Stockholms universitet

Applying a GARCH Model to an Index and a Stock

Jacob Lindberg

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2016:4
Matematisk statistik
Juni 2016

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Mathematical Statistics
Stockholm University
Bachelor Thesis **2016:4**
http://www.math.su.se

Applying a GARCH Model to an Index and a Stock

Jacob lindberg*

June 2016

Abstract

Can a GARCH model be used to accurately forecast Value at Risk of a single stock and an index portfolio? To answer this question, we fit a GARCH(1,1)-t model to the Samp;P500 index and to a bank's stock. Using this model with a rolling window procedure, we perform 1000 one-day ahead Value at Risk forecasts. These forecasts are then backtested—mainly using Christoffersen's conditional coverage test—after which we draw the conclusion that the model is indeed appropriate for our return series.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: jali0922@student.su.se. Supervisor: Mathias Lindholm, Filip Lindskog and Joanna Tyrcha.

Contents

1 Introduction

In this section we introduce Value at Risk, mention the current literature as well as state the purpose of this thesis.

Market risk is the exposure that an investor has to changes in the market prices. A measure of market risk is Value at Risk (VaR). This risk measure summarizes how much a firm may lose due to unfavourable price changes under normal circumstances. It is widely implemented, for example it is used within financial institutions as a risk measurement tool (Jorion, 2007:22-27). Also, regulators use VaR to enforce capital requirements depending on risk exposure. The populatity and thereby the importance of VaR grew after J.P. Morgan published RiskMetrics in 1994 (Longerstaey and Spencer, 1996). Risk management is a big field within finance. The literature on VaR is large, see for example Manganelli and Engle (2001), Giot and Laurent (2004), and Kuester et al. (2006).

Estimating VaR is difficult. Many decision have to be made regarding what models to be used. In this thesis we use statistical models to calculate VaR and decide if the models are appropriate.

1.1 Purpose and Research Question

VaR is a heavily used risk measure. It is especially important for portfolios. When calculating portfolio VaR researchers and practitioners frequently use GARCH models. In this thesis we study the difference between using a GARCH model a single stock and on an index portfolio. Given that VaR in practice often is used on portfolios rather than single assets, the practitioner might be interested in knowing to what extent the risk models used in practice can be applied to a single asset as opposed to a portfolio such as an equity index.

The specific research question answered in this thesis is the following: Can a standard GARCH(1,1) model with t distributed innovations be used to accurately forecast VaR for a stock and an index?

2 Time Series

In this section we begin by introducing some fundamental definitions in the first subsection. In the second subsection we present some statistical tests used in Time Series Analysis. In the third subsection we introduce GARCH models.

2.1 Theoretical Background

In this subsection we define returns, conditional variance, ACF, white noise, volatility clustering and covariance stationarity.

2.1.1 Returns

Let P_t be the price of an asset (e.g. stock or index or portfolio) at time t. We define the return at time t as

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

The log returns are used, rather than the arithmetic returns, because of the log returns statistical properties. (Tsay 2013:4f)

2.1.2 Conditionality

Define $\mathcal{F}_{t-1}$ as the information set up until time $t-1$. In other words

$$\mathcal{F}_{t-1} = \{r_1, ...r_{t-1}\}.$$

Define the conditional mean as

$$\mu_t = E[r_t|\mathcal{F}_{t-1}].$$

Define the conditional conditional variance by

$$\sigma_t^2 = \text{var}(r_t|\mathcal{F}_{t-1}).$$

We will use statistical models to forecast μ_t and σ_t. The conditionality is important but prevalent in time series analysis and therefore we will refer to μ_t as the *mean* and to σ_t^2 as the *variance*. By dropping the prefix "conditional" we save time and ease the notation. It is important, however, the concept of conditionality. Especially so in section 2.3.4 where we describe how the conditional distribution, not the unconditional distribution, is used when estimating the parameters from the returns data using MLE.

2.1.3 Autocorrelation Function

The autocovariance is defined as

$$\gamma_k = \text{cov}(r_t, r_{t-k})$$

where r_{t-k} is the returns k^{th} lag, and of course k is a positive integer. In a stationary time series $\{r_t\}$ the autocorrelation is defined as

$$\rho_k = \frac{\text{cov}(r_t, r_{t-k})}{\sqrt{\text{var}(r_t)\,\text{var}(r_{t-k})}} = \frac{\gamma_k}{\gamma_0} \tag{2}$$

where the second equality holds because $\text{var}(r_t) = \text{var}(r_{t-k})$ in a stationary time series. (Tsay, 2013:45-47)

2.1.4 White Noise Process

Let z_t be a time series, then

$$z_t \sim \mathit{WhiteNoise}(0,\ \sigma_z^2)$$

if and only if $\gamma_0 = \sigma_z^2 \in \mathbb{R}$ and $\gamma_k = 0$ for lags $k > 0$. In other words: if a process has zero mean and no covariance between its values at different times, it is said to be a white noise process.

A white noise process can follow different distributions. One example is the Student's t-distribution (Brooks, 2002:232f). The PDF of a Student's t-distribution is

$$f(x) = \frac{\Gamma\left(\frac{v+1}{2}\right)}{\sqrt{\beta v \pi}\,\Gamma\left(v/2\right)} \left(1 + \frac{(x-\alpha)^2}{\beta v}\right)^{-\left(\frac{v+1}{2}\right)} \tag{3}$$

2.1.5 Volatility Clustering

Mandelbrot noted in 1963 that large changes tend to be followed by large changes—of either sign. Likewise, he observed that small changes by small changes tend to be followed by small changes—of either sign. This is what has come to be referred to as volatility clustering. Mandelbrot's observation implies that r_t^2 and $|r_t|$ should have high autocorrelation, which is in fact observed in the ACF plot that we will introduce later on.

2.1.6 Covariance Stationarity

Let $\{r_t\}$ be a time series process. For $\{r_t\}$ to be covariance stationary, three conditions must be satisfied we must have that

I $E[r_t] = \mu$

II $E[(r_t - \mu)^2] = \gamma_0 < \infty$

III $\text{cov}(r_t, r_{t+k}) = \gamma_k$

for all times $t \in \mathbb{Z}$ and lags k. That is to say, a time series is covariance stationary when (I) the mean, (II) the variance and (III) the autocovariance structure are all time invariant. When a time series is non-stationary, a unit root is present. (Brooks 2002:230f)

2.2 Statistical Tests

In this subsection we bein with introducing a statistical tests for stationarity (ADF-test). Then we discuss normality tests (Q-Q-Plots) and how to test for autocorrelation (Ljung-Box test and ACF plot).

2.2.1 Test of Unit Root

Let $\{r_t\}$ be a time series. The Augmented Dicke Fuller regression is

$$\Delta r_t = \alpha + \beta t + \gamma r_{t-1} + \sum_{k=1}^{p} \delta_k \Delta r_{t-k} + u_t$$

where α is a constant, u_t is the regression error term, and β is the coefficient on the time trend. In the Augmented Dicke Fuller test (ADF-test) we use this regression to test

$$H_0 : \gamma = 0 \quad \text{Unit root is present i.e. non-stationarity}$$

against the one-sided alternative $H_1 : \gamma < 0$ so the alternative states there is no unit root i.e. that the time series process is stationary. The test statistic

$$DF^{obs} = \frac{\hat{\gamma}}{Std.Err(\hat{\gamma})}$$

is compared with $DF^{critical} = -3.4$ which is the critical value at a significance level of 5 %. The more negative DF^{obs} the stronger the rejection of the null. If $DF^{obs} < DF^{critical}$ then the null is rejected at a 5 percent significance level, leading us to conclude that there is no unit root present. (Dickey & Fuller, 1979)

2.2.2 Test Normality

A Q-Q plot is often used to determine if data is normally distributed. It can also be used to test if data follows a t distribution (or other theoretical distributions). If the data follows the theoretical distribution, then the points in the Q-Q Plot lie on a straight line. In the exploratory data anlysis on page 19 in Figure 4 we draw four Q-Q Plots.

2.2.3 Test of Autocorrelation

To evaluate the autocorrelation of the returns and squared returns we can either use a Ljung-Box test or an ACF plot, or both

The null hypothesis in the Ljung-Box test is

$$H_0 : \gamma_k = 0$$

for the lags $k = 1, ..., K$ against $H_1 : \exists\ k\ s.t.\ \gamma_k \neq 0$. The test statistic in the Ljung-Box test is

$$Q(K) = T(T+2)\sum_{k=1}^{K} \frac{\gamma_k^2}{T-l} \sim^{H_0} \chi^2(K) \tag{4}$$

where T is the number of observations. We will use a type I error of 5% and $K = \ln(T)$. (Ljung and Box, 1978:297-303)

An ACF plot can be used as a visual test if data is autocorrelated. In and ACF plot γ_k is displayed on the y-axis and k on the x-axis. The dotted horizontal lines in an ACF plot are for $-1/T \pm 1.96/\sqrt{T}$ where 1.96 comes from $\lambda_{0.95}$. If a value γ_k is above the dotted line we deem lag k to be statistically significantly different from zero at the 5 percent level. On page 21 we have drawn four ACF plots.

The ACF will be used to both on returns and on squared returns, as will be discussed in section 4.1 where we do an exploratory data analysis.

2.3 GARCH Models

In this subsection we introduce the GARCH model, we are used to forecast volatility. Then we define standardized residuals. We specify how the GARCG parameters are estimated, and discuss information criterions as well as unconditional variance and persistence. Following that, we introdce the rolling window procedure. Lastly, RME is defined.

