Stockholms universitet

Forecasting One-Day-Ahead Value at Risk for a Stock Portfolio

Sebastian Baarsen

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2025:15
Matematisk statistik
Juni 2025

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Stockholm University

Mathematical Statistics
Stockholm University
Bachelor Thesis **2025:15**
http://www.math.su.se

Forecasting One-Day-Ahead Value at Risk for a Stock Portfolio

Sebastian Baarsen*

June 2025

Abstract

We compare three one-day-ahead VaR forecast methods, Delta-Normal, GARCH(1,1) with Student-t residuals, and EGARCH(1,1), on an equally weighted portfolio of SEB-A and Swedbank-A using daily data from 2010 to 2025. Forecasts are generated in a rolling 500-day window and evaluated by RMSE, tick loss, and VaR backtests (Kupiec's POF and Christoffersen's conditional coverage). Results show that GARCH and EGARCH both deliver violation rates close to the target of 10%, whereas Delta-Normal produces only around 7%. Moreover, EGARCH yields the lowest RMSE and tick loss, demonstrating its superior ability to capture time-varying and asymmetric volatility.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail: sebastian.baarsen@gmail.com. Supervisor: Kristofer Lindensjö and Jan Olov Persson.

Acknowledgements

I would like to express my deepest gratitude to my supervisors, Kristofer Lindensjö and Jan Olov Persson, for their guidance and support throughout the completion of this thesis. In particular, I am especially grateful to Kristofer for generously giving his time beyond our scheduled meetings to answer my questions and discuss ideas.

I would also like to acknowledge the use of ChatGPT as an aid for R programming and for optimizing computationally intensive scripts, as well as for assistance with structuring and refining the text and formatting tables and figures.

Contents

1 Introduction

Value at Risk (VaR) is one of the most common measures for quantifying financial risk. It forecast the maximum loss a portfolio can suffer over a given time horizon at a specified confidence level. This makes VaR an important tool for investors and risk managers to monitor their exposure to market fluctuations. In particular, a to conservative VaR forecast can lead to missed investment opportunities, whereas a too aggressive forecast may expose the portfolio to excessive risk.

The aim of this bachelor thesis is to compare three different one-day-ahead VaR Forecast methods: the Delta Normal method, a GARCH(1,1) model with Student-t residuals and an EGARCH(1,1) model that captures asymmetric volatility on an equally weighted portfolio of SEB-A and Swedbank-A.

Forecast accuracy is evaluated using Root Mean Square Error, tick loss and standard backtesting procedures such as Kupiec test and conditional coverage.

This study is limited to daily data for a two-stock portfolio and does not consider larger or more diversified portfolios. However, the theoretical framework could be easily extended to larger portfolios.

Data covering 2010 to 2025 are sourced from Yahoo Finance. We apply a rolling window with 500 trading days per window. All computations are performed in R using `tidyquant` for data handling and `rugarch` for the GARCH family models.

This is the following structure of the thesis: Chapter 2 reviews the theoretical background: time series models, key concepts, and diagnostic tests. Chapter 3 describes the data and our in-sample/out-of-sample split. Chapter 4 details the empirical modeling steps. Chapter 5 presents the results. Chapter 6 discusses findings, limitations, and potential improvements. Finally, Chapter 7 offers an extension, introducing a copula-based method for VaR forecasting.

2 Theoretical Background

In this section we review the necessary theoretical background, including important notions, concepts, and time series theory that will serve as the foundation for our analysis.

We begin with basic definitions and properties of time series data, such as stationarity, autocorrelation, and volatility. We then introduce the concepts behind ARCH and GARCH models, and discuss the extended EGARCH variants used in volatility forecasting.

This theoretical framework provides the context for the empirical methods discussed in later sections.

2.1 Notations and Concepts

Here we will describe some fundamental concepts that will be used throughout the study.

2.1.1 Returns

Let P_t denote the daily closing price of a stock at time t. The simple return is defined in (Tsay, 2005, p. 3) as

$$R_t = \frac{P_t}{P_{t-1}} - 1.$$

While simple returns are intuitive, it is common in financial modeling to work with log returns, which are scale-free and additive over time. The log return at time t is therefore defined by (Tsay, 2005, p. 5)

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

Throughout this thesis we will use log returns r_t as our primary return series.

2.1.2 Value at Risk (VaR)

Value at Risk (VaR) is a widely used risk measure in financial analysis. As Jorion (2007, p. 17) states, "*Value at Risk summarizes the worst loss over a target horizon that won't be exceeded with a given level of confidence*".

