2019

4.7 Computational Complications

Absolute majority of the computations run in the programming language R. GAS models are possible to be estimated and analyzed using the recently developed package *GAS*. Since the package is still fresh, published in 2016 by Catania et al., there are some errors and imperfections. It turned out that there is a problem with the convergence of other types of scaling besides identity. Despite proving that the scaling type seems not to be important, it would be interesting to really analyze it. Unfortunately, none of the models with other than identity scaling did not converge and therefore, only identity scaling was put into consideration.

Conclusion

Significant differences among volatility estimates from GAS and GARCH models confirmed not merely graphical analysis. The distinction led to the belief that GARCH models tend to extremely increase the variance in days with price shocks, while GAS models with the same observation distribution estimate the volatility significantly lower. VaR estimates for GARCH models showed that the values significantly fluctuate and reach extreme values in the period with highly volatile returns, while values of VaR for GAS models were not influenced by the shocks as extremely as GARCH despite of observable growth of risk as well. If the models were applied to a tranquil period without shocks or jumps, both models estimated the volatility very similarly as well the values of VaR.

GAS models with conditional Student's t and skew-Student's t distributions evinced very similar results. Differences of AIC Δ_h were negligible and the models considered to be equally good in all assessed cases for all four indices. Estimated volatility and values of both risk measures led to similar estimates, which differed very slightly. On the contrary, GAS (GARCH) model with the Gaussian distribution increased the estimated volatility substantially in days of price shocks and jumps, which is not correct since AIC was significantly higher comparing to AIC of GAS models with Student's t or skew-Student's t distribution. The values of corresponding VaR and ES copied the fluctuations and significantly increase in days of price shocks or jumps.

The elements of vector of constants $\boldsymbol{\kappa}$ stayed unchanged for most of the GAS models with all conditional probability distributions comparing all pairs of chosen time periods. Matrix $\boldsymbol{A}$, which represents the step of the update was the same over time periods for the indices DJIA and S&P 500, where the price returns were more volatile, while they are mostly equal for remaining two indices, where the volatility was more stable without shocks or spikes. Matrix $\boldsymbol{B}$, which controls the persistence of the process was very strongly supported to be the same over time periods in all cases.

The length of the refit step for rolling windows did not show significant differences for GAS models in all considered cases assuming Student's t or skew-Student's t distributions. The estimated values of VaR and ES were almost indifferent. The length of the refit step influenced more the GAS model with the Gaussian distribution, where the estimated values of VaR and ES tended to be significantly lower in days with price shocks and jumps than the values of VaR and ES calculated for GAS models without rolling windows. Dynamic quantile test resulted to reject the correct coverage of the GAS model in almost all cases for 1 % risk level. The less strict test at 5 % risk level confirmed correct specification for all GAS models with the length of the refit step equal to 1. In eight cases out of 72 (for both risk levels separately) the null hypothesis of correct model specification is rejected for GAS models with the Gaussian distribution but not rejected for GAS models with Student's t and skew-Student's t distribution. However, the test seemed to rather reject or not reject all the GAS models in the given time periods for a given market index. This is a well known issue of DQ test.

GAS models with the Student's t or skew-Student's t distribution outperformed the GAS models with the Gaussian distribution in the model performance according to the ratios of quantile loss function, which evaluate the corresponding VaR estimates for given pair of GAS models. There were few pairs where the models resulted to perform equally. GAS models with the Gaussian distribution were outperformed the most by the other two distributions for TOPIX in the first two periods. The ratios of FZL function, which take into account both estimated risk measures VaR and ES jointly led to less cases of equality of pairs of GAS models and less cases, where GAS models with the Gaussian distribution outperformed GAS models with the Student's t or skew-Student's t distributions.

Various additional exercises can be applied starting with the different setting of the time periods, the length of the refit steps for rolling windows or the size of the in-sample and out-of-sample. Additional parametric models can be used to estimate volatility or the model base can be changed to semi-parametric models without the observation distribution assumption. Patton et al. (2019) tested the length of the out-of-sample for NIKKEI and concluded that longer forecasting windows lead to worse model performances according to FZL function. Moreover, additional tests can be used to test the model performance and quality. An example can be Diebold-Mariano tests of differences in average loss (Patton et al., 2019).

References

ACERBI, Carlo; TASCHE, Dirk, 2002. Expected shortfall: a natural coherent alternative to Value at Risk. *Economic notes*. Vol. 31, no. 2, pp. 379–388.

ARDIA, David; BOUDT, Kris; CATANIA, Leopoldo, 2018. Downside risk evaluation with the R package GAS. *R Journal*. Vol. 10, no. 2, pp. 410–421.

ARDIA, David; BOUDT, Kris; CATANIA, Leopoldo, 2019. Generalized autoregressive score models in R: The GAS package. *Journal of Statistical Software*. Vol. 88, no. 6, pp. 1–28.