2.3.1 The GARCH model

The autoregressive conditional heteroscedastic (ARCH) model developed by Engle (1982). This model was generealized to a Generalized ARCH (GARCH) by Bollerslev (1986). For details see Tsay (2010:109-173).

In the GARCH modelling framework we let

$$r_t = \mu_t + \varepsilon_t = \mu_t + \sigma_t z_t \tag{5}$$

where ε_t is the innovation (also called shock). The random variable z_t will—in this thesis—be assumed to follow a t distribution. If the variance $\sigma_t^2 = \text{var}(r_t|\mathcal{F}_{t-1})$ can be described by

$$\sigma_t^2 = \omega + \alpha_1\varepsilon_{t-1}^2 + \beta_1\sigma_{t-1}^2 \tag{6}$$

then we say that ε_t follows a GARCH(1,1) process, given that the constraints $\omega > 0$, $\alpha_1 \geq 0$, $\beta_1 \geq 0$ and $\alpha_1 + \beta_1 < 1$ are fulfilled. Equation (6) is also called the volatility equation whereas equation (5) is called the mean equation. For daily data we can often assume that $\mu_t = 0$ for the mean equation, which implies $\varepsilon_{t-1} = r_{t-1}$.

2.3.2 The ARCH model

The ARCH(m) model was developed by Engle in 1982. It has a similar setup as the GARCH model but differs mostly in the volatility equation which is

$$\sigma_t^2 = \omega + \sum_{k=1}^{m} \alpha_k \varepsilon_{t-k}^2$$

in other words that today's variance depend on past squared innovations. The value of m is often decided using maximum likelihood estimation.

2.3.3 Standardized Residuals

In the GARCH model we impose a theoretical distribution on the white noise term z_t for example a t distribution. This assumption is tested by plotting the standardized residuals

$$\tilde{\epsilon}_t = \frac{\hat{\epsilon}_t}{\hat{\sigma}_t} \tag{7}$$

where $\hat{\epsilon}_t = r_t - \hat{\mu}_t$. By comparing the white noise term

$$z_t = \frac{\varepsilon_t}{\sigma_t}$$

and the formula for $\tilde{\epsilon}_t$ we see that it is logical to evaluate if the assumed distribution of z_t is an appropriate assumption by plotting $\tilde{\epsilon}_t$.

To evaluate whether the standardized residuals seem to be white noise, the autocorrelation function, in which there should be no apparent autocorrelation, as well as the Q-Q plot, where the standardized residual should follow the assumed distribution. This is done in the Data Analysis under subsection 4.3.

2.3.4 Parameter Estimation: MLE

Given that the random variable Y takes on the value y the conditional density of the random variable X is $f_{x|y}(x;\theta) = f_{x,y}(x,y;\theta)/f_y(y;\theta)$ or equivalently

$$f_{x,y}(x,y;\theta) = f_y(y;\theta) \cdot f_{x|y}(x;\theta).$$

This fact from probability theory is used in MLE. For a time series $\{r_t\}$ we have

$$f(r_1, r_2, ..., r_T; \theta) = f(r_1;\theta) \cdot f(r_2 \mid r_1;\theta) \cdot f(r_3 \mid r_1, r_2;\theta) \cdot ... \cdot f(r_T \mid r_1, r_2, ..., r_{T-1};\theta)$$

or more compactly written

$$f(r_1, r_2, ..., r_T; \theta) = f(r_1;\theta) \cdot \prod_{t=2}^{T} f(r_t \mid r_1, ..., r_{t-1};\theta)$$

where $f(r_1;\theta)$ is the marginal density of the very first observation, and $f(r_t \mid r_1, ..., r_{t-1};\theta)$ is the conditional distribution.

If the conditional distribution is $N(\mu_t, \sigma_t)$ then $\theta = (\mu_t, \sigma_t)$ so the likelihood is

$$f(r_1, r_2, ..., r_T; \theta) = f(r_1;\theta) \cdot \prod_{t=2}^{T} \frac{1}{\sigma_t\sqrt{2\pi}} \exp\left(-\frac{(r_t-\mu_t)^2}{2\sigma_t^2}\right)$$

implying that the log-likelihood is

$$\ln f(r_1, r_2, ..., r_T; \theta) = \ln f(r_1;\theta) - \frac{1}{2}\sum_{t=2}^{T}\left(\ln(2\pi) + \ln(\sigma_t) + \frac{(r_t-\mu_t)^2}{\sigma_t^2}\right) \tag{8}$$

(Tsay, 2010:15). The maximum likelihood estimate $\hat{\theta}_{ML}$ is the value of θ that maximizes equation (8) in other words we define the maximum likelihood estimate as

$$\hat{\theta}_{ML} = \arg\ \max_{\theta} \ln f(r_1, r_2, ..., r_T; \theta). \tag{9}$$

2.3.5 AIC and BIC

The parameters are estimated so that the log-likelihood function is maximized. Even so, the log-likelihood is not just a tool that gives us the estimated parameters—it can also be used when comparing two models. If model w have a higher likelihood than model b then the former fits better to the data. Of course, we want parsimonious models and that is where information criterions come in. An information criterion is a function of the likelihood value (high value is good) and the number of parameters in the model (high number is bad). A few frequently used information criterions are AIC, BIC, Shibata and Hannan-Quinn. We will use AIC and BIC. The definitions are

$$\text{AIC} = 2p - 2\hat{\ell} \quad \text{and} \quad \text{BIC} = -2\hat{\ell} + p\ln(T) \tag{10}$$

where T is the sample size, p is the number of estimated parameters in the model, and $\hat{\ell}$ is the log-likelihood value. As we see in the formulas, AIC and BIC penalizes a model with many parameters but rewards a high likelihood.

2.3.6 Unconditional Variance and Persistence

The weight γ is put on the unconditional variance V_L. We have that

$$\omega = \gamma V_L. \tag{11}$$

This form is used for estimating the parameters, i.e. ω, α_1 and β_1. The relationship between the parameters are

$$\gamma + \alpha_1 + \beta_1 = 1 \text{ and } \alpha_1 + \beta_1 < 1$$

so in order for the weight on γ to take on a positive value needed for a stationary process. If $\gamma \leq 0$ then the process will not have mean reversion. Mean reversion is when the level of variance returns to normal levels after experiencing some shock. (Hull, 2012:502f)

We will also use the concept of half life. Define

$$halflife = \frac{-\ln(2)}{\ln Persistence}$$

$$Persistence = \alpha_1 + \beta_1$$

which is interpreted as the number of days it takes for half of the reversion back to the unconditional volatility V_L.

2.3.7 Rolling Window Procedure

When calculating the GARCH parameters and forecast the volatility, we will use a rolling window procedure.

Two Windows

To backtest the models the data set is divided into two different parts. For easier notation we distinguish between t and t^* according to Table 1.

t^*	t	window name
1	$1 - n(w_E)$	estimation window
...	...	estimation window
$n(w_E)$	0	estimation window
$n(w_E) + 1$	1	forecasting window
...	...	forecasting window
$n(w_E + w_F)$	$n(w_F)$	forecasting window

Table 1: Window definitions.

In the estimation window w_E the parameters for the GARCH model are estimated. We use these estimates to forecast volatility and mean—which are then used to calculate VaR_t. The forecasting window goes from $t = 1$ to $t = n(w_F)$ as defined in Table 1. In this thesis $n(w_F) = 1000$.

Volatility and Mean Forecasts

The estimate of $\hat{\sigma}_0$ is set equal to the sample standard deviation in the estimation window. The first estimate of the variance can then be calculated using

$$\hat{\sigma}_1^2 = \hat{\omega} + \hat{\alpha}_1 \varepsilon_0^2 + \hat{\beta}_1 \sigma_0^2$$

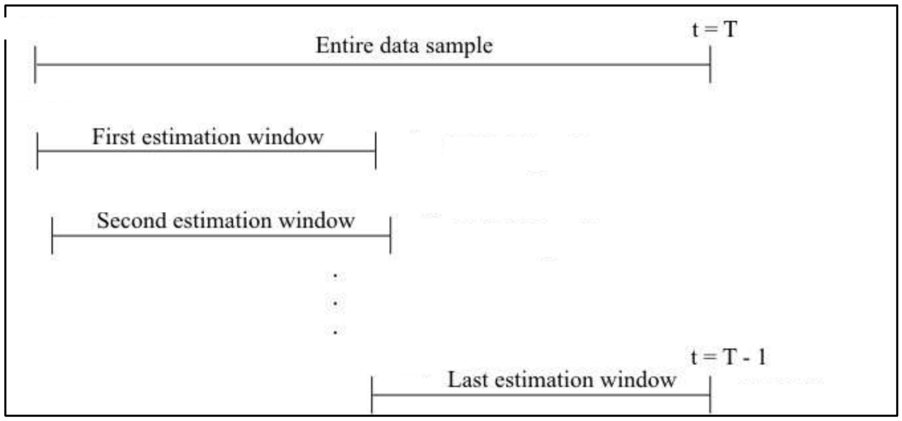

Figure 1: Rolling window procedure. Using the first estimation window we forecast the first VaR. This continues until the last estimation window which is used to forecast the last VaR.

where ε_0^2 is the last squared innovation in the estimation window (or r_0^2 if the mean is assumed to be zero).

This procedure continues until we have forecasts of $\hat{\sigma}_t^2 \quad \forall t \in w_F$. The procedure is illustrated in Figure 1. Using the first estimation window we forecast the first VaR. This continues until the last estimation window which is used to forecast the last VaR. In the forecasting window we get the VaR foreacsts $VaR_1, VaR_2, ..., VaR_{1000}$ and these are then backtested.