We formally define VaR, in line with Jorion (2007, p. 107), at a confidence level $1-\alpha$ as the level $VaR_{1-\alpha}$ such that

$$P\left(r_t \leq \text{VaR}_{1-\alpha}\right) = \alpha,$$

where r_t is the log return at time t. Thus, for $\alpha = 0.1$ (a 90% confidence level), $\text{VaR}_{0.90}$ is the 10th percentile of the distribution r_t, see Figure 1.

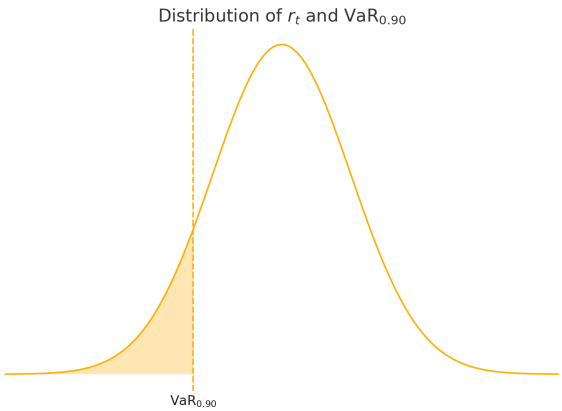


Figure 1: Illustration of return distribution with marking for $\text{VaR}_{0.90}$.

Sometimes the risk-free rate is also included in the calculation of the future portfolio value. However, in this analysis, the risk-free rate is omitted since we are going to compute VaR for a one-day horizon. Over such a short period, the effect of the risk-free rate is negligible and does not significantly impact the results.

2.1.3 Volatility

Following Tsay (2013, p. 177-178), we define volatility as the conditional standard deviation of the daily return:

$$\sigma_t = \sqrt{\text{Var}(r_t \mid \mathcal{F}_{t-1})},$$

where $\mathcal{F}_{t-1}$ is all the information available at time $t-1$ (formally a σ-algebra, but we use the informal "information set" notion following Tsay).

Furthermore, volatility exhibits some well-known characteristics (see Tsay (2005, p. 97-98)). One such feature is clustering: large market fluctuations tend to be followed by high-volatility days.

Moreover, stock return volatility responds asymmetrically to shocks: negative movements typically leads to a larger increase in volatility than positive shocks, Tsay refer to this as the *leverage* effect. This asymmetry will be analyzed further in the EGARCH section.

2.2 Time Series Analysis

In this section, we review some basic concepts of time series analysis, such as autocovariance and autocorrelation. We then proceed with simpler time series models, leading up to volatility models such as ARCH, GARCH, and EGARCH. Finally, we conclude with diagnostic tests for time series and methods for model validation. This theory will provide the groundwork for our Value at Risk modeling.

2.2.1 What is a Time Series?

In Brockwell and Davis (1991, p. 1), a time series is a set of observations ordered in time. An example of a time serie is stock data, where the daily closing prices are observations from a set of random variables $\{r_t\}$.

2.2.2 Stationarity

Following Tsay (2005, p. 25), a time series $\{r_t\}$ is weakly stationary if its mean is constant,

$$E[r_t] = \mu,$$

and the covariance between r_i and r_j depends only on the lag:

$$\text{Cov}(r_i, r_j) = \gamma_{|i-j|}.$$

In practice, this means that the process fluctuates around a constant value μ, and deviations from μ occur with constant magnitude.

For simplicity, we will refer to a weakly stationary time serie simply as *stationary*.

2.2.3 Autocovariance- & Autocorrelation Functions

Further, for a stationary time series $\{r_t\}$, the autocovariance function at lag l is defined as

$$\gamma_l = \text{Cov}(r_t, r_{t+l}).$$

Moreover, the autocorrelation function (ACF) at lag l is

$$\rho(l) = \frac{\text{Cov}(r_t, r_{t+l})}{\sqrt{\text{Var}(r_t)\,\text{Var}(r_{t+l})}}.$$

If the series is stationary, this simplifies to

$$\rho(l) = \frac{\gamma_l}{\gamma_0},$$

since

$$\sqrt{\text{Var}(r_t)\,\text{Var}(r_{t+l})} = \sqrt{\text{Var}(r_t)\,\text{Var}(r_t)} = \gamma_0.$$

The autocorrelation function will play a particularly important role in later applications (Tsay, 2005, p. 25-26).

2.2.4 White Noise Process

A basic time series is the white noise process, where $\{\epsilon_t\}$ is i.i.d. with finite mean and variance. In a white noise process there are no correlations for lags greater than one, meaning that the series has no systematic structure (Tsay, 2005, p. 31).