ARLT, Josef; ARLTOVÁ, Markéta, 2003. *Finanční časové řady*. Grada Publishing.

BERNARDI, Mauro; CATANIA, Leopoldo, 2015. Comparison of Value-at-Risk models: the MCS package. *arXiv preprint arXiv:1502.04472*.

BOLLERSLEV, Tim, 1986. Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*. Vol. 31, no. 3, pp. 307–327.

BURNHAM, Kenneth P; ANDERSON, David R, 2004. Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*. Vol. 33, no. 2, pp. 261–304.

CANINA, Linda; FIGLEWSKI, Stephen, 1993. The informational content of implied volatility. *The Review of Financial Studies*. Vol. 6, no. 3, pp. 659–681.

CATANIA, Leopoldo; BOUDT, Kris; ARDIA, David, 2019. *Generalized Autoregressive Score Models*. Available also from: `https://cran.r-project.org/web/packages/GAS/GAS.pdf`. R package version 0.3.0.

CONT, Rama, 2001. Empirical properties of asset returns: stylized facts and statistical issues.

COX, David R; GUDMUNDSSON, Gudmundur; LINDGREN, Georg; BONDESSON, Lennart; HARSAAE, Erik; LAAKE, Petter; JUSELIUS, Katarina; LAURITZEN, Steffen L, 1981. Statistical analysis of time series: Some recent developments [with discussion and reply]. *Scandinavian Journal of Statistics*, pp. 93–115.

CREAL, Drew; KOOPMAN, Siem Jan; LUCAS, André, 2008. A general framework for observation driven time-varying parameter models.

CREAL, Drew; KOOPMAN, Siem Jan; LUCAS, André, 2013. Generalized autoregressive score models with applications. *Journal of Applied Econometrics*. Vol. 28, no. 5, pp. 777–795.

DAI, Yu-Hong, 2002. Convergence properties of the BFGS algoritm. *SIAM Journal on Optimization*. Vol. 13, no. 3, pp. 693–701.

DATASTREAM, 2019. Thomson Reuters Datastream. Available also from: `http://financial.thomsonreuters.com/`. (Accessed: March 2019).

DELBAEN, Freddy, 2002. Coherent risk measures on general probability spaces. In: *Coherent risk measures on general probability spaces. Advances in finance and stochastics*. Springer, pp. 1–37.

DUMITRESCU, Elena-Ivona; HURLIN, Christophe; PHAM, Vinson, 2012. Backtesting value-at-risk: from dynamic quantile to dynamic binary tests. *Finance*. Vol. 33, no. 1, pp. 79–112.

ENGLE, Robert F, 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, pp. 987–1007.

ENGLE, Robert F; MANGANELLI, Simone, 2004. CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business & Economic Statistics*. Vol. 22, no. 4, pp. 367–381.

FISSLER, Tobias; ZIEGEL, Johanna F, et al., 2016. Higher order elicitability and Osband's principle. *The Annals of Statistics*. Vol. 44, no. 4, pp. 1680–1707.

FTSE., 2003. *Ground Rules for the Management of the UK Series of the FTSE Actuaries Share Indices*.

GAO, Chun-Ting; ZHOU, Xiao-Hua, 2016. Forecasting VaR and ES using dynamic conditional score models and skew Student distribution. *Economic Modelling*. Vol. 53, pp. 216–223.

GHALANOS, Alexios; KLEY, Tobias, 2019. *Univariate GARCH Models*. Available also from: `https://cran.r-project.org/web/packages/rugarch/rugarch.pdf`. R package version 1.4-1.

GILL, Richard D; WELLNER, Jon A; PRÆSTGAARD, Jens, 1989. Non-and semi-parametric maximum likelihood estimators and the von mises method (part 1)[with discussion and reply]. *Scandinavian Journal of Statistics*, pp. 97–128.

GNEITING, Tilmann; RANJAN, Roopesh, 2011. Comparing density forecasts using threshold- and quantile-weighted scoring rules. *Journal of Business & Economic Statistics*. Vol. 29, no. 3, pp. 411–422.

HARVEY, Andrew C, 2013. *Dynamic models for volatility and heavy tails: with applications to financial and economic time series*. Cambridge University Press.

HARVEY, Andrew; ORYSHCHENKO, Vitaliy, 2012. Kernel density estimation for time series data. *International Journal of Forecasting*. Vol. 28, no. 1, pp. 3–14.

HELLEINER, Eric, 2011. Understanding the 2007–2008 global financial crisis: Lessons for scholars of international political economy. *Annual review of political science*. Vol. 14, pp. 67–87.

HUANG, Dashan; YU, Baimin; LU, Zudi; FABOZZI, Frank J; FOCARDI, Sergio; FUKUSHIMA, Masao, 2010. Index-Exciting CAViaR: a new empirical time-varying risk model. *Studies in Nonlinear Dynamics & Econometrics*. Vol. 14, no. 2.