Updating Parameters

In the previous paragraph we saw how to use these parameters to forecast the one-day-ahead VaR. In order for the parameters to not be outdated and produce bad forecasts, we will do a re-estimation of these parameters every 20 days—so that the values of these parameters will not be the same throughout w_F. The first parameters estimated using w_E are labelled by a (1) in the exponent to indicate the first estimate of the parameters so that we get

$$\hat{\omega}^{(1)} \quad \hat{\alpha}_1^{(1)} \quad \hat{\beta}_1^{(1)}.$$

For the first 20 days of $t \in w_F$ we use these parameters but for the next 20 days (i.e. day 21 to 40) we use $\hat{\omega}^{(2)}, \hat{\alpha}_1^{(2)}, \hat{\beta}_1^{(2)}$ and the 20 days after that (i.e. day 41 to 61) we use $\hat{\omega}^{(3)}, \hat{\alpha}_1^{(3)}, \hat{\beta}_1^{(3)}$ and so on. The length of w_F is 1000 and we refit every 20 days making it 1000/20=50 different volatility estimates for every parameter.

2.3.8 Root Mean Error

The GARCH model is used to produce volatility forecasts. To compare the forecasts between models, we need a measure. Root Mean Error (RME) is a forecast performance measure. We define

$$error_t = \hat{\sigma}_t^2 - r_t^2 \tag{12}$$

where $\hat{\sigma}_t^2$ is the forecasted variance and as a proxy for the true variance $\sigma_t^2 = \text{var}(r_t \mid \mathcal{F}_{t-1})$ we use squared realized return r_t^2. It is a noisy proxy but the reasoning behind using it is that when returns are close to zero, as it is with daily data, we have that $\sigma_t^2 \approx r_t^2$ because $E[(r_t - \mu_t)^2 \mid \mathcal{F}_{t-1}] \approx E[(r_t - 0)^2 \mid \mathcal{F}_{t-1}]$. Given this definition of the error, we have that

$$RME = \sqrt{\frac{1}{n(w_F)} \sum_{t=1}^{n(w_F)} error_t} \tag{13}$$

where $n(w_F) = 1000$ is the length of the forecasting window used in this thesis. RME have well documented statistical properties. It also has a nice economic interpretation that we will introduce later on. The forecast measure RME will be used together with the Kupiec's unconditional coverage test and Christoffersen's conditional coverage test which are both introduced under forthcoming section 3.

3 Value at Risk

In this section we discuss regulation, define VaR and describe how to calcualte it. Backtesting using Kupiec's unconditional coverage test and Christoffersen's conditional coverage Test is thoroughly described.

3.1 Regulation

VaR is *"the worst loss over a target horizon that will not be exceeded with a given level of confidence"*, Jorion (2007:viii). This risk measure can be used to estimate the risk in an investor's portfolio or for an entire financial institution. During times of financial crisis we are reminded of the importance of accurate risk management, and VaR is one of the most heavily used ways of measuring risk. (Duffie and Pan, 1997).

For financial institutions to be prepared to incur losses, regulators enforce capital requirements. One important regulation is the Basel Accords. The accords are developed by Basel Committee on Banking Supervision. It is currently adopted by for example the United States and the European Union. This regulation pressures the financial institution's measurements of their risk via backtesting. Financial institutions that fails to meet the validity requirements are penalized. This makes is essential for banks to accurately calculate their VaR.

3.2 Formula for Value at Risk

Even though the quote from Jorion above is correct, we need a more precise definition of VaR in order to be able to work with it. As mentioned in 2.1.1, losses recorded in L_t will be denoted as positive numbers. VaR is concerned with the upper of the loss distribution. VaR is defined as

$$VaR_{1-p} = \inf\{L_t \mid CDF_t(L_t) \geq 1 - p\}. \tag{14}$$

In other words

$$Pr[L_t \leq VaR_{1-p}] \geq 1 - p$$

In this thesis we use a value of $p = 0.05$. The more correct $VaR_{0.95}$ is sometimes abbreviated VaR for ease of notation. Figure 2 displays the VaR as the dotted line.

Let $\hat{\mu}_t$ be the estimated mean in the rolling window. Let $t_{0.95}(\nu)$ be a 95 percent quantile from the t distribution with $\hat{\nu}$ degrees of freedom—for example $t_{0.95}(7) \approx 1.9$ and $t_{0.95}(999) \approx 1.65$. Let $\hat{\sigma}_t$ be the estimated volatility from the GARCH model. Given these three values $\hat{\mu}_t$, $t_{0.95}(\hat{\nu})$ and $\hat{\sigma}_t$ the formula for calculating VaR is

$$VaR_t = \hat{\mu}_t - t_{0.95}(\hat{\nu})\hat{\sigma}_t. \tag{15}$$

The knowledgeable reader might know the usual VaR formula is $VaR_t = (\hat{\mu}_t - t_{\hat{\nu},0.95}\hat{\sigma}_t)Value_{t-1}$ where $Value_{t-1}$ is the asset (or portfolio) value. In this thesis we will assume $Value_t = 1$ for simplicity, and this assumption can be made without loss of generality.

3.3 Backtest Value at Risk

The estimated VaR needs to be evaluated to decide if they are good or not. One way to evaluate it is to define an indicator variable I_t for the forecasted at time VaR_t as

$$I_t = \begin{cases} 0, & \text{if } L_t \geq \text{VaR}_{1-p,t} \\ 1, & \text{if } L_t < \text{VaR}_{1-p,t} \end{cases}. \tag{16}$$

This indicator variable I_t counts every time the observed return is lower than the forecasted VaR, so when $I_t = 1$ we say that a *violation* has occured.

An effective VaR measure satisfies two properties according to Christoffersen (1998). Property 1: $E[I_t] = p$ in other words the expected number of violations is indeed the stated p. Property 2: $E[I_t \mid \mathcal{F}_{t-1}] = p$ in other words that violations not are clustered. To test these properties we will be using Kupiec's Unconditional Coverage test for property 1 and Christoffersen's Conditional Coverage test for property 2. The two tests are summarized by Jorion (2007:143-152).

3.4 Unconditional Coverage Test

To test property 1 we use the unconditional coverage test—from here on abbreviated uc-test. This test evaluates if the actual number of violations are the same as the stated number. (Kupiec, 1995)

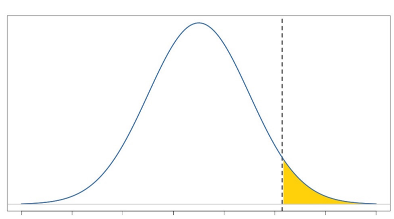

Figure 2: PDF of a random loss variable L_t. VaR is at the dotted line, with 5 percent density to the right of the dotted line.

Let p be the VaR coverage ratio from the subsection above. Let T be the number of out-of-sample estimates and N be the number of violations. [1] We test the null hypothesis

$$H_{0,cc} : p = 0.05$$

against $H_1 : p \neq 0.05$. The test statistic is

$$LR_{uc} = -2 \ln \frac{(1-p)^{T-N} p^N}{\left(1 - \frac{N}{T}\right)^{T-N} \left(\frac{N}{T}\right)^N} \tag{17}$$

Under the null hypothesis LR_{uc} is asymptotically distributed as $\chi^2(1)$. We will perform the test for a 95% confidence region, so the critical value is $LR_{uc}^{critical} = 3.84$. A higher LR value leads to the rejection of the null, which is logical since a LR is high when the numerator and denomator are different indicating that $p \neq 0.05$.

We wish to accept the null hypothesis, since we want the actual number of violation to be close to the stated 5 percent. The higher the p-value of the cc test the better the model is, because a high p-value indicates that the null is in fact correct.

3.5 Conditional Coverage Test

To test property 1 and 2 we use a the conditional coverage test—from here on abbreviated cc test. This test evaluates actual number of violations are the same as the stated number *and* if violations are clustered. (Christoffersen, 1998)

The violations I_t can be modelled with a Markov chain having transition probabilities

$$\begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix} \tag{18}$$

where π_{00} is the probability to go from a non violation at day $t-1$ to a non-violation at day t. In other words the first number in the subscript refers to the value of I_{t-1} and the second number in the subscript refers to the value of I_t. More formally this can be written as $\pi_{ij} = Pr(I_t = j \mid I_{t-1} = i)$. Obviously $\pi_{00} + \pi_{01} = 1$ and $\pi_{10} + \pi_{11} = 1$.

We want to test if violations are clustered or not, and it is possible to test this at the same time as we test if the stated p is the correct one. A cc test achieves this. The null hypothesis is

$$H_{0,cc} : \begin{bmatrix} \pi_{00} & \pi_{01} \\ \pi_{10} & \pi_{11} \end{bmatrix} = \begin{bmatrix} 1-p & p \\ 1-p & p \end{bmatrix} \tag{19}$$

and in this thesis the coverage rate is $p = 0.05$. Let T_{ij} be the actual number of

[1] In this thesis T has the value of 1000 and since $p = 5\%$ we will observe $N = 50$ or $N = 49$ or $N = 51$ for a good model.

times $I_{t-1} = i$ is followed by $I_t = j$ so that T_{ij} is the "empirical" π_{ij}. The test statistic is

$$LR_{cc} = -2 \ln \frac{(1-p)^{T_{00}} p^{T_{01}} (1-p)^{T_{10}} p^{T_{11}}}{(1-\pi_{00})^{T_{00}} \pi_{01}^{T_{01}} (1-\pi_{10})^{T_{10}} \pi_{11}^{T_{11}}}. \tag{20}$$

Christoffersen showed that under the null LR_{cc} is distributed as $\chi^2(2)$. For a 95 percent confidence level we reject the null hypothesis if $LR_{cc} > LR_{cc}^{critical} = 6$. Thus we can test property 1 and property 2 at the same time with a cc test. In this sense, the cc test incorporates the uc-test.

