UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Value at Risk Estimation using GARCH Family Models:
A Comparison of Different Specifications and Distributions

Khaled Jrideh

Bachelor's thesis in Statistics, 15 credits
Spring 2023
Supervisor: Mattias Sundén

Department of Economics and Statistics
School of Business, Economics and Law
University of Gothenburg

Abstract:

The objective of this study is to compare the performance of different GARCH models, under different conditional distribution assumptions, to predict one-day-ahead Value-at-Risk (VaR) for three stocks: Swedbank, Handelsbanken, and SEB over the Covid-19 period. The performance is evaluated using Kupiec, Christoffersen tests and the Quadratic Loss. The results show that the assumed distribution, model specification, and confidence level together play an important role in VaR estimation, as these factors can substantially affect the accuracy of the estimate. Models that assume the skew t-distribution and the t-distribution generally perform well, while the performance of models that assume the normal distribution changes dramatically as a function of the confidence level. Regarding the models, the study shows that the EGARCH specification resulted in the lowest losses and that the worst performing model is ARCH, especially when the assumed distribution is normal.

Contents

1 Introduction

1.1 Background

Risk is the possibility that something negative will occur in the future. In finance, it refers to the probability of loss or uncertainty in financial markets or investments. Major events such as the Asian financial crisis in 1997, the global financial crisis in 2008, and the recent Covid-19 pandemic highlight the importance of effective risk management in the financial industry. According to McNeil et al. (2005, p. 4), financial institutions need to manage different types of risk to maintain their stability and viability. These include credit risk, which is the risk of not receiving the promised repayment of investments due to the default of a borrower; operational risk, which is the risk of loss due to inadequate internal processes, people, systems, or external events; and market risk, which is the risk of potential loss due to changes in the value of the underlying assets. In this thesis we focus on market risk, specifically one of its risk measures. A popular method to help financial actors manage their market risk is Value-at-Risk (VaR), first introduced by JP Morgan in the early 1990s. Later, the publication of the G30 report and other reports such as that of the Basel Committee on Banking Supervision highlighted the usefulness of the VaR system and its potential (Dowd 2005, p. 10). There are different definitions of VaR depending on how the calculation is performed. According to the definition of Xekalaki and Degiannakis (2010, p. 239), VaR is the predicted amount of financial loss of a portfolio over a given time horizon and at a given probability level. According to Danielsson (2011, p. 93), there are two main methods for predicting VaR: the nonparametric and the parametric. The nonparametric method, also called historical simulation, is a simple, straightforward method that uses the empirical distribution of an asset's return to calculate VaR. This method does not require statistical model assumptions or parameter estimates. In the parametric method that we will use in this thesis, we need to specify the statistical distribution of an asset's return. Then we can predict the VaR by simply estimating the volatility for the assumed distribution.

Volatility is a term commonly used in the financial market that refers to the dispersion of an asset's return. That is, the fluctuation in the price of an asset. In statistics, volatility is the standard deviation. Accurate prediction of volatility is essential for calculating reliable VaR measures. For this purpose, we will use the GARCH (Generalized Autoregressive Conditional Heteroscedasticity) model, as it is a widely used method for estimating volatility. The GARCH model, developed by Bollerslev (1986), is particularly effective in capturing the phenomenon of volatility clustering, which refers to the tendency for large price movements to be followed by other large price movements. This implies that volatility is not constant but fluctuates over time, and there is strong empirical evidence for the existence of volatility clusters (Slim et al., 2017; Cont, 2001). Other factors can also affect the accuracy of VaR measures. For example, the choice of the underlying statistical distribution of an asset's return can substantially affect VaR estimates because VaR is, by definition, the quantile of that distribution, i.e., the threshold below which a given percentage of returns is expected to fall. For example, if a portfolio has a VaR of $1 million at a 99% confidence level, the probability of losing less than $1 million is one percent over a given time horizon. Therefore, the 1% quantile in this example is the VaR. If we assume a normal distribution, the returns are symmetric and the model is unable to capture extreme values because the probability of observing values that are 3 standard deviations away from the mean is highly unlikely. Therefore, the model may not be able to capture extreme shocks if we assume a normal distribution (Slim et al., 2017). On the other hand, there are other distributions that can account for skewness and excess kurtosis, such as the t-distribution and skew t-distribution, which can be particularly important when working with financial data because it accounts for the presence of "fat tails" and skewness in the return distribution, meaning that negative extreme values are more likely to occur than assumed with the normal distribution. However, it can also be difficult to determine the exact shape of the tails of the return distribution (Cont, 2001). Several studies have applied different statistical distributions to financial data and have reached similar conclusions about the distribution. For example, Bao et al. (2007), Köeksal and Orhan (2012), and Angelidis et al. (2004) found in their studies that distributions that account for skewness and leptokurtosis perform better than the normal distribution. Another factor that determines the accuracy of VaR is the choice of GARCH model, as many models have been developed for different stylized statistical properties, such as those mentioned above. For example, Hansen and Lunde (2005) found that more complex GARCH models do not necessarily perform better than simpler models. In contrast, Oberholzer and Venter (2005) found that more complex model specifications performed better during periods of high volatility. However, accurately predicting volatility can be difficult due to the complexity of financial markets.

There are also other factors that must be considered, such as the time period chosen to estimate the parameters. For example, if a period with high volatility is used to estimate the parameters, this can lead to overfitting and inaccurate predictions for new observations. Another factor is the length of the window used to fit the model: A short window may cause the parameters to be more sensitive to change, while a long window may have the opposite effect.

In this study, we will look at different GARCH models and examine how they perform in combination with different conditional distributions in a period of high volatility. We will then compare the results with the most commonly used methods based on backtesting measures according to Xekalaki and Degiannakis (2010, p. 246), namely Kupiec and Christoffersen tests to evaluate Value-at-Risk (VaR).

1.2 Purpose

The objective of this study is to compare the performance of four generalized autoregressive conditional heteroscedasticity (GARCH) models for predicting one-day-ahead Value-at-Risk (VaR) for three stocks: Swedbank, Handelsbanken, and SEB. The four conditional volatility structures are ARCH (1), GARCH(1,1), APARCH(1,1) and EGARCH(1,1). The performance of the models will be evaluated at both 99% and 95% VaR levels using the Kupiec, Christoffersen tests and the Quadratic Loss. The study will also examine the impact of three different distributions on the performance of the GARCH models: Normal distribution, t-distribution, and skew t-distribution.

2 Theory

This chapter presents the theoretical foundations and methodology used in this study. The first section explains the notation of the various GARCH models. Section 2.2 provides an overview of Value-at-Risk (VaR), followed by a discussion of the probability density functions used. Section 2.4 presents the Kupiec and Christoffersen tests, two important statistical tools for evaluating the performance of VaR models. The final section outlines the forecasting procedure and describes the out-of-sample forecasting method used.

2.1 Conditional Heteroscedastic Models

Many models used to forecast volatility belong to the GARCH family. The first such model, introduced by Engle (1982), is the autoregressive conditional heteroscedasticity (ARCH) model, which allows the conditional variance to vary over time based on squared past returns. A modified version, the generalized ARCH (GARCH) model proposed by Bollerslev (1986), includes lagged volatility and is the most commonly used volatility model. However, both the GARCH and ARCH models have the weakness that they assume that positive and negative shocks have the same impact on volatility (Tsay 2010, p.119). To address these weaknesses, Nelson (1991) introduced the exponential GARCH (EGARCH) model, which allows positive and negative shocks to have different effects on variance. To determine the basic structure for the ARCH framework, we first define the daily return as:

$$Y_t = 100(\ln P_t - \ln P_{t-1})$$

Where Y_t is the percentage log return and P_t is the closing price of an asset. According to Xekalaki and Degiannakis (2010, p.11), the dependent variable Y_t can be decomposed into two parts, the predictable component μ_t and the unpredictable component ε_t, thus:

$$Y_t = \mu_t + \varepsilon_t$$

$$\mu_t = \mu(\theta|I_{t-1})$$

$$\varepsilon_t = \sigma_t Z_t$$

$$\sigma_t = g(\theta|I_{t-1})$$

Here $\mu(\theta|I_{t-1})$ and $g(\theta|I_{t-1})$ are the functional forms of the conditional mean and conditional variance, where by conditional is meant that, μ_t and σ_t depend on the information set available up to time I_{t-1} and depend on the parameter vector θ. This is in contrast to the unconditional variant, where we simply use the entire sample. The conditional mean $\mu(\theta|I_{t-1})$ can be estimated by an autoregressive moving average (ARMA), where the return can be expressed as a function of the conditional variance. Consistent with the conclusion of Angelidis et al. (2004), that the mean process does not add anythings significant to the models except the complexity in the estimation, we assume in this study that the conditional mean is zero. The return on day t then becomes:

$$Y_t = \sigma_t Z_t$$

Where Z_t is i.i.d with mean 0 and variance 1 and σ_t is the standard deviation, can be estimated using conditional variance models. The parameters of the models are estimated using the so-called Maximum Likelihood method (ML). The ML is typically used to estimate the GARCH parameters and aims to find the parameters that maximise the likelihood that the model will produce the observed data, (see A.3 in the Appendix)

2.1.1 ARCH

The ARCH specification is giving as follow:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i Y_{t-i}^2$$

Where q is the number of lags that is included to estimate the variance. The α_i parameters explain how fast the conditional variance at time t reacts to news. The ARCH(1) given as:

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2$$

For the conditional variance to be positive, the parameters should satisfy $a_0 > 0$ and $a_i \geq 0$ for $i = 1, ..., q$.