HUŠEK, Roman; PELIKÁN, Jan, 2003. *Aplikovaná ekonometrie: teorie a praxe*. Professional Publishing.

CHARLES, Amélie; DARNÉ, Olivier, 2014. Large shocks in the volatility of the Dow Jones Industrial Average index: 1928–2013. *Journal of Banking & Finance*. Vol. 43, pp. 188–199.

CHRISTOFFERSEN, Peter, 2010. Backtesting. *Encyclopedia of Quantitative Finance*.

CHRISTOFFERSEN, Peter F, 1998. Evaluating interval forecasts. *International economic review*, pp. 841–862.

IHAKA, Ross; GENTLEMAN, Robert, 1996. R: a language for data analysis and graphics. *Journal of computational and graphical statistics*. Vol. 5, no. 3, pp. 299–314.

KAUFMAN, Robert L, 2013. *Heteroskedasticity in regression: Detection and correction*. Sage Publications.

KUPIEC, Paul, 1995. Techniques for verifying the accuracy of risk measurement models. *The J. of Derivatives*. Vol. 3, no. 2.

LINSMEIER, Thomas J; PEARSON, Neil D, 2000. Value at risk. *Financial Analysts Journal*. Vol. 56, no. 2, pp. 47–67.

MANGANELLI, Simone; ENGLE, Robert F, 2004. A comparison of value-at-risk models in finance. *Risk measures for the 21st century*, pp. 123–44.

MORGAN, JP et al., 1996. Riskmetrics - technical document.

PATERNOSTER, Raymond; BRAME, Robert; MAZEROLLE, Paul; PIQUERO, Alex, 1998. Using the correct statistical test for the equality of regression coefficients. *Criminology*. Vol. 36, no. 4, pp. 859–866.

PATTON, Andrew J; ZIEGEL, Johanna F; CHEN, Rui, 2019. Dynamic semiparametric models for expected shortfall (and Value at Risk). *Journal of Econometrics*. Vol. 211, no. 2, pp. 388–413.

QIU, Debin, 2015. *Alternative Time Series Analysis*. Available also from: `https://www.rdocumentation.org/packages/aTSA`. R package version 3.1.2.

RAZALI, Nornadiah Mohd; WAH, Yap Bee, et al., 2011. Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. *Journal of statistical modeling and analytics*. Vol. 2, no. 1, pp. 21–33.

ROCKAFELLAR, R Tyrrell; URYASEV, Stanislav, 2002. Conditional value-at-risk for general loss distributions. *Journal of banking & finance*. Vol. 26, no. 7, pp. 1443–1471.

SATCHELL, Stephen; KNIGHT, John, 2011. *Forecasting volatility in the financial markets*. Elsevier.

SCHWARZ, Gideon et al., 1978. Estimating the dimension of a model. *The annals of statistics*. Vol. 6, no. 2, pp. 461–464.

SVOBODA, Martin, 2008. *Index investing*. Computer press.

TAYLOR, Stephen J, 1986. Forecasting the volatility of currency exchange rates. *International Journal of Forecasting*. Vol. 3, no. 1, pp. 159–170.

VASCONCELOS, Gabriel, 2017. Formal ways to compare forecasting models: Rolling windows. Available also from: `https://www.r-bloggers.com/formal-ways-to-compare-forecasting-models-rolling-windows/`.

WAGENMAKERS, Eric-Jan; FARRELL, Simon, 2004. AIC model selection using Akaike weights. *Psychonomic bulletin & review*. Vol. 11, no. 1, pp. 192–196.

Appendix

A. Supplementary results

Tables A.1 and A.2: Estimated parameters testing the skewness

DJIA, first 500 observations				
	Estimate	SE	t value	p-value
κ_1	0.0548	0.0538	1.018	0.154
κ_2	0.0171	0.0152	1.124	0.130
κ_3	0.2162	0.4061	0.532	0.297
κ_4	-2.8647	0.6967	-4.111	0.000
A_2	0.0897	0.0335	2.677	0.003
A_3	1.4729	3.4617	0.425	0.335
B_2	0.9008	0.0709	12.703	0.000
B_3	0.0000	0.0000	4042.770	0.000

Table A.1: Illustration 1: The values of estimated parameters testing the skewness

S&P, 2000-2009				
	Estimate	SE	t value	p-value
κ_1	0.0442	0.0167	2.638	0.004
κ_2	0.0013	0.0010	1.207	0.113
κ_3	-0.0882	0.0897	-0.983	0.162
κ_4	-1.9221	0.3460	-5.554	0.000
A_2	0.0450	0.0052	8.629	0.000
A_3	0.0523	0.3110	0.168	0.433
B_2	0.9918	0.0030	326.560	0.000
B_3	0.0128	0.0009	13.443	0.000