We wish to accept the null hypothesis. As with the simpler uc-test we have for the cc test that a high p-value is a sign of a good model. The reason is that we do not want violations to be clustered and we want to have the $\pi_{01} = \pi_{11} = 0.05$ since we want the actual number of violation to be close to the stated 5 percent.

As a final warning, from the definition $\pi_{ij} = Pr(I_t = j \mid I_{t-1} = i)$ we see that the definition of violation is crude since it only have a "memory" of one day. For financial reasons it might be interesting to look two days back and consider the probability $Pr(I_t = j \mid I_{t-2} = i)$ but this is not implemented and is a flaw.

4 Data Analysis

In this section begin with an exploratory data analysis. Then a GARCH(1,1)-t model is fitted to data from the estimation window of the two return series, after which the differences are discussed. Lastly, we test if the GARCH model assumptions are met.

4.1 Exploratory Data Analysis

In this subsection we perform seven steps, each with a conclusion that takes us one step further in arriving at what model is suitable for modelling the return series.

Step 1: Plot prices together with returns and test stationarity using the ADF-test. Conclusion: Returns are stationary.

Step 2: Tabulate summary statistics and Q-Q plots of the return. Conclusion: Returns are closer to a t distribution than a normal distribution.

Step 3: Plot ACF of return to view autocorrelation and test it using Ljung-Box test. Conclusion: Returns are correlated.

Step 4: Test if the mean return is zero with a t-test. Conclusion: Returns are not statistically significantly different from zero, hence we use a constant for the mean equation.

Step 5: Plot ACF of squared return to view autocorrelation and test it using Ljung-Box test. Conclusion: Squared returns are correlated.

Step 6: Test if the mean squared return is zero with a t-test. Conclusion: Squared returns are statistically significantly different from zero, hence we need to model the variance equation using an ARCH or GARCH model.

Step 7: Discuss if ARCH or GARCH should be used. Conclusion: A GARCH model is better than an ARCH model.

4.1.1 Description of Data

Data from 2001-01-04 to 2016-01-01 on the daily prices of Morgan Stanley stock and the GSPC index is downloaded from Yahoo Finance. From the daily prices we calculate the returns. We reserve the 1000 data points (approximately four years) for backtesting making the estimation window go from 2001-01-04 to 2014-01-07. GSPC is often called S&P500. It consists of the 500 most frequently traded stocks in the US. The GSPC index is weighted by market capitalisation. Note that Morgan Stanley is included in the index.

4.1.2 Stationary of Prices and Returns

In Figure 3 we plot the prices and returns of the index and the stock. During the period 2009 to 2013 the index seem to trend more upwards than stock does. According to the different scaling of the y-axis we see that stock is more volatile than the index is.

We need the data to be stationary. According to an ADF-test the prices are not stationary since the p-values are close to zero. The return series is stationary so we can model the return series.

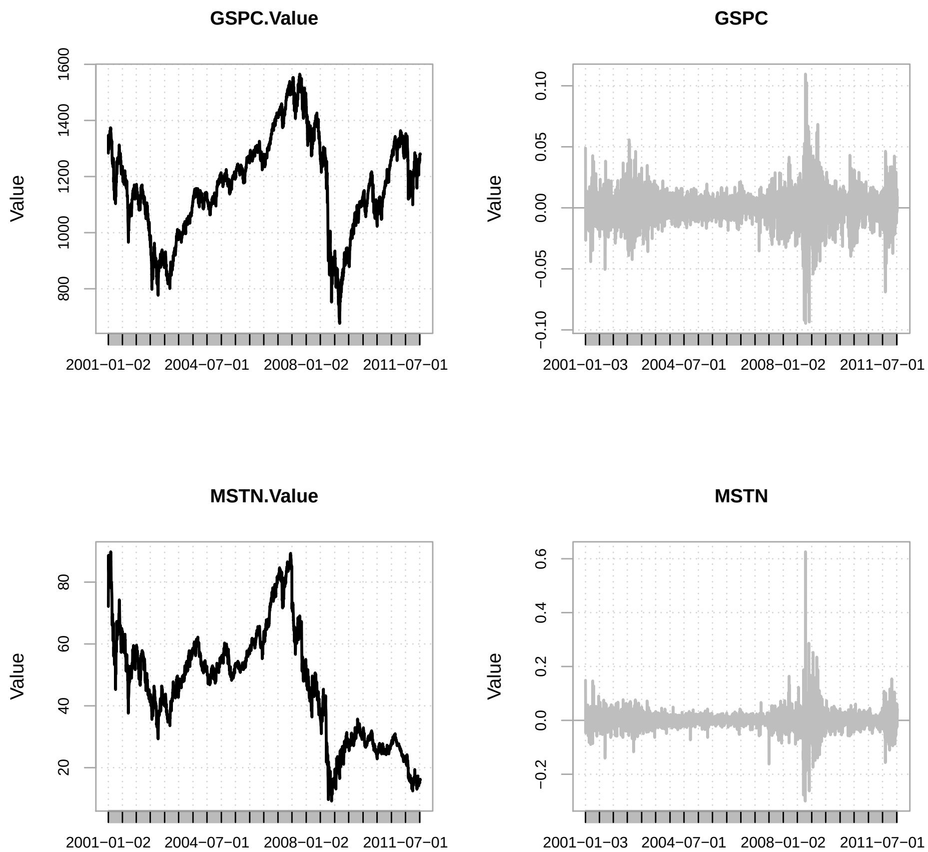

Figure 3: Price (left) and return (right) for the index (upper) and the stock (lower). Note that the different scales of the y-axis. Only data in estimation window have been used.

4.1.3 Summary Statistics and Q-Q Plots

In Table 2 the volatility of GSPC is lower than that the bank's which is logical since GSPC is an index and carries lower risk than an individual stock. Both means are close to zero, as expected since we have daily data and there is no time for a return to occur. Moreover, neither mean is statistically significantly different from zero—performing a one-sample t-test of the null hypothesis "mean is zero" give a p-value of less than 1 percent for both assets. Hence the higher return of the stock is not statistically significant. Lastly, the table presents skewness and kurtosis. Q-Q plots will be used to draw conclusions that data is not normally distributed. The kurtosis is much larger than 3 which is the kurtosis of a standard normal, in other words both return series exhibit fat tails.

	n.Obs	NA	Min	Mean	Max	Var	Stdev	Skew	Kurt
index	2772	0	-0.0947	0.00000	0.1096	0.00020	0.01380	-0.17	7.61
stock	2772	0	-0.2997	-0.00050	0.6259	0.00130	0.03650	1.41	42.58

Table 2: Summary statistics for our returns data in the estimation window.

In Figure 4 four Q-Q plots are shown. The upper part is for index returns and the lower part for stock returns. On the y-axis we have the theoretical quantiles and on the y-axis we have the sample quantiles. Theoretical distribution in the Q-Q plot is the normal distribution (left part) as well as the t distribution (right part). For both index returns and stock returns we can see that a t distribution better fits the return data than a normal distribution does. It is a better fit because the points lie closer to the line—for a perfect fit all points would lie on the line since then the theoretical quantiles would exactly match the sample quantiles.

This is important to check because in the GARCH-model on page 6 the white noise term z_t is assumed to follow some distribution (for example normal or std) and as is evident from the formula the assumed distribution of z_t is connected with the distribution of returns. Our data clearly shows that returns are closer to follow a t distribution than a normal, hence these plots suggest that it is more reasonable to assume $z_t \sim t(\nu)$ than to assume $z_t \sim N(0, 1)$. Of course the shape parameter ν needs to be estimated using MLE which increases the number of parameters, but given the huge difference in the Q-Q plots between normal and t distribution in the estimation window it should be better to use a t distribution. Our MLE of the shape parameter for from data in the estimation window is $\hat{\nu}_1 = i.df = 2.63$ and $\hat{\nu}_2 = s.df = 2.28$ for index returns and stock returns respectively.

None of the return series, however, show a perfect fit. Especially the tails deviate from the t distribution. For probabilities less than 1 percent or greater than 99 percent (as indicated by the y-axis) our data do not fit a t distribution since the points are far away from the line. If one cares about a 1 percent VaR or 1 percent CVaR [2] then it would be more appropriate to use another distribution or perhaps model the tails using Extreme Value Theory. This thesis is concerned with 5 percent VaR so the departure from normality at the 1 percent level it is not a major issue. [3]

[2] CVaR is a measure similar to VaR but concerns the expected tail loss and is calculated by integrating the return distribution from $-\infty$ to VaR.