One example of a white noise process is a standardized Student-t white noise process. Let

$$\epsilon_t \sim \sqrt{\frac{\nu-2}{\nu}}\, t_\nu,$$

where t_ν is a standard Student-t random variable with ν degrees of freedom, mean zero and variance $\frac{\nu}{\nu-2}$. The scaling factor $\sqrt{\frac{\nu-2}{\nu}}$ ensures that ϵ_t has unit variance.

Note that a white noise process does not necessarily have unit variance. However, when modeling volatility with ARCH-family models, we assume the residuals are standardized to have unit variance, more on this in Section 2.2.8-2.2.11.

2.2.5 The ARMA model

Next we briefly consider the generalized ARMA model. We will show in Section 4 that our implementation does not require fitting an ARMA model to the financial time series. Therefore, we will not discuss the details here. Interested readers are referred to Tsay (2005, p. 58), upon which this section is based.

An ARMA(p, q) model decomposes the series r_t into a mean equation μ_t and a residual part a_t:

$$r_t = \mu_t + a_t,$$

where a_t is a white noise process, and the mean equation is

$$\mu_t = \phi_0 + \sum_{i=1}^{p} \phi_i \, r_{t-i} - \sum_{j=1}^{q} \theta_j \, a_{t-j}.$$

We can interpret each ϕ_i parameter as a measures of how much of the past value r_{t-i} the process remembers, and each θ_j parameter as how much the process remembers from past shock terms.

One can also show that an ARMA process is stationary only if all the roots of the characteristic equation

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p$$

lie outside the unit circle, that is, $|z| > 1$.

2.2.6 Time Series Model Building for the Log Returns

We assume that the log returns follows the form

$$r_t = \mu_t + a_t,$$

where a_t is white noise with $E[a_t] = 0$ and $\mathrm{Var}(a_t) = \sigma^2$.

For the mean equation μ_t, Tsay (2005, p. 101) recommends using an ARMA(p, q) model. Empirically, daily stock log returns exhibit low serial correlation, so it is sometimes sufficient to use the simplest case, ARMA$(0, 0)$, which yields

$$\mu_t = \bar{r},$$

where $\bar{r}$ is the sample mean of the log returns.

It is often the case that $\bar{r}$ is very close to zero. Therefore, we can set

$$r_t = a_t.$$

To justify this assumption, we conduct a one-sample t-test on $\bar{r}$ (see details of the t-test procedure in Appendix A). As shown in Section 4.2.2, the test confirms $\bar{r} \approx 0$. Thus, we proceed with $r_t = a_t$.

Hence, our primary focus is on the white noise process a_t, which captures the volatility dynamics of the log returns. In the following sections, we model the conditional variance ($\sigma_t^2 = \mathrm{Var}(a_t \mid \mathcal{F}_{t-1})$) using ARCH, GARCH and EGARCH models to account for time-varying and asymmetric volatility.

2.2.7 Conditional Variance

To capture the time-varying volatility in stock returns, we need to broaden our modeling approach beyond the ARMA model.

The conditional variance of r_t is defined as

$$\sigma_t^2 = \text{Var}(r_t \mid \mathcal{F}_{t-1}),$$

where $\mathcal{F}_{t-1}$ represents the information set available up to time t, for us this means all the observations $r_{t-1}, ..., r_1$ up until time t (Tsay, 2005, p. 100).

For a time series, the unconditional variance, $\text{Var}(r_t)$, can be constant for all t, but the conditional variance is still allowed to change over time, capturing periods of both high and low volatility.

To account for this dynamic, we extend the model (see Tsay (2005, p. 103)) for the return series as

$$r_t = \mu_t + a_t,$$

with

$$a_t = \sigma_t \epsilon_t,$$

where μ_t is the mean equation and $\sigma_t \epsilon_t$ denotes a scaled white noise process. In this formulation, σ_t captures the conditional volatility at time t, while ϵ_t is a white noise process with zero mean and unit variance. This setup forms the basis for ARCH, GARCH, and EGARCH models.

2.2.8 Auto Regressive Conditional Heteroskedasticity Modeling

In the follow section, we go through three different volatility models, that is often used on financial time series data. First, we discuss the standard ARCH/GARCH model, which models the conditional variance as a function of past squared shocks and past variances. Then we cover the EGARCH model, which extends the GARCH model by allowing for asymmetric responses to positive and negative movements.

2.2.9 ARCH Model

The ARCH(p) model, first proposed by Engle (1982), is one of the simplest models for capturing conditional heteroscedasticity in time series data. An ARCH(1) process, a_t, is defined in Tsay (2005, p. 103) as

$$a_t = \sigma_t \epsilon_t,$$

with

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

where, ϵ_t is a white noise process with zero mean and unit variance, and the parameters satisfy

$$\alpha_0 > 0, \quad \alpha_1 \geq 0, \quad \alpha_1 < 1.$$

to ensure stationarity. It is common to assume ϵ_t follows either a standard normal, or a standardized Student-t distribution.