Table A.2: Illustration 2: The values of estimated parameters testing the skewness

Tables A.3 and A.4: Estimated parameters testing the kurtosis assuming the Student's t distribution

DJIA, 1985 – 1994				
	Estimate	SE	t value	p-value
κ_1	0.0626	0.0141	4.419	0.000
κ_2	-0.0055	0.0028	-1.964	0.024
κ_3	-0.7657	0.7713	-0.992	0.160
A_2	0.1089	0.0196	5.552	0.000
A_3	3.0295	3.8527	0.786	0.215
B_2	0.9929	0.0031	311.584	0.000
B_3	0.8009	0.2014	3.976	0.000

Table A.3: Illustration 1: The values of estimated parameters testing the kurtosis

S&P 500, 2010 – 2019				
	Estimate	SE	t value	p-value
κ_1	0.0852	0.0129	6.600	0.000
κ_2	-0.0293	0.0095	-3.070	0.001
κ_3	-0.9903	1.9174	-0.516	0.302
A_2	0.3554	0.0427	8.316	0.000
A_3	1.4140	2.8959	0.488	0.312
B_2	0.9649	0.0090	107.115	0.000
B_3	0.7183	0.5458	1.316	0.094

Table A.4: Illustration 2: The values of estimated parameters testing the kurtosis

Tables A.5 and A.6: Estimated parameters testing the kurtosis assuming the skew-Student's t distribution

	S&P 500, 2010 – 2019			
	Estimate	SE	t value	p-value
κ_1	0.0830	0.0135	6.111	0.000
κ_2	-0.0069	0.0033	-2.034	0.0209
κ_3	-0.0380	0.0960	-0.396	0.346
κ_4	-0.0637	0.0582	-1.094	0.136
A_2	0.0886	0.0099	8.906	0.000
A_4	1.3420	1.4755	0.909	0.181
B_2	0.9626	0.0091	105.671	0.000
B_4	0.9805	0.0168	58.226	0.000

Table A.5: Illustration 1: The values of estimated parameters testing the kurtosis distribution

	DJIA, 2010 – 2019			
	Estimate	SE	t value	p-value
κ_1	0.0828	0.0131	6.293	0.000
κ_2	-0.0103	0.0039	-2.608	0.004
κ_3	0.0149	0.0990	-0.151	0.439
κ_4	-0.0504	0.0506	-0.996	0.159
A_2	0.0935	0.0100	9.280	0.000
A_4	1.6576	1.3678	1.211	0.112
B_2	0.9574	0.0099	96.532	0.000
B_4	0.9839	0.0152	64.408	0.000

Table A.6: Illustration 2: The values of estimated parameters testing the kurtosis

Tables A.7 and A.8: The values of AIC, Δ and BIC for the GAS models

2000 – 2010

	DJIA			**S&P 500**		
Distribution	AIC	Δ	BIC	AIC	Δ	BIC
Student's t	7966	0	7996	8275	0	8304
skew-Student's t	7968	1.75	8004	8276	1.05	8311
Gaussian	8073	106.5	8097	8354	79.27	8377
	FTSE 100			**TOPIX**		
Distribution	AIC	Δ	BIC	AIC	Δ	BIC
Student's t	8177	0.23	8207	8962	0	8991
skew-Student's t	8177	0	8212	8966	4.12	9011
Gaussian	8199	22.74	8223	9005	42.98	9028

Table A.7: The values of AIC, Δ_h and BIC for each model over 2000 – 2010

2010 – 2019

	DJIA			**S&P 500**		
Distribution	AIC	Δ	BIC	AIC	Δ	BIC
Student's t	5267	0	5296	5457	0	5486
skew-Student's t	5269	1.99	5304	5459	1.84	5494
Gaussian	5418	151.28	5441	5629	171.84	5652
	FTSE 100			**TOPIX**		
Distribution	AIC	Δ	BIC	AIC	Δ	BIC
Student's t	5819	0	5847	6825	0	6854
skew-Student's t	5820	1.49	5855	6826	1.03	6861
Gaussian	5885	66.8	5908	6948	123.27	6971

Table A.8: The values of AIC, Δ_h and BIC for each model over 2010 – 2019

Tables A.9 and A.10: Average values of VaR and ES for GAS models

2000 – 2010

	DJIA		**S&P 500**		**FTSE 100**		**TOPIX**	
Distribution	$\overline{VaR}$	$\overline{ES}$	$\overline{VaR}$	$\overline{ES}$	$\overline{VaR}$	$\overline{ES}$	$\overline{VaR}$	$\overline{ES}$
Student's t	-2.82	-3.48	-2.98	-3.64	-2.82	-3.37	-3.25	-3.9
skew-Student's t	-2.84	-3.5	-3.02	-3.69	-2.87	-3.42	-3.23	-3.75
Gaussian	-2.61	-2.99	-2.77	-3.18	-2.73	-3.13	-3.12	-3.58