[3] It would, though, be an issue for this thesis to use a normal distribution; if we see look on the plots to the left for index returns (the top left plot) and focus closely on the y-axis at the 5 percent level we can see that the points the points lie to the left of the line. By doing the same exercise for stock returns (the bottom left plot) we see that also there the return depart from normality at the 5

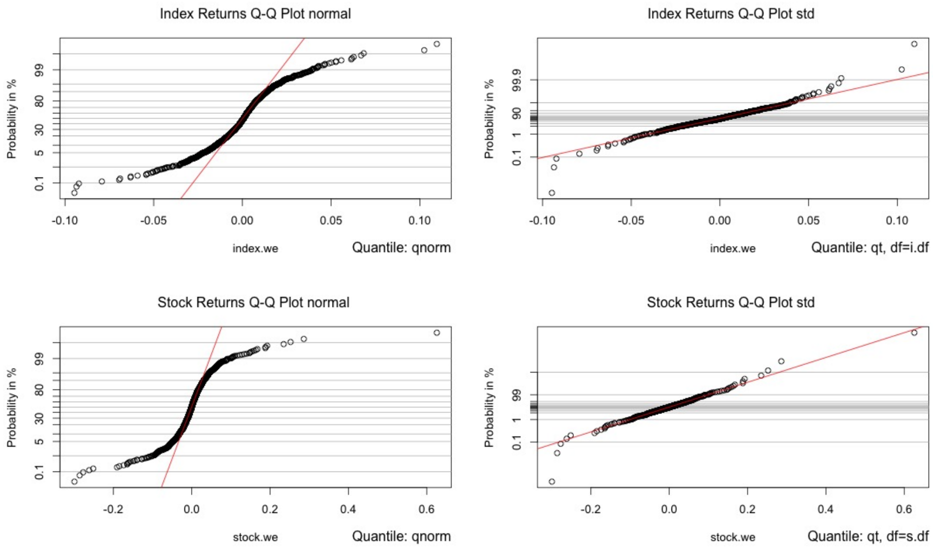

Figure 4: Q-Q plots of index returns (above) and stock returns (below) for two different distributional assumptions. A t distribution (right) is clearly more appropriate than a normal distribution (left).

4.1.4 Mean Equation and ACF of Returns

In Figure 5 we show four ACF plots. The upper part is for index returns and the lower part for stock returns. On the y-axis we have the value of γ_k and on the x-axis we have the number of lags k. The left part is an ACF plot of the returns. The right part is an ACF plot of the squared returns. An ACF plot is used to see if our data is serially uncorrelated. If the γ_k lies within the dotted lines for many values of k then data is serially uncorrelated. A formal statistical tests may be done—in this case the Ljung-Box test—and we use it together with the ACF plots.

By looking at the left part of Figure 5 we see that returns are correlated according to the ACF plots. But a t-test show that they are not statistically significantly different from zero. Moreover, the autocorrelation of returns are weak in comparison to the autocorrelation of squared returns as can be seen by at the y-axis of a plot to the left compared with the x-axis of a plot to the right ($0.05 < 0.3$). Even though returns appear to be correlated according to ACF, we will not model the mean equation μ_t with an ARMA model but instead use a constant mean $\mu_.$ in our models.

percent level. This is important since VaR is nothing but a quantile—see equation (14)—so in order to get the correct VaR estimates we should definitely assume a t distribution for our innovations.

4.1.5 Volatility Equation and ACF of Squared Returns

In the right part of Figure 5 we see that r_t^2 are correlated according to the ACF plots. A t-test show that r_t^2 are significantly different from zero since the p-value is close to zero. So we need to model the volatility equation.

Note that the reason for looking at the ACF of squared return to determine whether there exists ARCH effects is that the true variance of r_t is $\text{var}(r_t \mid \mathcal{F}_{t-1}) = E[(r_t - E[r_t])^2 \mid \mathcal{F}_{t-1}] = E[\varepsilon_t^2 \mid \mathcal{F}_{t-1}]$.

The Ljung Box-test's p-value is close to zero, so squared returns are correlated. Conclusively, according to formal statistical tests and ACF plots an ARCH-type model is appropriate for our data.

4.1.6 Conclusion of EDA and Model Discussion

We have established that there exists ARCH effects in the data. When choosing an appropriate model we need to decide if we will use an ARCH or GARCH model, how many parameters the model should have, and what distribution we should assume for the white noise term z_t.

Should we use an ARCH or a GARCH model to capture the conditional heteroskedasticity? We would need over $m = 10$ lags in an ARCH-model. But we don't want to include so many lagged squared returns, therefore we use a GARCH model instead.

How many parameters should we use? We will use a GARCH(1,1) model because it's the most parsimonious GARCH model there is. Of course some alternatives are (2,1) and (1,2) or even (2,2) but that may be subject to another thesis. Besides being the most parsimonious—making the model more interpretable for us—for financial data the GARCH(1,1) model is the most used, so we are not alone in choosing this model for the volatility equation.

What distribution should be assumed for the white noise term? According to Figure 4 using a t distribution is a better fit than a normal distribution for both returns series. Since the assumed distribution of z_t impacts the assumed distribution of r_t via $r_t = \mu_t + \sigma_t z_t$ and we have seen that a t distribution is the better choice for r_t we will assume a t distribution for the white noise term z_t.

Conclusively, we will use a GARCH(1,1)-t model for both index returns and stock returns.

4.2 Fitting the GARCH model

In this subsection we fit a GARCH(1,1)-t model to both return series in the estimation window. The parameter estimates as well as the information criteria for the models are tabulated. We also plot the mean forecasts and volatility forecasts generated by the GARCH model.

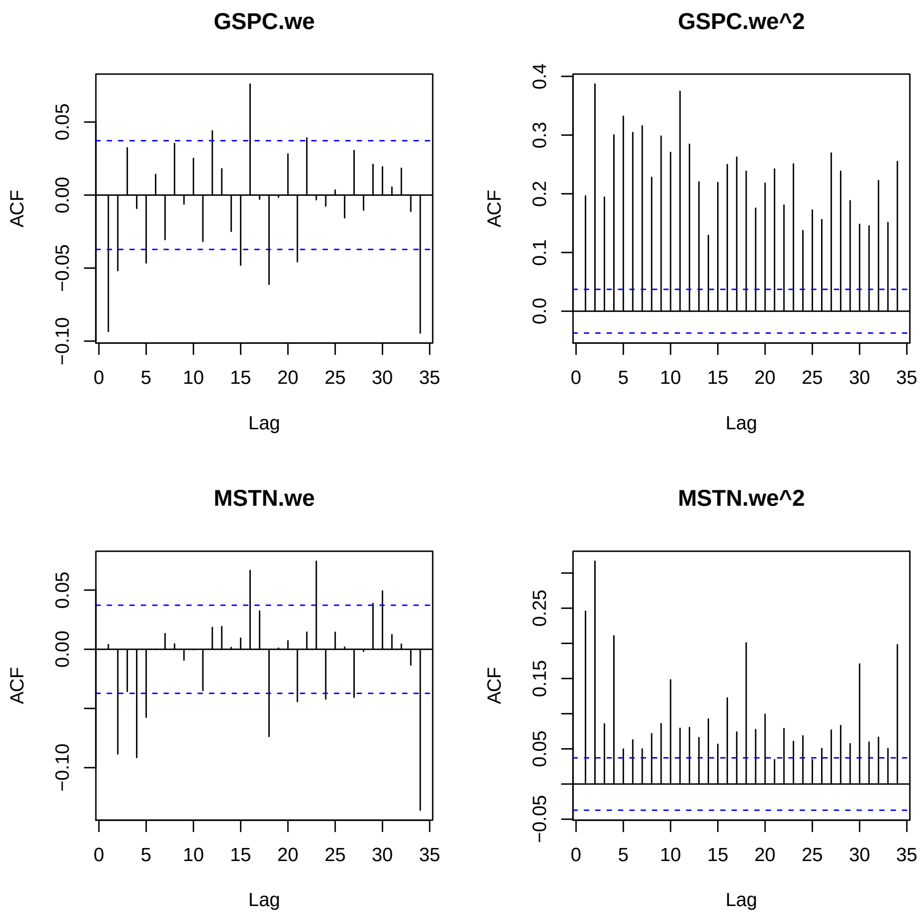

Figure 5: ACF plots Squared returns (right) and returns (left) for GSPC (upper) and the bank stock MSTN (lower). Squared return have high ACF so the volatility equation need to be modeled. Only data in estimation window have been used.

4.2.1 Estimated Coefficients

In Table 3 we have labeled the model fitted to index returns as *m.index* and the model fitted to stock returns as *m.stock*. We use data in the estimation window to fit models to our return series. The purpose of using only the estimation window, and not the entire sample is that we want to reserve the forecasting window to test our model.

	m.index	m.stock
Unc. Mean	0.0009	0.0002
Unc. Variance	0.0002	0.0017
alpha1	0.0721	0.0679
beta1	0.9212	0.9300
Persistence	0.9933	0.9979
Halflife	103	334
Likelihood	8433	6430
AIC	-6.007	-4.635
BIC	-5.997	-4.625

Table 3: Coefficient Comparison.

In Table 3 the most important values are $alpha1, beta1, Persistence$. The $Persistence$ is a bit higher in *m.stock* than in *m.index* but they are very close, i.e. $\hat{\alpha}_1+\hat{\beta}_1 \approx \hat{\alpha}_1+\hat{\beta}_1$, so let's study these in detail. For both assets, $beta1$ is close to 1 and $alpha1$ close to zero. This is to be expected with financial data. Recall that β_1 is the coefficient in front of σ_{t-1} so a high beta means that volatility is persistent in the sense that yesterday's volatility greatly impacts today's volatility. Recall that α_1 is the coefficient in front of ε_{t-1} so a high alpha means that volatility is spiky since a shock yesterday (caused by an unusually high or low return yesterday) affects the volatility of today.

The $beta1$ is higher in *m.stock* than in *m.index*, and the interpretation of this is that the volatility is more persistent for *m.stock* than for *m.index*.