2.1.2 GARCH

According to Danielsson (2011, p.38) the problem with ARCH is the long of lag length that is required to estimate the volatility. Despite that ARCH is simple, it often needed many parameters to define the volatility (Tsay, 2010, p.119). The GARCH model is basically the same as ARCH, but with the inclusion of lagged volatility. GARCH model's specification:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i Y_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$$

Where $\alpha_0 > 0, \alpha_i \geq 0$ for $i = 1, ..., q$ and $\beta_j \geq 0$ for $j = 1, ..., p$.
The most popular GARCH model is GARCH(1,1):

$$\sigma_t^2 = \alpha_0 + \alpha_1 Y_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

A high value of β_j means that the volatility takes long time to change and a high value of α_i means that volatility reacts fast to market movements (Dowd, 2005).

2.1.3 EGARCH

According to Xekalaki and Degiannakis (2010, p.13), Fischer Black was the first to observe that changes in asset's returns are often negatively correlated with changes in the volatility of returns. Volatility tends to increase on negative news and decrease on positive news, a phenomenon known as the leverage effect. As a result, it can be assumed that a negative return today will lead to higher volatility tomorrow. The GARCH model is effective at capturing fat-tailed return and volatility clusters, but it has a significant limitation: it uses the square of today's return to estimate tomorrow's volatility, without considering the sign of the return. To address this issue, Nelson (1991) introduced the EGARCH model, which captures the effects of both negative and positive shocks on variance, and accounts for the asymmetry of volatility ,the fact that negative shocks have a greater impact on volatility than positive shocks. The EGARCH model is expressed in Xekalaki and Degiannakis (2010, p.43) as follows:

$$log(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{q} g\left(\frac{\varepsilon_{t-i}}{\sigma_{t-i}}\right) + \sum_{j=1}^{p} \beta_j \; log \; \sigma_{t-j}^2$$

$$g(\varepsilon_t/\sigma_t) = \alpha_1(\varepsilon_t/\sigma_t) + \gamma_1(|\varepsilon_t/\sigma_t| - E|\varepsilon_t/\sigma_t|)$$

The function g is defined as a combination of two terms: the product of the parameter α_1 and the ratio of the error term to the variance, and the product of the parameter γ_1 and the difference between the absolute residuals and the expected value of the absolute residuals.

The terms $\alpha_1(\varepsilon_t/\sigma_t)$ and $\gamma_1(|\varepsilon_t/\sigma_t| - E|\varepsilon_t/\sigma_t|)$ represent the sign and the magnitude effect, respectively. Also, the term $\alpha_1(\varepsilon_t/\sigma_t)$ stands for the leverage effect. When $\alpha_1 = 0$, it means that bad news and good news have the same effect on volatility, and when $\alpha_1 < 0$ it means that bad news causes higher volatility than positive news, so the model is asymmetric. On the other hand, γ_1 is a parameter that captures the effect of the absolute value of the residuals on the variance.

Note that $\varepsilon_t = \sigma_t Z_t$ and $Z_t \sim i.i.d.$ with expectation 0 and variance 1, thus $Z_t = \varepsilon_t/\sigma_t$. Unlike the GARCH model, no inequality restrictions need to be imposed for model estimation, as the logarithm of the variance is estimated.

2.1.4 APARCH

The APARCH (Asymmetric Power ARCH) model was introduced by Ding et al. (1993) as a way to combine the power model proposed by Taylor (1986) with the leverage effect proposed by Nelson (1991). The APARCH model is defined by the following equation:

$$\sigma_t^\delta = \alpha_0 + \sum_{i=1}^{q} \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{j=1}^{p} b_j \sigma_{t-j}^\delta,$$

where

$$\alpha_o > 0, \delta > 0, b_j \geq 0, j = 1, ..., p, \alpha_i \geq 0$$

and

$$-1 < \gamma_i < 1, i = 1, ..., q$$

In this equation, σ_t^δ is the conditional standard deviation at time t, and ε_{t-i} is the residual at time $t - i$. The terms α_i and γ_i capture the magnitude and the sign effects, respectively, while δ is the size of the power transformation. To understand how the sign effect works lets consider the asymmetric ARCH(1) model. We have:

$$\sigma_t^2 = \begin{cases} \alpha_0 + \alpha_1(1 - \gamma_1)\varepsilon_{t-1} & \text{if } \varepsilon_{t-1} \geq 0, \\ \alpha_0 + \alpha_1(1 + \gamma_1)\varepsilon_{t-1} & \text{if } \varepsilon_{t-1} \leq 0 \end{cases}$$

Hence, we see that the effect on the current volatility is higher when the ε_{t-1} is negative and $\gamma_1 > 0$ than if the ε_{t-1} was positive. In other words, bad news has a higher effect on volatility than good news and the constraint $\gamma_i \geq 0$ make sure to capture the asymmetric property in financial series. The APARCH model uses an asymmetric absolute value instead of the squared residual, which allows it to take into account the absolute return rather than just the squared return. According to Danielsson (2011, p.52), the autocorrelation function (see section 4.1) of absolute return and return square sometimes shows that absolute return has stronger autocorrelations than squared return. This is why the APARCH model may be more effective at estimating volatility than the GARCH model, which only uses the past squared return. The APARCH model includes seven ARCH models as special cases, including the GARCH model (obtained by setting $\gamma_i = 0$ and $\delta = 2$).

2.2 Value-at-Risk

VaR has become a popular risk measure in financial institutions due to its simplicity, as it reduces the risk associated with any portfolio to a single number. According to Xekalaki and Degiannakis (2010, p.240), for a given probability $(1-\alpha)$, VaR is the predicted amount of financial loss of a portfolio over a given time horizon. The mathematical definition of VaR can be expressed as follows:

$$Pr(Y_t < VaR(1-\alpha)) = \alpha$$

and

$$VaR(1-\alpha) = \mu - \sigma_t z_a$$

Here Y_t is the daily log return as defined in Section 2.1, and $(1-\alpha)$ is the confidence level. The μ is the mean return and we have assumed that it is equal to zero. The critical value for the assumed distribution is z_a for area α and σ_t is the conditional variance, which can be estimated using various GARCH models. If we assume that the return on an asset is normally distributed, the probability of a loss less than VaR(95%) = -1.645 is equal to $\alpha = 5\%$. Figure 1 shows a theoretical normal distribution, with the probability density function in blue. The red areas under the curve represent the α values at a VaR of 95% and 99%, indicated by the dashed red vertical lines.

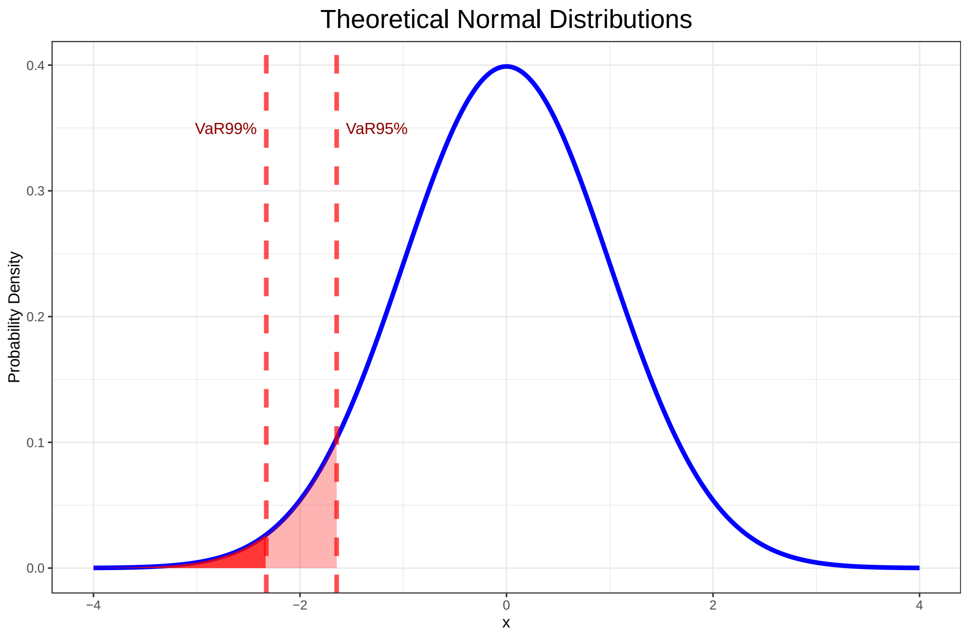

Figure 1: *Theoretical normal distributions with α values at a 95% and 99% VaR, which are indicated by the dashed red vertical lines.*

It is important for investors and financial institutions to estimate VaR accurately. If VaR is overestimated, investors put in more reserve than necessary and miss out on potential gains. On the other hand, if VaR is underestimated, the capital may not be enough to cover the risk. The simplest way to evaluate VaR is to count the number of losses that exceed VaR and compare them to the expected loss. If the difference is small, the VaR forecast is calculated correctly. According to Xekalaki and Degiannakis (2010, p.246), the methods proposed by Kupiec 1995 and Christoffersen 1998 is the most commonly used methods to evaluate Value-at-Risk (VaR).

2.3 Distributions

In this thesis, we will use three different distributional assumptions in the GARCH models to estimate VaR: the normal distribution, the t-distribution, and the skew t-distribution. By using these three distributions, we can compare the results and determine which distribution provides the most accurate VaR estimates.