Table A.9: Average values of VaR and ES in the period 2000 – 2010

2010 – 2019

	DJIA		**S&P 500**		**FTSE 100**		**TOPIX**	
Distribution	$\overline{VaR}$	$\overline{ES}$	$\overline{VaR}$	$\overline{ES}$	$\overline{VaR}$	$\overline{ES}$	$\overline{VaR}$	$\overline{ES}$
Student's t	-2.11	-2.77	-2.21	-2.92	-2.23	-2.78	-2.88	-3.65
skew-Student's t	-2.11	-2.78	-2.22	-2.93	-2.25	-2.82	-2.93	-3.71
Gaussian	-1.85	-2.13	-1.94	-2.23	-2.04	-2.34	-2.69	-3.10

Table A.10: Average values of VaR and ES in the period 2010 – 2019

Tables A.11 and A.12: The results of the Z-test for GAS models

S&P 500: Student's t						
	1985–1994/2000–2010		**1985-1994/2010+**		**2000–2010/2010+**	
Coeff	Z-score	p-value	Z-score	p-value	Z-score	p-value
κ_1	0.193	0.577	-1.237	0.108	-1.311	0.095
κ_2	1.030	0.849	-1.959	0.025	-2.554	0.005
κ_3	-2.178	0.015	-0.334	0.369	1.924	0.973
A_2	-1.745	0.040	-4.223	0.000	-2.968	0.001
B_2	0.230	0.591	2.493	0.994	2.397	0.992
S&P 500: skew-Student's t						
	1985–1994/2000–2010		**1985-1994/2010+**		**2000–2010/2010+**	
Coeff	Z-score	p-value	Z-score	p-value	Z-score	p-value
κ_1	0.212	0.584	-1.176	0.120	-1.315	0.094
κ_2	-0.011	0.496	-1.291	0.098	-1.224	0.111
κ_3	-0.287	0.387	-0.015	0.494	0.277	0.609
κ_4	-2.181	0.015	-0.360	0.359	1.894	0.971
A_2	-1.755	0.040	-4.188	0.000	-2.937	0.002
B_2	0.239	0.595	2.500	0.994	2.395	0.992
S&P 500: Gaussian						
	1985–1994/2000–2010		**1985-1994/2010+**		**2000–2010/2010+**	
Coeff	Z-score	p-value	Z-score	p-value	Z-score	p-value
κ_1	0.621	0.733	-0.518	0.302	-1.128	0.130
κ_2	2.876	0.998	0.315	0.624	-2.605	0.005
A_2	1.202	0.885	-1.243	0.107	-2.462	0.007
B_2	-4.245	0.000	-1.308	0.095	3.693	0.999
FTSE 100: Student's t						
	1985–1994/2000–2010		**1985-1994/2010+**		**2000–2010/2010+**	
Coeff	Z-score	p-value	Z-score	p-value	Z-score	p-value
κ_1	0.328	0.629	0.493	0.689	0.156	0.562
κ_2	1.890	0.971	-0.251	0.401	-2.015	0.022
κ_3	-0.733	0.232	1.144	0.874	1.648	0.950
A_2	-0.632	0.264	-1.465	0.071	-1.035	0.150
B_2	-1.856	0.032	-0.121	0.452	1.808	0.965

Table A.11: Z-test of coefficient equality and corresponding p-values

FTSE 100: Gaussian

	1985–1994/2000–2010		1985-1994/2010+		2000–2010/2010+	
Coeff	Z-score	p-value	Z-score	p-value	Z-score	p-value
κ_1	0.320	0.626	0.461	0.678	0.146	0.558
κ_2	0.871	0.808	-0.581	0.281	-1.223	0.111
A_2	-5.551	0.000	-4.198	0.000	0.266	0.605
B_2	-1.668	0.048	0.353	0.638	1.954	0.975

TOPIX: Student's *t*

	1985–1994/2000–2010		1985-1994/2010+		2000–2010/2010+	
Coeff	Z-score	p-value	Z-score	p-value	Z-score	p-value
κ_1	0.512	0.696	-0.984	0.163	-1.292	0.098
κ_2	1.122	0.869	1.066	0.857	0.200	0.579
κ_3	-2.813	0.002	-1.123	0.131	1.911	0.972
A_2	2.428	0.992	0.552	0.710	-1.639	0.051
B_2	-0.882	0.189	0.691	0.755	1.427	0.923

TOPIX: Gaussian

	1985–1994/2000–2010		1985-1994/2010+		2000–2010/2010+	
Coeff	Z-score	p-value	Z-score	p-value	Z-score	p-value
κ_1	1.939	0.974	0.906	0.818	-0.943	0.173
κ_2	-0.892	0.186	-0.747	0.227	0.123	0.549
A_2	3.85	0.999	3.441	0.999	-0.455	0.325
B_2	-4.953	0.000	-2.860	0.002	1.850	0.968