The $alpha1$ is higher in *m.index* than in *m.stock*. The interpretation of this is that the volatility is more spiky for *m.index* than *m.stock* (however both α_1 coefficients are insignificant as seen by their high p-values in so we should be cautious in drawing conclusions about the spikiness).

In Table 3 the AIC and BIC have higher absolute values for *m.index* than for *m.stock* which suggest that in the estimation window the fitted model is more appropriate (w.r.t AIC and BIC measures) for *m.index* than *m.stock*.

In Table 3 we see that or the unconditional mean μ we have $\mu^{m.index} > \mu^{m.stock}$ which is also reported in Table 4 on the first row.

m.index	Estimate	Std. Error	t value	Pr(>|t|)
mu	0.0009	0.00	4.77	0.000
omega	0.0000	0.00	0.39	0.700
alpha1	0.0721	0.05	1.37	0.169
beta1	0.9212	0.05	17.31	0.000
shape	8.1801	1.97	4.15	0.000
m.stock	Estimate	Std. Error	t value	Pr(>|t|)
mu	0.0002	0.00	0.79	0.432
omega	0.0000	0.00	0.36	0.718
alpha1	0.0679	0.05	1.34	0.179
beta1	0.9300	0.05	18.41	0.000
shape	6.0121	0.93	6.44	0.000

Table 4: Coefficients with robust standard errors.

In Table 4 we display the maximum likelihood estimates of the GARCH coefficients, together with the Std.Error of the estimates as well as their p-values. Understanding how these coefficients impact the voltility forecast is important, so we highlight some of the key numbers in a bullet list.

- The *omega* have a high p-value for both assets, which is common for financial data. The fact that *omega* is not different from zero means that the unconditional long run variance is zero and this was seen in Table 3.
- The *beta*1 is significant as is seen by the low p-value. So it is not only large, as we already knew from Table 3, but also significant. This points toward the conclusion that yesterday's volatility does in fact impact today's volatility.
- The estimate of *alpha*1 is, on the other hand, not significant. The large Std. Error result in a p-value so high that we cannot unfortunately reject the null $alpha_1 = 0$ so the value of *alpha*1 need to be taken with a grain of salt.
- The *shape* parameter is the estimated degrees of freedom in the t distribution of our innovation ε. We denote these shape paremeters ν_1 for *m.index* and ν_2 for *m.stock*. Their estimated values are seen in the table and $\nu_1 > \nu_2$ but only slightly. If the values are far enough from each other one can see a difference when plotting the PDF, but these values of $\hat{\nu}_1$ and $\hat{\nu}_2$ are so close that a density plot would look the same.

The coefficients are not the same in our estimation window, nor are the standard errors and thereby the p-values. During volatile periods the standard errors increase since we get less certain about our estimates. This is one reason why models are not as suitable during periods of high volatility. Therefore, during a crisis, we might trust our models less.

4.3 Asses Model Fit and Assumptions

For the model assumptions to be fulfilled we must perform some diagnostical checks on the standardized residuals. We would like to conclude that the standardized residuals

$$\tilde{\epsilon}_t = \frac{\hat{\epsilon}_t}{\hat{\sigma}_t}$$

to follow a white noise process. For details see equation (7) on page 6. Before we can say that we believe $\tilde{\epsilon}_t$ follows a white noise process we want to test some assumptions.

Assumption 1: There are no autocorrelation in the standardized residuals. This is the case if the the ACF of standardized residuals for most lags k lies within the dotted lines in an ACF plot for most values k or it can be tested using a Ljung-Box test.

Assumption 2: There are no ARCH effect in the standardized residuals. This is tested the same way as above but with squared standardized residuals instead. [4].

Assumption 3: The assumed distribution for the innovations is close to the empirical distribution of the standardized residuals. This is the case if the Q-Q Plot looks like a straight line.

4.3.1 Test of First and Second Assumption

In Figure 6 we do an ACF plot to test the assumption 1 and 2. The left part is for standardized residuals and is used to test assumption 1. The right part is for the squared standardized residuals and is used to test assumption 2. The upper part of the for *m.index* and the lower part for *m.stock*.

Assumption 1 is fulfilled for both *m.index* and *m.stock* since the value of ACF lies within the dotted lines for all lags. The plot is supported by a Ljung-Box test with p-values seen in the plot.

Assumption 2 is considered to be fulfilled for both *m.index* and *m.stock* although this is not as clear since the value of ACF sometimes is higher than the dotted line indicating that there is some ARCH effect left meaning that our GARCH model did not pick up all of the conditional heteroskedasticity that existed in the returns data. The p-values of the Ljung-Box test are high for both *m.index* and *m.stock* indicating the same thing as we can see i the plot, namely that we accept the null although we are not as certain that assumption 2 is fulfilled as we are with assumption 1.

To summarize, we see that the Ljung-Box test complement the ACF plot and support us in our conclusion that assumption 1 and 2 are fulfilled.

[4] This is by same argument as described in the EDA section when we argued that ARCH effects are found by plotting ACF of Squared Return

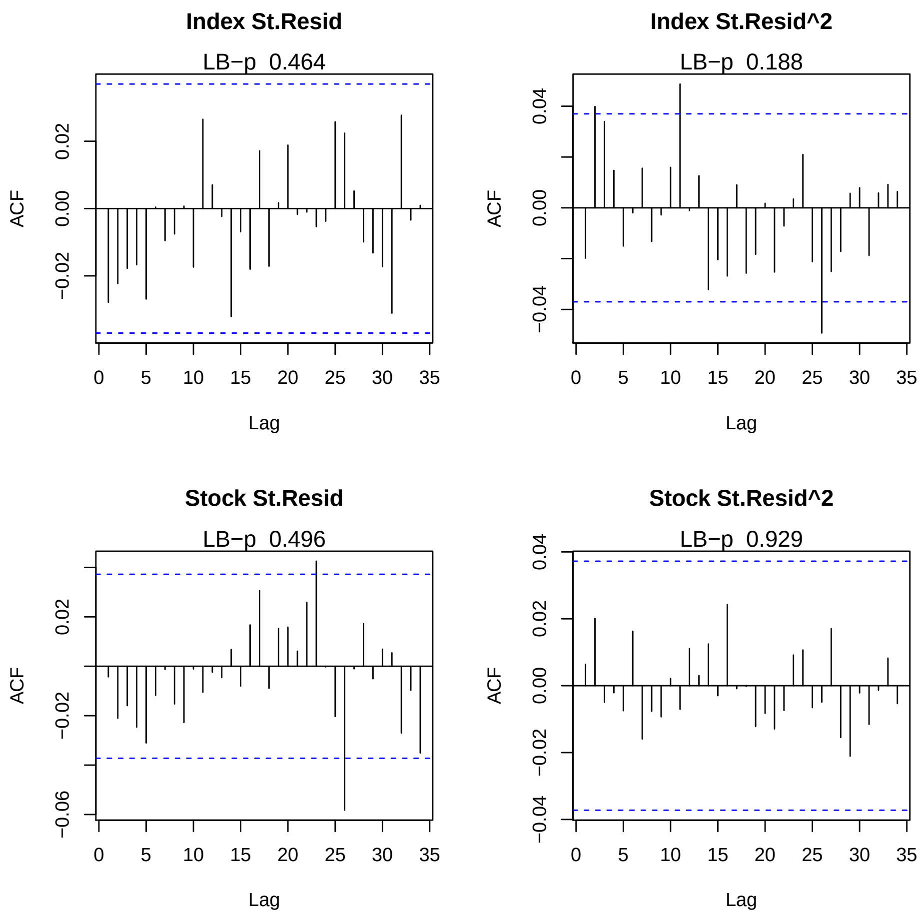

Figure 6: ACF of Standardized residuals (left) and Standardized residuals squared (right) for *m.index* (upper) and *m.stock* (lower). Ljung-Box p-values are shown in each plot. Conclusion: Assumption 1 and 2 are fulfilled.

4.3.2 Test of the Third Assumption

In Figure 7 we test assumption 3 using a Q-Q plot with the theoretical quantile on the y-axis and the empirical quantiles on the x-axis, *m.index* is in the upper part and *m.stock* in the lower part. Normal distribution is to the left and t distribution is to the right. Firstly and most importantly the plots to the right indicate that our standardized residuals seem to follow a t distribution hence our assumption that $\tilde{\varepsilon} \sim t(\nu)$ is supported by data. Thus assumption 3 is fulfilled. Secondly, and not as important, we note that *m.index* is closer to the line than *m.stock*—so *m.index* fulfills assumption 3 better than *m.stock* mainly due to the outlier. Thirdly, as a side note, we can compare the left part to the right part and conclude that the t is a better fit hence it was good to assume t distribution.