2.3.1 Normal distribution

The normal distribution, also known as the Gaussian distribution, is a continuous probability distribution that is defined by a bell-shaped curve. The distribution is completely described by the mean μ and the variance σ^2. Figure 2 shows different normal distributions with different values for the mean and variance. The probability density function (PDF) of the normal distribution is given by the following equation:

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

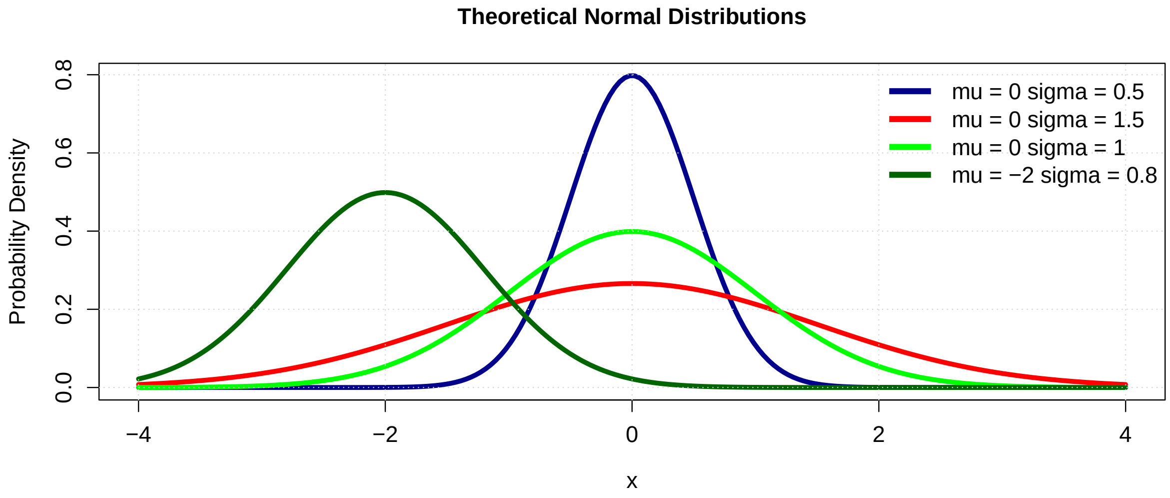

Figure 2: *Theoretical normal distributions with different means and standard deviations.*

2.3.2 t-distribution

The t-distribution, also called the Student's t-distribution as used by Bollerslev (1987), has a heavier tails than the normal distribution, which makes it more suitable for modelling financial data that may contain extreme values. In addition, the tail of the t-distribution can be adjusted by the degrees of freedom parameter. When the degrees of freedom approach infinity, the t-distribution becomes a normal distribution. By looking at Figure 3, we can see how the degrees of freedom affect the distribution. The density for t-distribution is given by:

$$f(x;v) = \frac{\Gamma(\frac{v+1}{2})}{\sqrt{v\pi}\Gamma(\frac{v}{2})}\left(1+\frac{x^2}{v}\right)^{-\frac{v+1}{2}}$$

The gamma function $\Gamma(.)$ is used to give the distribution a specific shape and behavior. The degrees of freedom v determines the shape of the distribution, and the gamma function helps to ensure that the distribution is properly normalized.

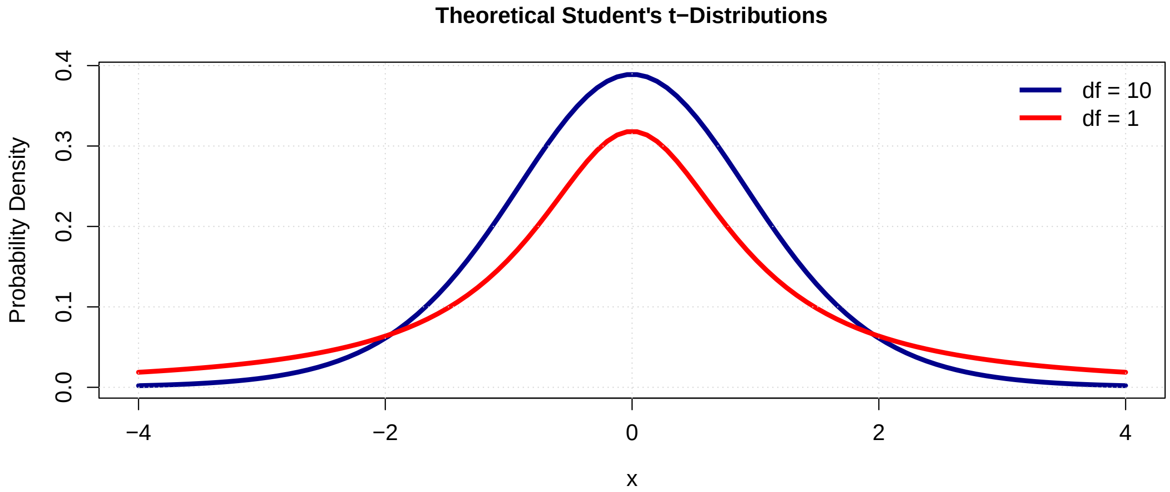

Figure 3: *Theoretical t-distributions with Different Degrees of Freedom.*

2.3.3 Skew t-distribution

The main difference between the t-distribution and the skew t-distribution is that the latter allows for the possibility of skewness in the distribution, while the standard t-distribution is a symmetric distribution. According to Fernandez and Steel (1998), the density function for the skew t-distribution is given by:

$$f(x;\gamma) = \frac{2}{\gamma + \frac{1}{\gamma}}\left\{ f\left(\frac{x}{\gamma}\right) I_{[0,\infty)^{(x)}} + f(\gamma x) I_{(-\infty,0)^{(x)}} \right\}$$

The skew t-distribution is a t-distribution scaled by the skewness parameter γ. The skewness parameter $\gamma \in (0, \infty)$ determines the degree of skewness of the distribution, where a value of 1 indicates a symmetric distribution and values greater than or less than 1 indicate a skewed distribution. The function $f(.)$ is the density function of the t-distribution as defined in the previous section and $I(.)$ is an indicator function. In Figure 4, we see that the distribution (in red) is symmetric when $\gamma = 1$, and that the distribution is right-skewed, i.e., it has a thicker tail to the right when $\gamma > 1$ and vice versa.

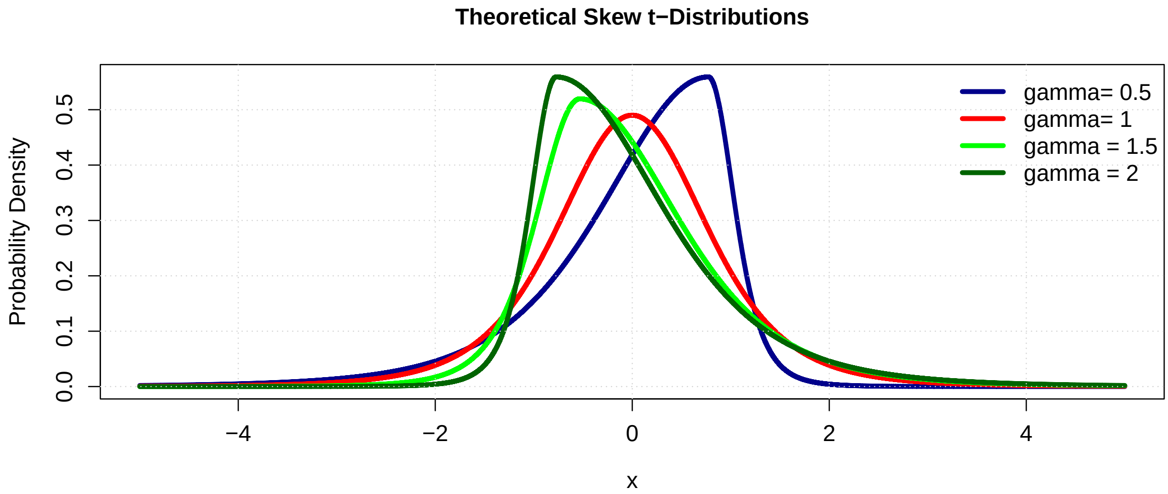

Figure 4: *Theoretical Skew t-distributions with different values for skewness parameter γ and fixed degrees of freedom.*

2.4 Backtesting

In this section, we will introduce the Kupiec and Christoffersen tests, two important statistical tools for evaluating the performance of VaR models. The Kupiec test compares the observed number of VaR exceedances with the expected number of exceedances based on the confidence level, while the Christoffersen test also considers the magnitude of the exceedances.

2.4.1 Kupiec Test

To evaluate the performance of VaR, we create an indicator variable:

$$I_t = \begin{cases} 1 & \text{if } Y_t < VaR_t \\ 0 & \text{otherwise} \end{cases}$$

Where $I_t = 1$ if the daily return exceeds VaR, and 0 otherwise.
Let $N = \sum_{t=1}^{T} I_t$ be the number of trading days in the given period T on which

the return exceeds VaR. Thus, the null hypothesis of the test is $H_0 : \frac{N}{T} = \alpha$. That is, the empirical quantity is equal to the nominal quantity. Accordingly, N under the null hypothesis follows a binomial distribution with parameters T and α, i.e., $N \sim binomial(T, a)$. Thus, according to Xekalaki and Degiannakis (2010, p. 245-247), the likelihood ratio under the null hypothesis is as follows:

$$LR_{un} = 2log\left((1-\frac{N}{T})^{T-N}(\frac{N}{T})^{N}\right) - 2log\left((1-\alpha)^{T-N}\alpha^{N}\right)$$

The likelihood ratio for unconditional coverage LR_{un} approximately follows the χ^2 distribution with one degree of freedom. This test yields zero under the null hypothesis and the more $\frac{N}{T}$ deviates from α, the larger the test becomes. The critical values of χ_2 for 10%, 5%, and 1% are 2.706, 3.841, and 6.635, respectively. If LR_{uc} is greater than 6.635, i.e., the null hypothesis is rejected at 1%, it means that the model specification is not appropriate to estimating VaR. Kupiec test is limited by two shortcomings. First, the test does not provide information on the magnitude of losses. This means that a 1% violation carries the same weight as a 5% violation. Second, Kupiec test only considers the frequency of losses and not the timing of their occurrence, i.e., it is not able to detect the independence of violations (Dowd, 2005).