	FTSE 100: skew-Student's *t*		TOPIX: skew-Student's *t*	
	2000-2010/2010+		1985-1994/2010+	
Coeff	Z-score	p-value	Z-score	p-value
κ_1	0.190	0.575	-0.528	0.299
κ_2	-1.000	0.159	-0.582	0.280
κ_3	0.357	0.639	0.379	0.648
κ_4	1.704	0.956	-1.211	0.113
A_2	-1.020	0.154	0.645	0.741
B_2	1.788	0.963	0.621	0.733

Table A.12: Z-test of coefficient equality and corresponding p-values

Tables A.13 and A.14: The resulted p-values of 1 % and 5 % DQ test

1985 – 1994:

DJIA	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % DQ	1 % DQ	5 % DQ	1 % DQ	5 % DQ	1 % DQ
Student's t	0.079	0.111	0.043	0.1	0.037	0.685
skew-Student's t	0.080	0.848	0.007	0.034	0.002	0.032
Gaussian	0.084	0.299	0.000	0.039	0.000	0.034
FTSE 100	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % DQ	1 % DQ	5 % DQ	1 % DQ	5 % DQ	1 % DQ
Student's t	0.085	0.994	0.206	0.995	0.202	0.994
Gaussian	0.086	0.993	0.056	0.957	0.074	0.996
TOPIX	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % DQ	1 % DQ	5 % DQ	1 % DQ	5 % DQ	1 % DQ
Student-t	0.128	0.052	0.113	0.053	0.114	0.053
Skew-Student-t	0.127	0.000	0.059	0.000	0.061	0.000
Gaussian	0.137	0.000	0.816	0.000	0.803	0.000

2000 – 2010:

DJIA	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % DQ	1 % DQ	5 % DQ	1 % DQ	5 % DQ	1 % DQ
Student's t	0.161	0.017	0.001	0.017	0.004	0.018
skew-Student's t	0.161	0.011	0.001	0.017	0.004	0.007
Gaussian	0.162	0.000	0.000	0.000	0.000	0.000
S&P 500	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % DQ	1 % DQ	5 % DQ	1 % DQ	5 % DQ	1 % DQ
Student's t	0.183	0.000	0.000	0.000	0.000	0.000
skew-Student's t	0.183	0.000	0.000	0.000	0.000	0.000
Gaussian	0.184	0	0.017	0.000	0.023	0.000
FTSE 100	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % DQ	1 % DQ	5 % DQ	1 % DQ	5 % DQ	1 % DQ
Student's t	0.168	0.000	0.260	0.000	0.257	0.000
skew-Student's t	0.167	0.000	0.248	0.000	0.244	0.000
Gaussian	0.166	0.000	0.329	0.000	0.347	0.000

Table A.13: p-values for 1 % and 5 % DQ test of the model correct coverage

2000 – 2010:

TOPIX	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % DQ	1 % DQ	5 % DQ	1 % DQ	5 % DQ	1 % DQ
Student's t	0.175	0.123	0.000	0.122	0.000	0.120
Gaussian	0.174	0.031	0.000	0.031	0.000	0.068

2010 – 2019:

S&P 500	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % DQ	1 % DQ	5 % DQ	1 % DQ	5 % DQ	1 % DQ
Student's t	0.099	0.000	0.308	0.000	0.281	0.000
skew-Student's t	0.099	0.000	0.255	0.000	0.247	0.000
Gaussian	0.100	0.000	0.181	0.000	0.181	0.000
TOPIX	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % DQ	1 % DQ	5 % DQ	1 % DQ	5 % DQ	1 % DQ
Student's t	0.145	0.005	0.000	0.005	0.000	0.005
skew-Student's t	0.145	0.017	0.000	0.033	0.000	0.017
Gaussian	0.147	0.001	0.000	0.000	0.000	0.000

Table A.14: 1 % and 5 % DQ test of the model correct coverage

Tables A.15 and A.16: The values of 1 % and 5 % QL function

1985 – 1994:

DJIA	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % QL	1 % QL	5 % QL	1 % QL	5 % QL	1 % QL
Student's t	0.079	0.026	0.079	0.026	0.079	0.026
skew-Student's t	0.08	0.026	0.08	0.027	0.08	0.027
Gaussian	0.084	0.026	0.085	0.026	0.085	0.026
S&P 500	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % QL	1 % QL	5 % QL	1 % QL	5 % QL	1 % QL
Student's t	0.075	0.024	0.076	0.024	0.075	0.024
skew-Student's t	0.076	0.024	0.076	0.024	0.076	0.024
Gaussian	0.081	0.025	0.08	0.025	0.08	0.025
TOPIX	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % QL	1 % QL	5 % QL	1 % QL	5 % QL	1 % QL
Student's t	0.128	0.038	0.128	0.038	0.128	0.038
skew-Student's t	0.127	0.038	0.127	0.038	0.127	0.038
Gaussian	0.137	0.043	0.138	0.044	0.138	0.044