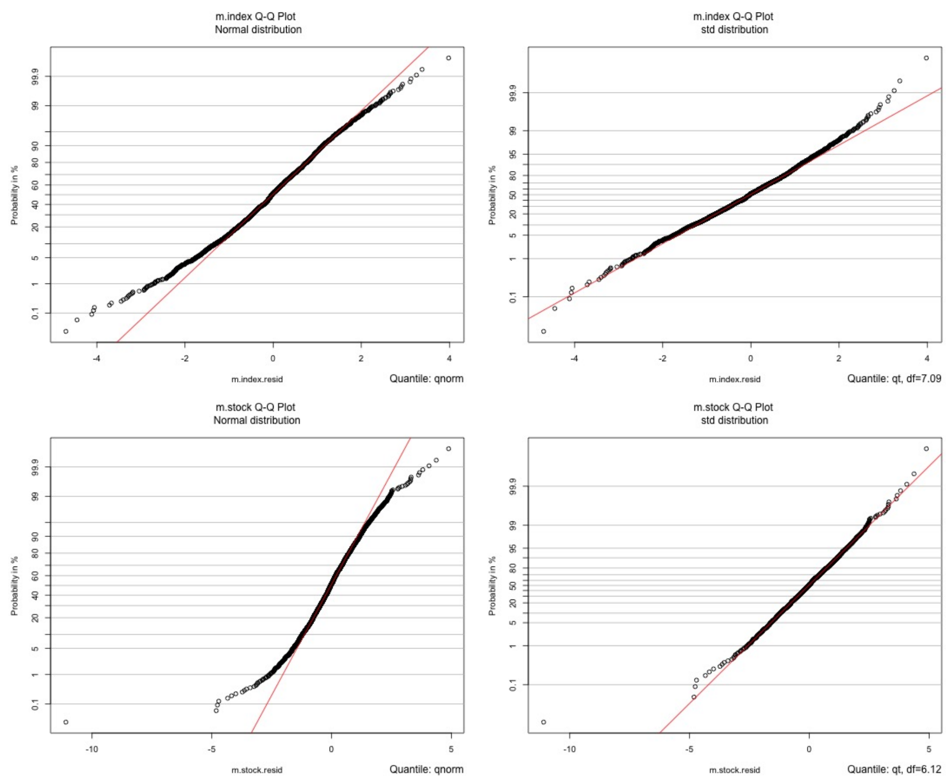

Figure 7: Q-Q plot for *m.index* (upper) and *m.stock* (lower) using theoretical distributions normal (left) and t (right). Conclusion: Assumption 3 is fulfilled.

4.3.3 Conclusion of Diagnostical Tests

No model is perfect, but based on these Diagnostics checks of the residual assumption we would argue that the GARCH(1,1)-t model is appropriate.

5 Results

In this section we use the two models discussed in the previous section to forecast and backtest VaR. In the first subsection one-day ahead VaR forecasts are evaluated with standard backtesting procedures, and the results are summarized in a table. In the second subsection we plot the VaR forecast for both models. In the third subsection we calculate the RME. In the last subsection we conclude that the GARCH(1,1)-t model is better for a stock than for an index.

5.1 Backtest Comparison

The size of the forecasting window is $n(w_F) = 1000$ and $p = 0.05$ so the expected number of violations is

$$expected = p \cdot n(w_F) = 50$$

but the actual violations for a model is given by

$$actual = \sum_{t=1}^{n(w_F)} I_t$$

where I_t is defined in equation (16) on page 13. The calculation of VaR_t is defined in equation (15). If $actual < expected$ the model is said to overestimate risk. If $actual > expected$ the model is said to underestimate risk. Overestimating the risk implies that the bank has unnecessary capital reserves (which is bad for the society because capital is used ineffectively). Underestimating the risk is also bad for a bank because it leads to punishment's from regulators (and it is bad for society because historically bank's have been bailed out at tax payer's cost). A good model will have $actual \approx expected$.

	expected	actual	cc.p	uc.p
m.index	50	60	0.3620	0.1589
m.stock	50	47	0.4710	0.6603

Table 5: VaR-testing with uc-test and cc-test. Higher p-value is better. Column uc.p is the most important.

The uc test essentially tells us if the sum of I_t is correct. The cc test essentially tells us if violations are clustered at the same time as the number of violations are correct. For more details of these test see section 3.3 on page 13.

The calculation of $(VaR_1, ..., VaR_{w_F})$ and the violations $(I_1, ..., I_{w_F})$ are thoroughly described in section 2.3.7. The results from the cc test and uc test are found in Table 5. A key take away from Table 5 is that both models are successful, our GARCH(1,1)-t is appropriate for both return series—especially so for the stock.

We wish for failure of rejecting the null. Given the formulation of the null hypothesis of each test we wish to "accept" the null hypothesis (or more formally "fail to reject"). We pick a significance level of 5 percent for our test, and do note that this percentage has nothing to do with the 5 percent of VaR. Both our models accept the null at a significance level of 5 percent, although *m.stock* is close to failing the cc test since the p-value is close to 5 percent. If both models fail to reject the null hypothesis we may decide a winner by comparing the p-value. The higher the p-value the better the model. With this idea, *m.stock* is better than *m.index* accoridng to VaR backtesting because of the the higher p-value in Table 5.

5.2 VaR Forecast Plot

In Figure 8 and 9 we plot the forecasted value at risk VaR_t together with the realized return rr_t. The first figure is for *m.index* and second is for for *m.stock*. The returns are more volatile for the stock. The VaR forecast varies over time, and when returns are more volatile the VaR forecast increases. It does so because the sigma forecast increase because a high ε_{t-1} leads to a high $\hat{\sigma}_t$. This plot is important for understanding that the violations occur whenever the realized return (grey) is lower than the VaR forecast (black) and this is the basis of our backtests.

5.3 RME Comparison

The Root Mean Error (RME) is a performance measure. In this thesis, it can be interpreted as a measure of how snugly the estimated VaR line is to the returns. A low RME measure will be seen in Figure 8 and 9 as having very little space between the realized return and the VaR line. RME matters because a high RME means a financial institution has a lot of money in its reserves. The lower the RME the better the model. [5]

As seen from the formula for RME on page 11 we use the VaR forecasts and the realized returns to compute RME value. The fraction $RME^{m.index}/RME^{m.stock} \approx 0.40$ is small so $RME^{m.index} < RME^{m.stock}$. According to the RME measure *m.index* is much better than *m.stock*. In financial terms this means that *m.index* have lower capital charge than *m.stock* does. This result can be understood from looking Figure 8 and 9, since the stock have a larger range of y-values making the VaR line fit less tightly to the realized returns.

[5] If person A put away 50 USD and person B put away 60 USD as a reserve and everything goes well then person A will have had 10 USD more to invest hence you might argue that person B have wasted 10 USD. This argument is not entirely accurate but it serves as a description of why RME matters.

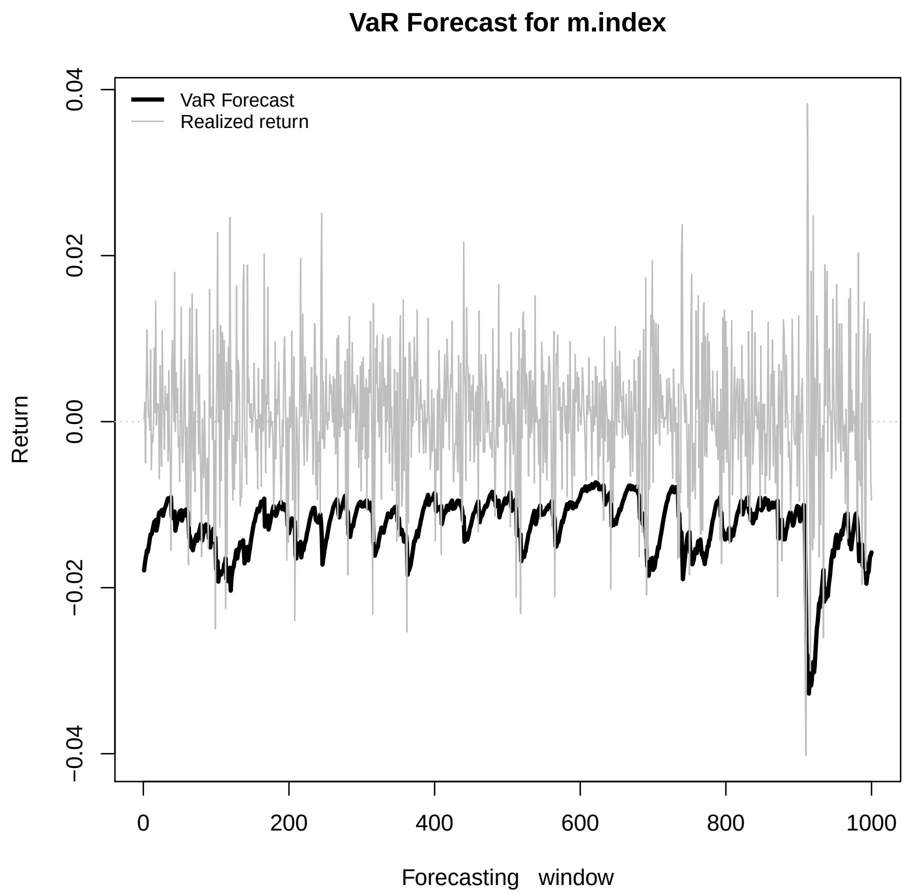

Figure 8: VaR Forecast Plot for $m.index$. The forecasted VaR is a black line and the realized returns are gray.