2.4.2 Christoffersen Test

Christoffersen (1998) suggests a model that allows us to test the independence of VaR exceedances. In other words, the test checks whether the probability of observing a violation today is independent of observing a violation yesterday. The null hypothesis for this test is that the probability of observing a VaR violation today, given that a VaR violation occurred yesterday, is equal to the probability of observing a VaR violation today. This null hypothesis can be formulated as follows:

$$\pi_{ij} = P(I_t = j | I_{t-1} = i) = P(I_t = j), \qquad i, j = 0, 1$$

That is , whether the violation is independent or not. Thus, the null hypothesis is $H_0 : \pi_{01} = \pi_{11} = p$ where π_{01} is the probability of observing a VaR violation today given that there was a non-VaR violation yesterday, π_{11} is the probability of observing a VaR violation today given that there was a violation yesterday, and p is the desired percentage of failures. The indicator variable $\{I_t\}_{t=1}^{T}$ is defined in the same way as in the previous section. The test is a likelihood ratio test with χ^2 distribution and 1 degree of freedom. Thus, the likelihood ratio for independence LR_{ind} under the null hypothesis according to Xekalaki and Degiannakis (2010, p. 247) is obtained as:

$$LR_{ind} = 2\Bigg(log((1-\hat{\pi}_{01})^{n_{00}}\ \hat{\pi}_{01}^{n_{01}}(1-\hat{\pi}_{11})^{n_{10}}\ \hat{\pi}_{11}^{n_{11}})$$

$$-log\left(\left(1-\frac{N}{T}\right)^{n_{00}+n_{10}}\left(\frac{N}{T}\right)^{n_{01}+n_{11}}\right)\Bigg)$$

Where $\hat{\pi}_{ij} = n_{ij}/\sum_j n_{ij}$ is the estimated probability for state j on any given day, given that state i occurred on the previous day, and n_{ij} is the number of days that j occurred after state i occurred on the previous day. For example, if n_{01} is the number of days that state 1 occurred after state 0 occurred the day before, and $n_{00} + n_{01}$ is the total number of days that state 0 occurred the day before, then $\hat{\pi}_{01} = n_{01}/(n_{00} + n_{01})$. Similarly, $\hat{\pi}_{11} = n_{11}/(n_{10} + n_{11})$. In addition, Christoffersen (1998) introduces a modified version that can be used to simultaneously check whether the percentage of failures is equal to the desired value and whether the violations are independently distributed, i.e..:

$$H_0 : p^* = p \text{ and } \pi_{01} = \pi_{11} = p,$$

where p^* is the true percentage of failures and p is the desired percentage. The alternative hypothesis is:

$$H_1 : H_0 : p^* \neq p \text{ or } \pi_{01} \neq \pi_{11} \neq p,$$

The likelihood ratio test with χ^2 is now with 2 degrees of freedom. The likelihood ratio for the conditional coverage LR_{cc}, under the null hypothesis according to Xekalaki and Degiannakis (2010, p. 247), is given as:

$$LR_{cc} = -2log\left((1-p)^{T-N}p^N\right) + 2log\left((1-\hat{\pi}_{01})^{n_{00}}\hat{\pi}_{01}^{n_{01}}(1-\hat{\pi}_{11})^{n_{10}}\hat{\pi}_{11}^{n_{11}}\right)$$

where N is the number of violations and T is the total number of observations The conditional coverage test addresses the Kupiec test's shortcoming of not being able to detect independence of violations, which we discussed earlier. The conditional coverage test rejects a model that produces either too many or too few clustered violations. The critical value for the χ^2 distribution with 2 degrees of freedom for a significance level of 10%, 5%, and 1% is 4.61, 5.99, and 9.21, respectively.

2.5 Forecast Procedure

The out-of-sample forecasting method is commonly used to evaluate the predictive power of a model. In this method, data are divided into an in-sample period, an out-of-sample period, and the data used to estimate model parameters are different from the data used to evaluate model performance. In this study, we used the rolling-window approach with fixed length to implement the out-of-sample forecast. In this approach, the estimation window is fixed and is moved forward by one day at a time. In our study, the in-sample period consists of the first 1000 observations, while the remaining 759 observations constitute the out-of-sample period. To clarify, let n be the total size of the sample, see Figure 5, and w is the length of the rolling window. The first window consists of the first observation up to w and is used to train the model and forecast one-day-ahead h for day $w + 1$. The next day, the window moves by one day, excluding the first observation and including the second observation up to $w + 1$. This process continues until all 759 values have been forecasted. We then compare the predicted values with the actual observations using a specific tester.

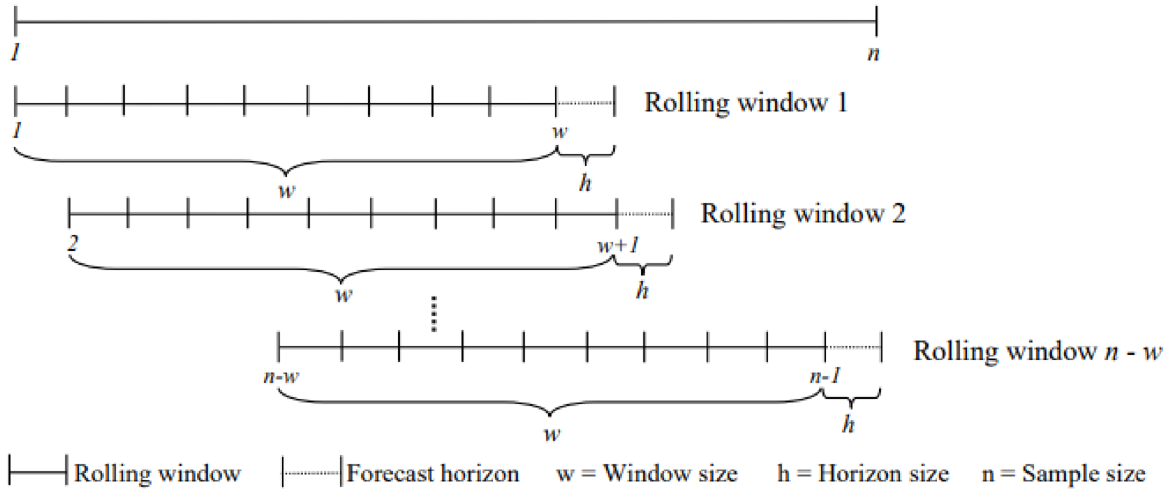

Figure 5: *Illustrating the Rolling window approach.; Source: Titov, C. (2022)*

Determining the optimal rolling window's length can be a challenge. While large rolling windows can improve the generalizability of the model and reduce its sensitivity to small changes in the data distribution, they can also reduce the flexibility of the model and slow its response to new observations. Previous research has reached conflicting conclusions about the optimal window length. Brownlees, Engle, and Kelly (2011) found that the longest possible estimation window produced the best results, while Köksal and Orhan (2012) and McNiel and Frey (2000) used window lengths of 1000 observations. Given the high volatility in our out-of-sample period due to the Covid outbreak, we have opted for a rolling window of 500 observations. We will also readjust the model parameters every 10 days to maintain flexibility, as recommended by Brownlees, Engle, and Kelly (2011).

3 Previous studies

There are numerous studies on VaR in the literature. The reasons for this are the property of VaR to reduce the risk associated with an asset to a single number and, as Iorgulescu (2012) described it, that it is considered as a benchmark for risk measurement. Several studies have been conducted on this topic, using different time periods and models. The results of these studies reach different conclusions, but there are also some commonalities.

Köksal and Orhan (2012) compared a list of comprehensive GARCH models to estimate VaR in a period of high volatility. The authors used data for stock market indices from both emerging markets (Brazil and Turkey) and developed markets (Germany and the United States) over the period of the global financial crisis. The study applies different GARCH specifications to estimate one-day-ahead variance and calculate VaR. The results show that the specification ARCH, followed by GARCH(1,1), provides the most favorable results. In addition, the t-distribution performed slightly better than the normal distribution. The study also shows that the worst performance is the non-linear power GARCH. Orhan and Akin (2011) conducted a similar study. They compared the performance of VaR for different indices of the Istanbul Stock Exchange during the global crises. The authors concluded that GARCH with normal distribution gives the best results, in contrast to GARCH with t-distribution. It is important to note that the authors used only the unconditional Kupiec test to evaluate the performance.

Berkowitz and O'Brien (2002) in their study, evaluated the performance of VaR for 6 large banks in the USA. They compared the performance of the bank's models with a simple ARMA-GARCH model. The VaR prediction of this model outperformed the models provided by the banks. The author attributes this shortcoming to the legislation that banks must follow, which makes the internal model less flexible, making it more difficult to account for changes in volatility. This result is confirmed by the study of Iorgulescu (2012), based on a portfolio consisting of four of the most liquid stocks traded on the Bucharest Stock Exchange (BSE). The volatility of the portfolio was predicted using a constant conditional correlation model (GARCH-CCC) and a dynamic conditional correlation model (GARCH-DCC). To account for leptokutosis Iorgulescu (2012) used the following tools: the t-distribution, the generalized hyperbolic distribution (GH), extreme value theory (EVT), and the Cornish-Fischer approximation for quantiles (CF). The result showed that all VaR models easily passed the unconditional coverage test, but only four of the original 8 models passed the independence test at a significant level of 10%. According to the author, the result proves that volatility cluster is a real problem in determining VaR for both simple and complex models and that the time period used to estimate the models has a crucial impact on the performance of the models.

In a similar study, Hansen and Lunde (2005) compared a large number of volatility models, such as the Exponential GARCH (EGARCH) model, the Asymmetric Power GARCH (APGARCH) model, and the FIGARCH model, in terms of their ability to describe the volatility of financial data. To evaluate the models, the authors used a variety of statistical methods, including mean square error, mean absolute error, and also considered the out-of-sample forecasting performance of the models over different time horizons. The authors found that when evaluating exchange data, none of the models outperformed the GARCH(1,1) model. However, when evaluating the returns of IBM stock, the authors found that the GARCH(1,1) model was outperformed by other models that could accommodate a leverage effect.