2000 – 2010:

DJIA	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % QL	1 % QL	5 % QL	1 % QL	5 % QL	1 % QL
Student's t	0.161	0.044	0.161	0.044	0.161	0.044
skew-Student's t	0.161	0.044	0.161	0.044	0.161	0.044
Gaussian	0.162	0.046	0.162	0.046	0.162	0.046
S&P 500	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % QL	1 % QL	5 % QL	1 % QL	5 % QL	1 % QL
Student's t	0.183	0.049	0.183	0.049	0.183	0.049
skew-Student's t	0.183	0.049	0.182	0.048	0.183	0.049
Gaussian	0.184	0.052	0.184	0.052	0.185	0.053

Table A.15: Results of 1 % and 5 % QL functions in 1985 – 1994 and 2000 – 2010

2010 – 2019:

DJIA	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % QL	1 % QL	5 % QL	1 % QL	5 % QL	1 % QL
Student's t	0.099	0.031	0.099	0.031	0.099	0.031
skew-Student's t	0.099	0.031	0.099	0.031	0.099	0.031
Gaussian	0.100	0.032	0.100	0.031	0.100	0.032
S&P 500	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % QL	1 % QL	5 % QL	1 % QL	5 % QL	1 % QL
Student's t	0.099	0.033	0.099	0.033	0.099	0.033
skew-Student's t	0.099	0.033	0.099	0.033	0.099	0.033
Gaussian	0.100	0.034	0.100	0.034	0.101	0.034
FTSE 100	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % QL	1 % QL	5 % QL	1 % QL	5 % QL	1 % QL
Student's t	0.097	0.028	0.097	0.028	0.097	0.028
skew-Student's t	0.097	0.028	0.097	0.028	0.097	0.028
Gaussian	0.097	0.029	0.097	0.029	0.097	0.029
TOPIX	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % QL	1 % QL	5 % QL	1 % QL	5 % QL	1 % QL
Student's t	0.145	0.046	0.145	0.046	0.145	0.046
skew-Student's t	0.145	0.046	0.145	0.046	0.145	0.046
Gaussian	0.147	0.051	0.148	0.051	0.148	0.051

Table A.16: 1 % and 5 % QL functions in the period 2010 – 2019

Tables A.17 and A.18: The values of the ratios of 1 % and 5 % QL function

	refit step = 1		refit step = 5		refit step = 30	
DJIA	5 % QL ratio	1 % QL ratio	5 % QL ratio	1 % QL ratio	5 % QL ratio	1 % QL ratio
St/Gauss	0.94	0.99	0.93	0.98	0.93	0.97
skew-St/Gauss	0.95	1.02	0.95	1.01	0.94	1.01
S&P 500						
St/Gauss	0.94	0.98	0.94	0.98	0.94	0.98
skew-St/Gauss	0.94	0.99	0.95	0.98	0.95	0.98
FTSE 100						
St/Gauss	0.99	1.02	0.98	1.00	0.99	1.02
TOPIX						
St/Gauss	0.93	0.88	0.93	0.87	0.93	0.87
skew-St/Gauss	0.93	0.88	0.92	0.88	0.92	0.87

Table A.17: 1 % and 5 % ratios of the QL functions for the period from 1985 to 1994

	refit step = 1		refit step = 5		refit step = 30	
DJIA	5 % QL ratio	1 % QL ratio	5 % QL ratio	1 % QL ratio	5 % QL ratio	1 % QL ratio
St/Gauss	1.00	0.96	1.00	0.96	1.00	0.95
skew-St/Gauss	1.00	0.96	0.99	0.95	0.99	0.95
S&P 500						
St/Gauss	0.99	0.93	0.99	0.93	0.99	0.92
skew-St/Gauss	0.99	0.93	0.99	0.92	0.99	0.92
FTSE 100						
St/Gauss	1.01	0.99	1.01	0.98	1.01	0.99
skew-St/Gauss	1.01	0.98	1.01	0.98	1.01	0.98
TOPIX						
St/Gauss	1.00	0.99	1.00	0.99	1.00	0.99

Table A.18: 1 % and 5 % ratios of the QL functions for the period from 2000 to 2010

Tables A.19 and A.20: The values of the average FZL function at 1 % and 5 % risk level

1985 – 1994:

DJIA	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % QL	1 % QL	5 % QL	1 % QL	5 % QL	1 % QL
Student's t	0.449	0.950	0.451	0.958	0.445	0.935
skew-Student's t	0.465	0.977	0.471	0.991	0.463	0.995
Gaussian	0.510	0.985	0.516	1.013	0.520	1.020
S&P 500	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % QL	1 % QL	5 % QL	1 % QL	5 % QL	1 % QL
Student's t	0.406	0.883	0.407	0.887	0.405	0.883
skew-Student's t	0.412	0.886	0.407	0.875	0.409	0.872
Gaussian	0.469	0.939	0.460	0.943	0.462	0.940
TOPIX	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % QL	1 % QL	5 % QL	1 % QL	5 % QL	1 % QL
Student's t	0.896	1.298	0.896	1.297	0.895	1.292
skew-Student's t	0.897	1.310	0.896	1.309	0.891	1.305
Gaussian	0.948	1.479	0.950	1.481	0.954	1.494

2000 – 2010:

DJIA	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % QL	1 % QL	5 % QL	1 % QL	5 % QL	1 % QL
Student's t	1.135	1.548	1.136	1.549	1.135	1.544
skew-Student's t	1.135	1.553	1.132	1.547	1.133	1.538
Gaussian	1.145	1.642	1.145	1.644	1.146	1.638
S&P 500	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % QL	1 % QL	5 % QL	1 % QL	5 % QL	1 % QL
Student's t	1.268	1.648	1.268	1.649	1.271	1.654
skew-Student's t	1.266	1.644	1.264	1.639	1.269	1.648
Gaussian	1.284	1.781	1.285	1.784	1.290	1.796

Table A.19: 1 % and 5 % average FZL functions for selected indices

2010 – 2019:

DJIA	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % QL	1 % QL	5 % QL	1 % QL	5 % QL	1 % QL
Student's t	0.603	1.155	0.603	1.157	0.602	1.145
skew-Student's t	0.604	1.158	0.604	1.159	0.602	1.147
Gaussian	0.625	1.240	0.624	1.238	0.629	1.243
S&P 500	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % QL	1 % QL	5 % QL	1 % QL	5 % QL	1 % QL
Student's t	0.636	1.289	0.635	1.287	0.632	1.276
skew-Student's t	0.636	1.288	0.634	1.285	0.631	1.273
Gaussian	0.642	1.364	0.641	1.357	0.644	1.352
FTSE 100	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % QL	1 % QL	5 % QL	1 % QL	5 % QL	1 % QL
Student's t	0.626	0.993	0.625	0.991	0.624	0.988
skew-Student's t	0.625	0.991	0.624	0.99	0.624	0.988
Gaussian	0.629	1.055	0.629	1.054	0.628	1.056
TOPIX	**refit step = 1**		**refit step = 5**		**refit step = 30**	
	5 % QL	1 % QL	5 % QL	1 % QL	5 % QL	1 % QL
Student's t	1.009	1.495	1.009	1.497	1.009	1.493
skew-Student's t	1.009	1.490	1.009	1.491	1.009	1.486
Gaussian	1.040	1.694	1.044	1.696	1.045	1.697

Table A.20: 1 % and 5 % average FZL functions in the period 2010 – 2019

Tables A.21 and A.22: The values of the ratios of 1 % and 5 % average FZL function

	refit step = 1		refit step = 5		refit step = 30	
DJIA	5 % FZL ratio	1 % FZL ratio	5 % FZL ratio	1 % FZL ratio	5 % FZL ratio	1 % FZL ratio
St/Gauss	0.88	0.96	0.87	0.94	0.85	0.92
skew-St/Gauss	0.91	0.99	0.91	0.98	0.89	0.98
S&P 500						
St/Gauss	0.87	0.94	0.89	0.94	0.88	0.94
skew-St/Gauss	0.88	0.94	0.88	0.93	0.89	0.93
FTSE 100						
St/Gauss	0.97	1.02	0.95	0.97	0.98	1.03
TOPIX						
St/Gauss	0.95	0.88	0.94	0.88	0.94	0.86
skew-St/Gauss	0.95	0.89	0.94	0.88	0.93	0.87

Table A.21: 1 % and 5 % ratios of the average FZL function for the period from 1985 to 1994

	refit step = 1		refit step = 5		refit step = 30	
DJIA	5 % FZL ratio	1 % FZL ratio	5 % FZL ratio	1 % FZL ratio	5 % FZL ratio	1 % FZL ratio
St/Gauss	0.99	0.94	0.99	0.94	0.99	0.94
skew-St/Gauss	0.99	0.95	0.99	0.94	0.99	0.94
S&P 500						
St/Gauss	0.99	0.93	0.99	0.92	0.99	0.92
skew-St/Gauss	0.99	0.92	0.98	0.92	0.98	0.92
FTSE 100						
St/Gauss	1.00	0.97	1.00	0.96	1.00	0.97
skew-St/Gauss	1.00	0.96	1.00	0.96	1.00	0.96
TOPIX						
St/Gauss	1.00	0.98	1.00	0.98	1.00	0.98

Table A.22: 1 % and 5 % ratios of the average FZL function for the period from 2000 to 2010