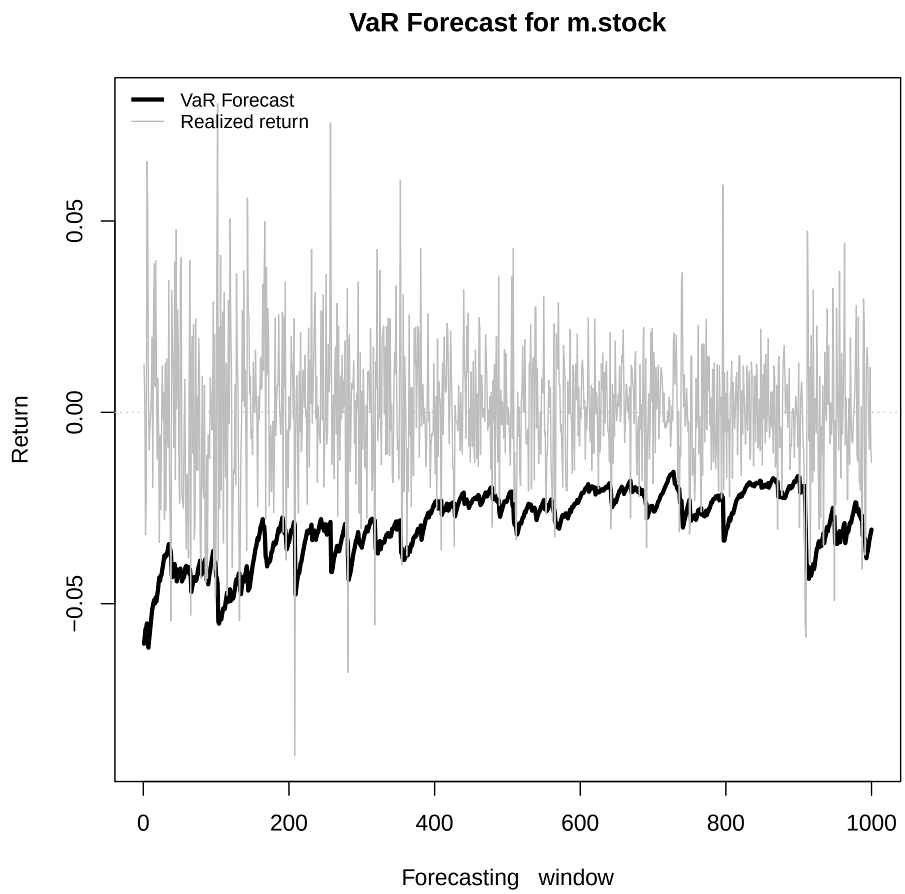

Figure 9: VaR Forecast Plot for *m.stock*. The forecasted VaR is a black line and the realized returns are gray.

5.4 Summary and Discussion of Results

Firstly, the p-value from the cc test suggests violations are less clustered in $m.stock$ than in $m.index$. Secondly, the actual violations are closer to the expected for $m.stock$. However, RME is lower for $m.index$, due to the larger range for stock returns as seen by the y-axis.

Is it possible to explain why? No, it is impossible to say for sure why this and it's very hard to pin down an exact reason for our results. Both the academic and the practioner will have to survive with the results that this simply *is* and we cannot know exactly *why*.

In fact it would, ex ante, be reasonable to believe that $m.index$ would be better according to VaR backtesting, becuase of two reasons: (a) volatility is higher for a single stock making it harder to predict (b) market news can hit a stock worse than a diversified portfolio.

(a) The input to the GARCH model is returns data. As seen in Figure 3 when comparing the index returns plot and the stock returns plot we observe that stock returns are more volatile. Moreover, stock specific events such as earnings and corporate actions like mergers, product releases should have an impact on the stock price, but none of these factors are incorporated in the GARCH model. Indices and portfolios can, since they average the returns, smooth out these stock specific events.

(b) In the GARCH framework bad news means that $\varepsilon_t < 0$. Bad economical news such as an abrupt change in FX rates hits a single stock harder than it hits a large portfolio. This is especially true for sector-specific news. To the investor holding a portfolio, bad news for one set of assets is good news for another set of assets e.g. a decrease in the EURUSD rate is good for exporter/importer but bad for the importer/exporter.

Nonetheless, VaR is a more suited risk management tool for our stock than for our index portfolio, at least according to the backtesting procedures used in this thesis. One possible explanation is that the tests are not very good; a lot of information is thrown away when we apply an uc-test or a cc-test. Critique has been raised towards these tests, and papers from Berkowitz as well as Hong and Li suggest ways of utilizing more information to gain statistical power. It will not be discussed further or used in this thesis, yet, if other statistical it is possible that we would have ended up with the opposite results namely that VaR is a more appropriate risk measure to the index.

Lastly, we should repeat that both the stock and the index passed the backtesting procedures. Hence, a financial regulator would deem the GARCH model used in this thesis to appropriately forecast risk for both the stock and the index.

6 Conclusions

In this section we draw conclusions from the results in the previous section. The first sentence states the main findings and answers our research question. The following text describe how this result was reached. A detailed discussion of our results are on the previous page.

The two main findings are that

1. the risk measure Value at Risk—where volatility is calculated with a standard GARCH(1,1)-t model—is a more suitable risk management tool for our stock than for our index portfolio, and that

2. according to cc test and uc test our GARCH(1,1)-t model adequately captures the risk for both the stock and the index.

Given the purpose of our thesis described in section 1.1 we hope these findings answer to what extent the risk models can be applied for a single asset as opposed to a portfolio. To be more precise what we mean by "suitable", we need to briefly describe how the index was compared to the stock. We had two different datasets: returns on a stock index and on a bank's stock. We used the same GARCH model for both return series, and labelled them *m.index* and *m.stock* respectively. To forecast the volatility we used a standard GARCH(1,1) model with t distributed innovations. Via the formula $VaR_t = \hat{\mu}_t - t_{0.95}(\hat{\nu})\hat{\sigma}_t$ the volatility forecast was used in the calculation of the $VaR_{0.95}$ The VaR forecast from these models is found in Figures 8 and 9. The VaR forecasts can be backtested using standard backtesting procedures, in this thesis the most important one is Christoffersen's conditional coverage test. The results from applying this test to our VaR forecasts is found on page 27 in Table 5 which showed that *m.stock* was better than *m.index* with respect to VaR backtesting.

7 Further research

In this section we suggest what can be done in order to built upon the research done in this thesis. The suggestions can be seen as potential subjects for another B.Sc. or M.Sc. thesis.

In this thesis $VaR_{0.95}$ i.e. $p = 5$ % have been used. In practice financial institutions often need to have $p = 1$ % or even $p = 0.1$ % so another thesis might consider a lower the value of p.

A standard GARCH(1,1) model have been used. Asymmetric GARCH models handle a negative innovation $\varepsilon_t < 0$ different from a positive innovation $\varepsilon_t > 0$ so that negative news increase the volatility forecast more than positive news does. Asymmetric GARCH models often perform better than symmetric ones. Examples are E-, GJR-, AV- and NAGARCH. It would be interesting to replace our standard GARCH with an Asymmetric one, such as EGARCH, and produce the same analysis.

Even though it is standard procedure to use two parameters (1,1) as we have done in this thesis, maybe considering ε_{t-2} and using a GARCH(2,1) would give extra forecasting power.

A t distribution have been assumed for the innovation. This choice was made both with EDA and with simplicity in mind. There are, however, numerous distribution that can be used such as: skew t, Normal Inverse Gaussian, Johnson's Su distribution etc.

We have used an equity index and a stock. There are many different asset classes such as fixed income, currencies, commodities and real estate. A financially interesting question is whether the results from this thesis is generalizable to other asset classes.

A lot of information is thrown away when applying the uc test and cc test. Other tests such as Berkowitz test or Hong and Li test have more statistical power and are more accurate forecasting measures.

References

Articles

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

Bollerslev, T., 1987, A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return, *The Review of Economics and Statistics* 69(3), 542-547.

Christoffersen, P. F., 1998, Evaluating Interval Forecasts, *International Economic Review* 39(4), 841-862.

Dickey, D. A., Fuller, W. A., 1979, Distribution of the Estimators for Autoregressive Time Series with a Unit Root, *Journal of the American Statistical Association* 74(366a), 427-431.

Engle, R. F., 1982, Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation, *Econometrica* 50(4), 987-1008.

Giot, P. and Laurent, S., 2004, Modelling daily Value-at-Risk using realized volatility and ARCH type models, *Journal of Empirical Finance* 11(3), 379-398.

Kuester, K. and Mittnik, S. and Paolella, M. S., 2006, Value-at-Risk Prediction: A Comparison of Alternative Strategies, *Journal of Financial Econometrics* 4(1), 53-89.

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

Ljung, G. M. and Box, G. E. P., 1978, On a measure of lack of fit in time series models, *Biometrika* 65(2), 297-303.

Mandelbrot, B., 1963, The Variation of Certain Speculative Prices, *The Journal of Business* 36(4), 394-419.

Books

Brooks, C., 2002, *Introductory Econometrics for Finance*, The Press Syndicate of the University of Cambridge, Cambridge, United Kingdom.

Hull, J. C., 2012, *Options, Futures, and Other Derivatives*, 8th ed., Pearson Education Limited, Essex, United Kingdom.

Jorion, P., 2007, *Value at Risk: The New Benchmark for Managing Financial Risk*, 3rd ed., The McGraw-Hill Companies, Inc., Two Penn Plaza, New York.

Tsay, R. S., 2010, *Analysis of Financial Time Series*, 3rd ed., John Wiley & Sons, Inc., Hoboken, New Jersey.

Tsay, R. S., 2013, *An Introduction to Analysis of Financial Data With R*, John Wiley & Sons, Inc., Hoboken, New Jersey.

Electronic resources

Ghalanos, A., 2015, *Introduction to the rugarch package.*, The Comprehensive R Archive Network, URL: `https://cran.r-project.org/web/packages/rugarch/vignettes/Introduction_to_the_rugarch_package.pdf` (accessed 2016-05-15).

Longerstaey, J. and Spencer, M., 1996, *RiskMetrics—Technical Document*, Morgan Guaranty Trust Company and Reuters Ltd, New York, URL: `http://pascal.iseg.utl.pt/~aafonso/eif/rm/TD4ePt_2.pdf` (accessed 2016-05-11).