4 Data

In this study, we analyze the daily percentage log returns of Swedbank (SWED-A. ST), Handelsbanken (SHB-A. ST), and SEB (SEB-A. ST) using data from Yahoo Finance for the period from January 1 2015 to December 30, 2021. The analysis was performed using R version 4.2.2.

4.1 Descriptive Statistics and Data Visualization

The results in Table 1 show that all stocks in the analysis have the same number of observations and their means are very close to zero. Skewness is a measure

Table 1: *Summary statistics for log return*

Stock	Obs	Mean	St.Deviation	Skewness	Kurtosis	Min	Max	JB
SWED-A.ST	1759	0.003%	1.67%	-0.56	12.81	-15%	9.8%	p<0.00
SHB-A.ST	1759	0.01%	1.60%	-0.70	6.96	-12%	8.8%	p<0.00
SEB-A.ST	1759	0.03%	1.66%	-1.12	10.69	-14%	8.7%	p<0.00

of the asymmetry of a distribution, with a value greater than 0 indicating right skewness and a value less than 0 indicating left skewness. The normal distribution has a skewness of zero. In this case, the skewness values in Table 1 indicate that the actual distribution for each stock is left skewed, meaning that the distribution has a long tail on the left and a short tail on the right. In other words, negative returns are expected to be more common than positive returns. This is one of the stylized facts about asset's returns mentioned earlier. It is worth noting that the fact of left skewness does not apply to the exchange rate, which tends to move up and down symmetrically Cont (2000).

The Kurtosis is a measure of the peakedness or flatness of a distribution, with a value of 3 for the normal distribution. The kurtosis values in Table 1 show that the actual distribution is leptokurtic, that is, it has a higher peak and thicker tails than the normal distribution, which means that most of the values in the distributions are centered around the mean. The Jarque-Bera test, which tests whether a sample comes from a normally distributed population, significantly rejects the null hypothesis of normality for each set in the analysis. This shows that the distributions of the three stock returns are not normally distributed. This is not surprising since one of the stylized properties of asset's returns is that its distribution is not Gaussian and has a thicker tails and a sharp peak, (Cont, 2001).

Figure 6 shows the percentage log returns of each stock over time. A notable feature of the data is the appearance of the leverage effect at the beginning of 2020, which is characterized by a downward trend in prices and an increase in volatility, likely due to the negative impact of the Covid 19 pandemic on the stock market. This result is consistent with the theory proposed by Black (1976), which states that volatility tends to increase in response to negative news

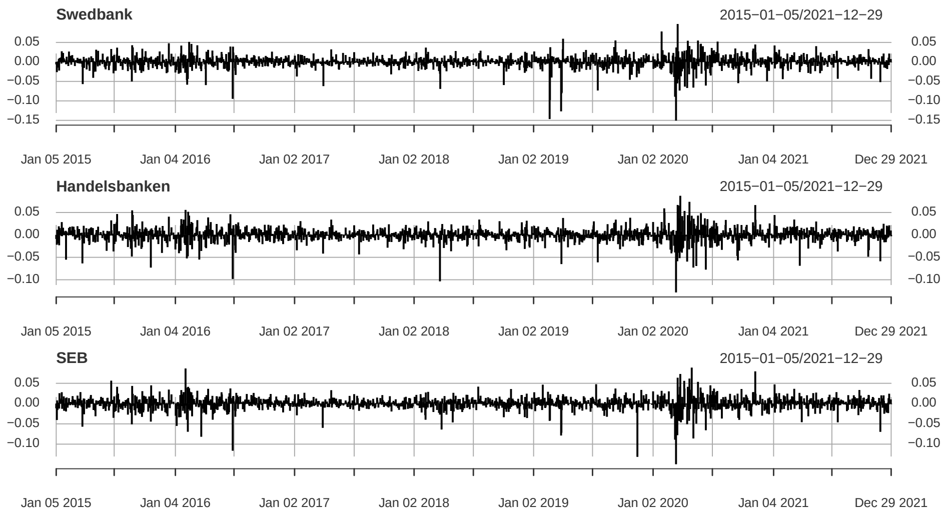

Figure 6: *Log returns from January 2015 to December 2021 for the three stocks*

and decrease in response to positive news. The returns in Figure 6 also show a clustering of volatility, i.e., the tendency for periods of large price changes to be followed by other large price changes and vice versa. This means that high levels of volatility tend to occur in clusters rather than being evenly distributed over the period.

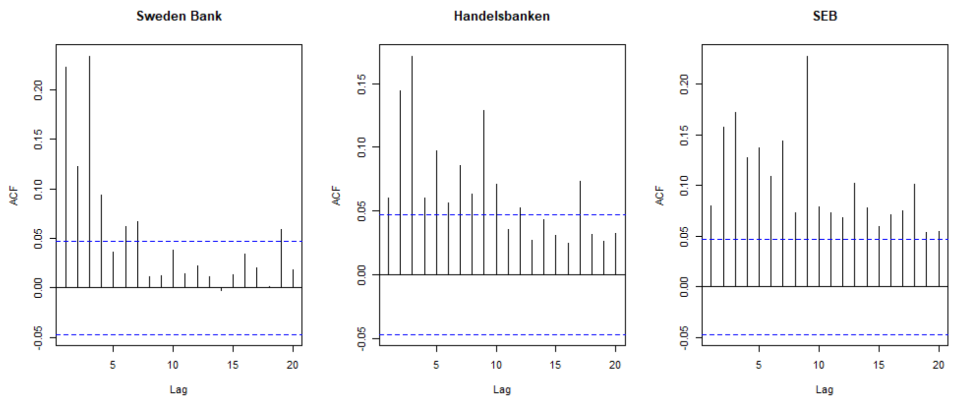

Figure 7: *Autocorrelation function (ACF) for squared return*

One way to test the suitability of using GARCH to predict VaR is to use the autocorrelation function (ACF) (see Section A.1 in the Appendix for the definition of ACF). Figure 7 shows the ACF for the squared return with a confidence interval of 95%. It can be seen that the correlations are significant even at long lags, as in the case of Handelsbanken and SEB, which is strong evidence of the predictability of volatility. Figure 8 shows the autocorrelation function for

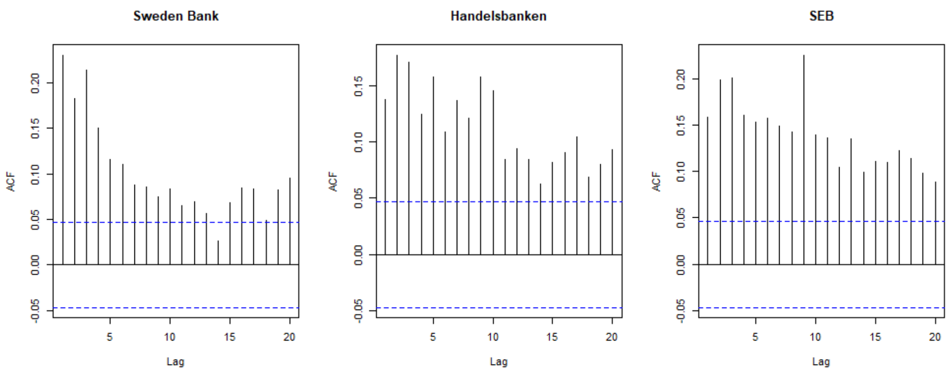

Figure 8: *Autocorrelation function (ACF) for absolute value of return*

absolute returns. The figure shows that the ARCH effect exists for all stocks, and the results of the Ljung-Box test also confirm this result (see Table 6 in the Appendix). As we can see, absolute returns have a stronger correlation than squared returns, which we discussed in Section 2.1.4.

To assess how well the distribution of daily stock returns fits the normal distribution, we use a Q-Q plot, a graphical tool that compares the empirical distribution of a data set to a reference distribution. In Figure 9, the black line represents the normal distribution and the plotted points show the empirical distribution of stock returns. It can be seen that stock returns deviate significantly from the normal distribution, as indicated by the deviation of the plotted points from the line. This is consistent with the results of the Jarque-Bera test in Table 1, which show that stock returns are not normally distributed.

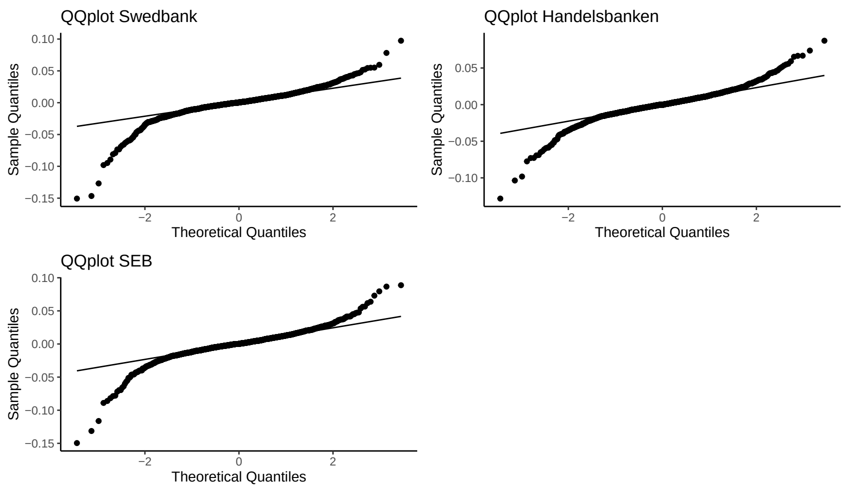

Figure 9: *Q-Q plot shows the observed distribution of our stock returns to the assumed normal distribution (straight line)*

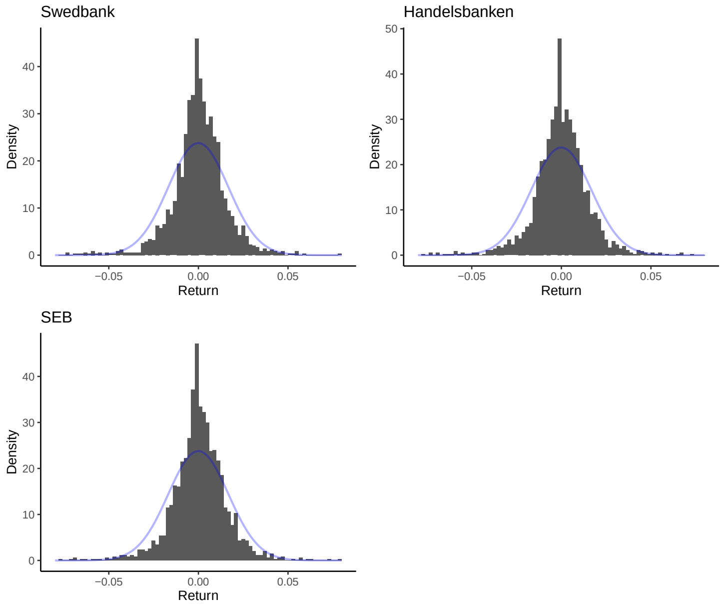

Figure 10: *Daily return histogram with a normal distribution density curve in blue.*

Figure 10 shows the histograms of daily returns along with the hypothesised normal density curves in blue. It is clear that the normal distribution does not accurately reflect the distribution of daily returns for all stocks, especially in the tails that are of particular interest in this study. This supports the conclusion that the normal distribution may not be an appropriate model for the distribution of stock returns.

5 Results

In this section we will use the Kupiec test to evaluate the performance of the models for unconditional convergence and the Christoffersen test for conditional convergence. The tests are performed at 99% and 95% VaR for each of the 12 models, separately for each stock. The significance level is set at 10% as recommended by Christoffersen, Iorgulescu (2012). This corresponds to a critical value of 2.7 and 4.6 for the Kupiec and Christoffersen tests, respectively. A ranking of the models based on the sum of statistical losses and a summary & discussion of the results follows. The cells shaded in green are those that were not rejected at a significance level of 10%, and the cells shaded in red are those with the worst performance.

5.1 Swedbank

In Table 2, the Kubiec and Christoffersen tests and the violation ratio are used to evaluate the performance of each model. The ideal violation ratio should correspond to the selected confidence level α, which in this case is 1% for 99% VaR and 5% for 95% VaR.

At the 99% VaR level, the specification ARCH with the skew t-distribution performed best, providing the lowest Kupice test statistic and the violation ratio was closest to the perfect ratio 1.2%, followed by APARCH and GARCH with the same distribution. On the other hand, the specification ARCH with a normal distribution performed the worst, having the highest violation ratio 4.9% and significantly rejecting the null hypothesis in both tests. Regarding the distribution, it is clear that the normal distribution performs the worst among the tested distributions, since the null hypothesis is significantly rejected in both tests for each specification that uses it, and we can see how the likelihood ratio for both tests gives a significantly higher value when the distribution is normal. On the other hand, the skew t-distribution gives the best result and the t-distribution is somewhere in between.

At the 95% VaR level, we see a different result. As for the distribution, the normal distribution surprisingly outperforms the other distributions and shows a significant result in every specification except ARCH, where it still has the worst performance with a violation ratio of 7.8%. On the other hand, the t-distribution gives the worst result and is rejected in every specification for both tests. The GARCH specification with normal distribution has the lowest violation ratio and provides the lowest Kupiec test statistic, followed by EGARCH and APARCH. The Skew t-distribution also has good performance here when used with a specification that accounts for the leverage effect, such as APARCH and EGARCH.

Table 2: *Kupiec and Christoffersen tests result - Swedbank*

	$\textbf{VaR}^{99\%}$			$\textbf{VaR}^{95\%}$		
Model	N/T	LR_{un}	LR_{cc}	N/T	LR_{un}	LR_{cc}
ARCH-N	4.9%	59.569	63.977	7.8%	10.58	13.11
ARCH-t	1.6%	2.22	4.01	7.0%	5.62	8.40
ARCH-skew-t	1.2%	0.25	0.46	6.3%	2.59	2.91
GARCH-N	2.6%	14.14	20.29	5.1%	0.03	1.79
GARCH-t	2.4%	10.41	17.76	7.2%	7.12	10.90
GARCH-skew-t	1.4%	1.36	3.48	7.1%	6.35	10.50
EGARCH-N	2.4%	10.41	11.01	5.4%	0.25	0.28
EGARCH-t	1.8%	4.38	4.90	7.4%	7.93	8.13
EGARCH-skew-t	1.6%	2.20	2.59	6.3%	2.59	2.58
APARCH-N	3.3%	25.19	29.00	5.7%	0.68	1.24
APARCH-t	1.8%	4.38	4.90	7.1%	6.35	6.35
APARCH-skew-t	1.4%	1.36	3.48	6.3%	2.34	2.59

* note: Green and red cells display the superior and poor performance respectively.

5.2 Handelsbanken

The results for Handelsbanken are shown in table 3. Again, we see that the performance of ARCH with normal distribution is the worst at both levels. The asymmetric power ARCH (APARCH) with the skew t-distribution gives the best result, followed by GARCH with skew t-distribution and t-distribution. Again, it can be seen that the performance of the normal distribution is poor at 99% VaR, as it is significantly rejected in every specification in both tests, while skew t-distribution performs well and the t-distribution is very close to it.

For the 95% VaR, all distributions, seem to perform equally well, as almost all specifications pass the tests. The null hypothesis for unconditional and conditional convergence could not be rejected for 95% VaR for all models except ARCH with normal distribution, as it showed the worst performance. The APARCH and EGARCH with normal distribution have the lowest violation ratio and provide the lowest Kupiec test statistic.

Table 3: *Kupiec and Christoffersen tests result - Handelsbanken*

	$\text{VaR}^{99\%}$			$\text{VaR}^{95\%}$		
Model	N/T	LR_{un}	LR_{cc}	N/T	LR_{un}	LR_{cc}
ARCH-N	7.4%	130.2	139.9	10.7%	39.3	49.6
ARCH-t	2.0%	5.69	6.29	6.6%	3.67	3.84
ARCH-skew-t	2.0%	5.69	6.296	6.3%	2.59	2.91
GARCH-N	2.9%	18.28	20.31	6.0%	1.69	4.96
GARCH-t	1.4%	1.36	1.68	5.7%	0.68	0.82
GARCH-skew-t	1.4%	1.36	1.68	5.7%	0.68	0.82
EGARCH-N	2.6%	14.14	15.23	5.3%	0.11	1.65
EGARCH-t	1.7%	3.21	3.66	6.3%	2.59	2.91
EGARCH-skew-t	1.6%	2.20	2.59	5.9%	1.30	1.96
APARCH-N	2.1%	7.14	7.83	5.2%	0.11	0.12
APARCH-t	1.4%	1.36	1.68	5.9%	1.30	1.51
APARCH-skew-t	1.3%	0.70	0.97	5.7%	0.68	0.77

* note: Green and red cells display the superior and poor performance respectively.

5.3 SEB

SEB Bank's results are shown in table 4. For a VaR of 99%, the specification ARCH with a normal distribution performed worst in both tests, with a violation ratio of 8.8%. However, the pattern for SEB looks different from the previous results, as the null hypothesis for unconditional and conditional convergence was rejected for all specifications at a significant level of 10%. Interestingly, the normal distribution still had the highest Kupiec and Christoffersen test statistic and the highest violation ratios. When looking at the VaR level of 95%, the models showed slightly better performance. The GARCH specification with normal and skew t-distributions had the lowest violation ratio, followed by EGARCH and ARCH specifications with skew t-distribution. On the other hand, the ARCH specification performed the worst with a violation ratio of 12%. Nevertheless, the normal distribution performed best among all model specifications, passing both tests for each specification. The skew t-distribution and the t-distribution have almost the same performance.

Table 4: *Kupiec and Christoffersen tests result - SEB*

	VaR$^{99\%}$			**VaR**$^{95\%}$		
Model	N/T	LR_{un}	LR_{cc}	N/T	LR_{un}	LR_{cc}
ARCH-N	8.8%	177.8	179.5	12.0%	57.1	58.1
ARCH-t	2.4%	10.4	11.1	5.4%	0.25	3.23
ARCH-skew-t	2.0%	5.69	6.786	5.9%	1.31	3.18
GARCH-N	2.8%	16.16	18.47	5.1%	0.03	3.68
GARCH-t	2.1%	7.14	11.30	6.2%	2.12	5.07
GARCH-skew-t	1.7%	3.21	8.96	5.7%	0.68	3.08
EGARCH-N	2.5%	12.22	15.18	5.9%	1.30	1.96
EGARCH-t	1.9%	5.69	10.33	7.2%	7.12	8.16
EGARCH-skew-t	1.7%	3.21	8.96	6.7%	4.28	6.15
APARCH-N	3.3%	25.19	32.36	6.1%	1.69	4.59
APARCH-t	1.9%	5.69	10.33	7.0%	5.62	6.10
APARCH-skew-t	1.8%	4.38	5.68	7.0%	5.62	6.10

* note: Green and red cells display the superior and poor performance respectively.

5.4 Quadratic Loss (QL)

One shortcoming of Christoffersen test is that, it does not take the size of the violation into account. That is, the distance between the observed return and the predicted VaR value. Therefore, we will use the Quadratic Loss (QL) that was developed by Lopez (1999) and was used by Köksal and Orhan (2012) and Angelidis et al. (2004), to measure the accuracy of VaR. The QL defined as:

$$QL_t = \begin{cases} 1 + (Y_t - VaR_t)^2 & \text{if } Y_t < VaR_t \\ 0 & \text{Otherwise} \end{cases}$$

A VaR model is penalized when a violation occurs. The model that results in the smallest total loss compared to the other models is the preferred model. For simplicity, we averaged the losses for the three stocks.

Table 5 shows the result for the QL. It is clear that the skew t-distribution and the t-distribution outperform the normal distribution at 99% VaR, since the first 7 specifications have one of these distributions. EGARCH has the lowest loss in predicting 99% VaR, followed by GARCH and APARCH with skew t-distribution. This is not surprising since these models performed best in the Kupiec and Christoffersen tests. The highest losses are for APARCH, ARCH and GARCH with normal distribution. For 95% VaR, the result is quite different EGARCH and GARCH with normal distribution perform best, followed by EGARCH with skew t-distribution, which is consistent with the results of Kupiec and Christoffersen tests. The specification ARCH with normal and t-distribution has the worst results. Looking at the 95% VaR, we can concluded that the distributions do not affect the performance of the models, since the result does not show a pattern like the 99% VaR.

Table 5: *Ranking the models by the sum of statistical losses*

Models	**VaR**$^{99\%}$	**VaR**$^{95\%}$
1	EGARCH-skew-t	EGARCH-N
2	GARCH-skew-t	GARCH-N
3	APARCH-skew-t	EGARCH-skew-t
4	EGARCH-t	GARCH-skew-t
5	GARCH-t	EGARCH-t
6	APARCH-t	GARCH-t
7	ARCH-skew-t	APARCH-skew-t
8	EGARCH-N	APGARCH-t
9	ARCH-t	APARCH-N
10	GARCH-N	ARCH-skew-t
11	ARCH-N	ARCH-t
12	APARCH-N	ARCH-N

5.5 Summary & Discussion

At the 99% VaR level, the GARCH specification with skew t-distribution performed best on average among the models tested, followed by the APARCH and EGARCH models with the same distribution. On the other hand, the ARCH model with a normal distribution performed the worst. However, when modeled with either the skew t-distribution or the t-distribution, acceptable results are obtained. This is in contrast to Köksal and Orhan (2012) who found that ARCH always performs best with both the normal and t-distributions. The EGARCH specification results in the lowest losses, which is consistent with Hansen and Lunde's (2005) finding that good out-of-sample performance requires a specification that can account for a leverage effect. In other words, the EGARCH specification can be considered a good model for predicting volatility in our out-of-sample period because it is able to give more weight to the negative shock caused by the Covid 19 outbreak.

As for the conditional distribution, the analysis shows that the normal distribution performs poorly on the 99% VaR. One possible reason for this is the presence of heavy tails in the distribution of the asset's return, which also explains the good performance of the t-distribution. On the other hand, the good performance of the skew t-distribution may be due to its ability to capture the asymmetry in the asset's return. These are some of the stylized facts discussed by Cont (2000).

At the 95% VaR level, the performance of the models is somewhat different. The normal distribution performs better than the other distributions, with a good result for every specification except ARCH, where it has the worst performance. In a similar study by Angelidis et al (2004), the authors come to the same conclusion. They also found that the normal distribution performed better on the 95% VaR than at a higher confidence level.

In summary, this study shows that the accuracy of VaR estimation can be good in some places and drastically wrong in others. Many factors such as model specifications, conditional distribution, and sample period (as shown in other studies such as Angelidis et al. (2004)) can affect the estimate. Therefore, we agree with Danielsson (2008) that the statistical model should not be considered a reliable factor for risk measurement because it often does not produce the desired results, especially for unpredictable events.

6 Conclusion

To predict one-day-ahead VaR the study found that the EGARCH specification is the most accurate estimate of Value-at-Risk (VaR), followed by the GARCH and APARCH models. On the other hand, ARCH shows the worst result. In terms of distribution, the normal distribution shows poor performance at the 99% VaR level, but better performance at the 95% VaR level, the skew t-distribution has good performance at both the 99% and 95% levels, followed by the t-distribution. The EGARCH specification resulted in the lowest losses, suggesting that it may be a good choice for modeling the leverage effect. In summary, it is important to consider the model specifications, confidence level, and conditional distribution when estimating VaR, as these factors can substantially affect the accuracy of the estimate. However, it should be noted that the reliability of VaR models for risk measurement should be questioned, especially in the case of extreme events.

7 References

Angelidis, T. Benos, A. & Degiannakis, S. (2004). The use of GARCH models in VaR estimation. Statistical Methodology. 1, 105-128 https://doi.org/10.1016/j.stamet

Bao, Y. Lee, T,. H. & Saltoğlu, B. (2007). Comparing density forecast models. Journal of Forecasting. 26, 203-255. https://doi.org/10.1002/for.1023

Berkowitz, J. & O'Brien, J. (2002). How Accurate Are Value-at-Risk Models at Commercial Banks?. The Journal of Finance. 57, 1093 - 1111 https://How-Accurate-Are-Value-at-Risk

Bollerslev, T (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics. 31, 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, 542–546. https://doi.org/10.2307..

Brownlees, C. & Engle, R. & Kelly, B. (2011). A Practical Guide to Volatility Forecasting Through Calm and Storm. Journal of Risk 14, 3–22. https://www.sas.upenn.edu

Cont, R. (2001). Empirical properties of asset returns: Stylized facts and statistical issues. Quantitative Finance. 1, 223–236. https://www-tandfonline-com.ezproxy.ub.gu.se/doi/epdf/

Danielsson, J. (2011). Financial Risk Forecasting. John Wiley & Sons Ltd

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

Engle, R., F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Journal of the Econometric Society. 50, 987-1007. https://doi.org/10.2307

Fernandez, C. & Steel, M, F. (1998). On Bayesian Modeling of Fat Tails and Skewness. Journal of the American Statistical Association. 93, 359-371. https://www.jstor.org

Hansen, P. R., & Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH(1,1)?. Journal of Applied Econometrics 20, 873–889. https://doi.org/10.1002/jae.800

Iorgulescu, F. (2012). Backtesting Value-at-Risk: Case study on the Romanian capital market. Procedia - Social and Behavioral Sciences. 62, 796–800. https://doi.org/10.1016/j.sbspro.2012.09.134

Laurent, S. (2003). Analytical derivates of the APARCH model. Computational Economics. 24, 51-57. https://link.springer.com/article

McNeil, A. J., & Frey, R. (2000). Estimation of tail-related risk measures for heteroscedastic financial time series, An extreme value approach. Journal of Empirical Finance. 7, 271–300. https://doi.org/10.1016

McNeil, A. J., Embrechts, P., & Frey, R. (2005). Quantitative risk management: Concepts, techniques and tools. Princeton University Press.

Nelson, B,. D. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. Econometrica. 59, 347-370. https://doi.org/10.2307/2938260

Oberholzer, N. & Venter, P. (2015). Univariate GARCH models applied to the JSE/FTSE stock indices. Procedia Economics and Finance. 24, 491 – 500 https://doi.org/10.1016/S2212-5671(15)00616-4

Orhan, M. & Akin, A. (2011). Performance of VaR in Assessing Risk: Evidence from the Istanbul Stock Exchange. Journal of Economic & Social Research. 13, 57-73. https://www.researchgate.net.

Orhan, M. & Köksal, B. (2012). A comparison of GARCH models for VaR estimation. Expert Systems with Applications. 39, 3582-3592. https://doi.org/

Slim, S. Koubaa, Y. & BenSaïda, A. (2017). Value-at-Risk under Lévy GARCH models: Evidence from global stock markets. Journal of International Financial Markets, Institutions and Money. 46, 30-53 https://doi.org/10.1016/j.intfin.2016.08.008

Titov, C (2022). Forecasting Exchange Rate Value-at-Risk and Expected Shortfall: A GARCH-EVT Approach. [Master Thesis, Lund University]. https://lup.lub.lu.se/luur

Tsay, R.S. (2010). Analysis of financial time series (Thired Edition). John Wiley & Sons Ltd.

Xekalaki, E. & Degiannakis, S. (2010). ARCH Models for Financial Applications. John Wiley & Sons Ltd.

Zivot, E. (2021). Introduction to Computational Finance and Financial Econometrics with R. https://bookdown.org/compfinezbook/introcompfinr/

A Appendix

A.1 Autocorrelation Function ACF

To understand the autocorrelation function, we must first define correlation. The correlation between two random variables X and Y is defined as:

$$\rho_{x,y} = \frac{Cov(X,Y}{\sqrt{var(X)Var(Y)}} = \frac{E[(X-\mu_x)(Y-\mu_y)]}{\sqrt{E(X-\mu_x)^2 E(Y-\mu_y)^2}}$$

Where μ_x and μ_y are the mean for X and Y, respectively. The correlation coefficient $\rho_{x,y}$ measures the linear dependence between X and Y and takes a value between -1 and 1. We can say that X and Y are uncorrelated when $\rho_{x,y} = 0$. Thus, for time series data, when the correlation between the return r_t and its past value r_{t-i} is of interest, the concept of correlation becomes autocorrelation. The definition of autocorrelation is then as follows:

$$\rho_\ell = \frac{Cov(r_t, r_{t-\ell})}{\sqrt{var(r_t)Var(r_{t-\ell})}} = \frac{Cov(r_t, r_{t-\ell})}{\sqrt{var(r_t)}}$$

According to Tsay (2010, p. 32), if r_t is a weakly stationary time series, $\hat{\rho}_\ell$ is normal with mean zero and variance $(1 + 2\sum_{i=1}^{q} \rho_i^2)/T$. Hence we can use this to test $H_0 : \rho_\ell = 0$ vs $H_a : \rho_\ell \neq 0$. The test is:

$$t\ ratio = \frac{\hat{\rho}_\ell}{\sqrt{(1 + 2\sum_{i=1}^{\ell-1} \hat{\rho}_i^2)/T}}$$

So we reject H_0 if $|t\ ratio| > Z_{\alpha/2}$.

A.2 Ljung-Box

The specification of the Ljung-Box test according to Tsay (2010, p. 32-34) is as follow:

$$Q(m) = T(T+2)\sum_{\ell}^{m} \frac{\hat{\rho}_{\ell=1}^2}{T-\ell}$$

Where $Q(m) \sim \chi_a^2$ with m degrees of freedom.

Table 6: *Ljung-Box test result*

Ljung-Box 20 lags	SWED-A.ST	SHB-A.ST	SEB-A.ST
squared return	p<0.00	p<0.00	p<0.00
absolute return	p<0.00	p<0.00	p<0.00

A.3 Maximum Likelihood Estimation

A.3.1 Normal Distribution

Assuming that the ε_t is normally distributed, the log-likelihood function according to Engle (1982) is given as:

$$l(\theta) = \frac{1}{T}\sum_{t=1}^{T} l_t(\theta),$$

$$l_t(\theta) = -\frac{n}{2}log(2\pi) - \frac{n}{2}log(\sigma^2) - \frac{1}{2}\frac{\varepsilon_t^2}{\sigma_t^2}$$

Where l is the average log-likelihood, l_t is the log-likelihood of the tth observation and T is the sample size. The partial derivation of the log-likelihood l_t with respect to the parameter vector θ is given as:

$$\frac{\partial l_t}{\partial \theta} = \frac{1}{2\sigma_t^2}\frac{\partial \sigma_t^2}{\partial \theta}\left(\frac{\varepsilon_t^2}{\sigma_t^2} - 1\right)$$

A.3.2 ARCH

ARCH model:

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

Define the parameter vector $\theta = (\omega, \alpha_1, ..., \alpha_q)$ which is the argument to be maximized. The partial derivation of the σ_t^2 with respect to θ:

$$\frac{\partial \sigma_t^2}{\partial \theta} = (1, \varepsilon_{t-1}^2, ..., \varepsilon_{t-q}^2)'$$

A.3.3 GARCH

We continue with the GARCH model and the parameter vector:

$$\theta = (\omega, \alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p)$$

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

$$\frac{\partial \sigma_t^2}{\partial \theta} = s_t + \sum_{j=1}^{p} \beta_j \frac{\partial \sigma_{t-j}}{\partial \theta}$$

where

$$s_t = (1, \varepsilon_{t-1}^2, ..., \varepsilon_{t-q}^2, \sigma_{t-1}^2, ..., \sigma_{t-p}^2)'$$

A.3.4 EGARCH

The EGARCH model with $\theta = (\alpha_0, \alpha_1, ..., \alpha_q, \gamma_1, ..., \gamma_q, \beta_1, ..., \beta_p)$:

$$log(\sigma_t^2) = \alpha_0 + \sum_{i=1}^{q} g\left(\frac{\varepsilon_{t-i}}{\sigma_{t-i}}\right) + \sum_{j=1}^{p} \beta_j \; log \; \sigma_{t-j}^2$$

$$g(\varepsilon_t/\sigma_t) = \alpha_1(\varepsilon_t/\sigma_t) + \gamma_1(|\varepsilon_t/\sigma_t| - E|\varepsilon_t/\sigma_t|)$$

The differentiating of $log \; \sigma_t^2$ with respect to θ:

$$\frac{\partial \; log\sigma_t^2}{\partial \theta} = v_t - \frac{1}{2} \sum_{i=1}^{q} \{\alpha_i(\varepsilon_{t-i}/\sigma_{t-i}) + \gamma_i(|\varepsilon_{t-i}/\sigma_{t-i}|)\} \frac{\partial \; log\sigma_{t-i}}{\partial \theta} + \sum_{j=1}^{p} \beta_j \frac{\partial \; log\sigma_{t-j}}{\partial \theta}$$

where $\varepsilon_t/\sigma_t = Z_t$ and
$v_t = (1, Z_{t-1}, ..., Z_{t-q}, |Z_{t-1}| - E|Z_{t-1}|, ..., |Z_{t-q}| - E|Z_{t-q}|, log \; \sigma_{t-1}, ..., log \; \sigma_{t-p})'$.

A.3.5 APARCH

$$\sigma_t^\delta = \alpha_0 + \sum_{i=1}^{q} \alpha_i(|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta,$$

We define the vectors $\gamma = (\gamma_1, ..., \gamma_q)$, $\theta = (\alpha_0, \alpha_1, ..., \alpha_q, \beta_1, ..., \beta_p)$ and the vector set $\eta = (\gamma, \theta, \delta)$. We can rewrite σ_t^2 as $(\sigma_t^\delta)^{\frac{2}{\delta}}$ as in Laurent (2004), we get:

$$\frac{\partial \sigma_t^2}{\partial(\theta,\gamma)} = \frac{2\sigma_t^2}{\delta \sigma_t^\delta} \frac{\partial \sigma_t^\delta}{\partial(\theta,\gamma)}$$

and

$$\frac{\partial \sigma_t^2}{\partial \delta} = \frac{2\sigma_t^2}{\delta \sigma_t^\delta} \Big(\frac{\partial \sigma_t^2}{\partial \delta} - \frac{\sigma_t^\delta \; ln(\sigma_t^\delta)}{\delta} \Big)$$

The partial derivation of σ_t^δ with respect to θ:

$$\frac{\partial \sigma_t^\delta}{\partial \theta} = d_t + \sum_{j=1}^{p} \beta_j \frac{\sigma_{t-j}^\delta}{\partial \theta}$$

where $d_t = (1, k(\varepsilon_{t-1})^\delta, ..., k(\varepsilon_{t-q})^\delta, \sigma_{t-1}^\delta, ..., \sigma_{t-p}^\delta)'$
and $k(\varepsilon_{t-i}) = (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})$.

The partial derivation of σ_t^δ with respect to γ:

$$\frac{\partial \sigma_t^\delta}{\partial \gamma} = \sum_{i=1}^{q} \alpha_i \frac{\partial k(\varepsilon_{t-i})^\delta}{\partial \gamma_i} + \sum_{j=1}^{p} \beta_j \frac{\sigma_{t-j}^\delta}{\partial \gamma}$$

$$\frac{\partial k(\varepsilon_{t-i})^\delta}{\gamma_i} = \begin{cases} -\delta k(\varepsilon_{t-i})^{\delta-1} \; \varepsilon_{t-i}, & \text{if } t > 0, \\ -\frac{\delta}{T} \sum_{s=1}^{T} (|\varepsilon_s - \gamma_i \varepsilon_s|)^{\delta-1} \varepsilon_s, & \text{if } t \leq 0 \end{cases}$$

and $\frac{\partial \sigma_t^\delta}{\partial \gamma} = 0 \; for \; t \leq 0$

The partial derivation of σ_t^δ with respect to δ:

$$\frac{\partial \sigma_t^\delta}{\partial \delta} = \sum_{i=1}^{q} \alpha_i [k(\varepsilon_{t-i})^\delta \; ln \; k(\varepsilon_{t-i})]^{F_{t-i}} \left[\frac{1}{T} \sum_{s=1}^{T} (|\varepsilon_s| - \gamma_i \varepsilon_s)^\delta \; ln(|\varepsilon_{t-i}| - \gamma_i \varepsilon_s) \right]^{1-F_{t-i}}$$

$$+ \sum_{j=1}^{p} \beta_j \left(\frac{\partial \sigma_{t-j}^\delta}{\partial \delta} \right)^{F_{t-j}} \left[0.5 \left(\frac{1}{T} \sum_{s=1}^{T} \varepsilon_s^2 \right)^{\frac{\delta}{2}} ln \left(\frac{1}{T} \sum_{s=1}^{T} \varepsilon_s^2 \right) \right]^{1-F_{t-j}}$$

A.3.6 t-distribution

Assuming that the ε_t follows t-distributed, the log-likelihood function is given as:

$$l_t(\theta) = log\left[\Gamma\left(\frac{\nu+1}{2}\right)\right] - log\left[\Gamma\left(\frac{\nu}{2}\right)\right] - \frac{1}{2}log(\pi(\nu-2))$$

$$-\frac{1}{2}\sum_{t=1}^{n}\left[log(\sigma_t^2) + (1+\nu)log\left(1+\frac{\varepsilon_t^2}{\sigma_t^2(\nu-2)}\right)\right]$$

A.3.7 Skew t-distribution

Assuming that the ε_t follows skew t-distributed, the log-likelihood function is given as:

$$l_t(\theta) = log\left[\Gamma\left(\frac{\nu+1}{2}\right)\right] - log\left[\Gamma\left(\frac{\nu}{2}\right)\right] - \frac{1}{2}log(\pi(\nu-2))$$

$$+ log\left(\frac{2}{\xi+\frac{1}{\xi}}\right) + log(s) - \frac{1}{2}\sum_{t=1}^{n}\left[log(\sigma_t^2) + (1+\nu)log\right.$$

$$\left.\left(1+\left(\frac{s\varepsilon_t}{\sigma_t^2(\nu-2)}+\frac{m}{\nu-2}\right)\xi^{-I_t}\right)\right]$$

Where ξ is the asymmetry parameters, ν the degree of freedom of the distribution and I is a indicator function takes two values as:

$$I_t = \begin{cases} 1 & \text{if } \varepsilon_t \geq -\frac{m}{s} \\ -1 & \text{if } \varepsilon_t < -\frac{m}{s} \end{cases}$$

Where

$$m = \frac{\Gamma(\frac{\nu+1}{2})\sqrt{\nu-2}}{\sqrt{\pi}\Gamma(\frac{\nu}{2})}(\xi - \frac{1}{\xi})$$

$$s = \sqrt{(\xi^2 + \frac{1}{\xi^2} - 1)} - m^2$$