In Appendix B we derive further properties of the ARCH(1) model. In summary, the ARCH(1) process a_t results in a process that has a white noise–like structure, in terms of its unconditional mean and variance, but since the unconditional variance is time-dependent, the ARCH model allow us to model the time-varying volatility, and its clustering.

The ARCH model can be generalized to ARCH(p), where p is the number of past terms in the process that should be included in the model where, that is

$$a_t = \sigma_t \epsilon_t,$$

with

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \, a_{i-1}^2,$$

similarly as above

$$\alpha_0 > 0, \quad \alpha_i \geq 0, \quad \sum_{i=1}^{p} \alpha_i < 1 \quad (i = 1, .., p),$$

to ensure stationarity, and where ϵ_t is once agian a white noise process with zero mean and unit variance.

As stated in Tsay (2005, p. 105), the general properties derived in Appendix B for ARCH(1) extends to the ARCH(p) process, although the calculations becomes cumbersome. That is, both the conditional and unconditional means are zero, the conditional variance is time-dependent, given by

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i a_{t-i}^2,$$

and the unconditional variance is constant and equal to

$$\sigma^2 = \frac{\alpha_0}{1 - \sum_{i=1}^{p} \alpha_i}.$$

A drawback of the ARCH(p) model is that it often requires a large number of lagged terms to sufficiently capture the volatility dynamics in financial data (Tsay, 2005, p. 113). A solution is to extend the ARCH model to the so-called GARCH model. This will be the topic of the next section.

2.2.10 GARCH Model

An extension of the ARCH model is the GARCH model, proposed by Bollerslev (1986). The idea is to not only to account for past values of the process,

a_{t-i}, as with ARCH(p), but also to incorporate the past conditional variances, σ^2_{t-i}. In a GARCH$(1,1)$ process, the model for a_t is defined as

$$a_t = \sigma_t \epsilon_t,$$

where

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

Here, $\{\epsilon_t\}$ is a white noise process with zero mean and unit variance. The model parameters must satisfy the constraints

$$\alpha_0 > 0, \quad \alpha_1 \geq 0, \quad \beta_1 \geq 0, \quad \alpha_1 + \beta_1 < 1$$

to guarantee stationarity (Tsay, 2005, p. 114).

Similarly to the ARCH(1) model, we have included some properties of the GARCH(1,1) model in Appendix B, and analogous to the ARCH(p) extension, we can extend the GARCH(1,1) model by including additional lagged terms.

The GARCH(p, q) model is defined similar as above:

$$a_t = \sigma_t \epsilon_t,$$

where instead

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

where again $\{\epsilon_t\}$ is a white noise process. To ensure stationarity, the parameters should satisfy

$$\alpha_0 > 0, \quad \alpha_i \geq 0 \quad (i = 1, \ldots, p), \quad \beta_j \geq 0 \quad (j = 1, \ldots, q),$$

and

$$\sum_{i=1}^{q} \alpha_i + \sum_{j=1}^{p} \beta_j < 1.$$

Even though GARCH(p, q) can include many lag terms, a simple GARCH(1,1) model is often sufficient for financial data (Tsay, 2005, p. 116).

2.2.11 EGARCH Model

Since the GARCH model models the squared conditional volatility, it weighs positive and negative shocks equally. But, as stated in Section 2.1.3, there is often a *leverage* effect. This means that negative shocks tend to have a larger impact on volatility than positive shocks of the same size. This asymmetry motivates the use of the EGARCH model, as proposed by Nelson (1991).

A representation of the EGARCH(m,s) model is given by

$$a_t = \sigma_t \epsilon_t,$$

where

$$\ln(\sigma_t^2) = \omega + \sum_{i=1}^{s} \alpha_i \frac{|a_{t-i}| + \gamma_i a_{t-i}}{\sigma_{t-i}} + \sum_{j=1}^{m} \beta_j \ln(\sigma_{t-j}^2),$$

(see Tsay (2005, p. 125)). Here ϵ_t is a white noise process with zero mean and unit variance. Here, ω, α_i, γ_i, and β_j are constants.

Note the term

$$|a_{t-i}| + \gamma_i a_{t-i} = \begin{cases} a_{t-i}(\gamma_i + 1), & \text{if } a_{t-i} > 0, \\ a_{t-i}(\gamma_i - 1), & \text{if } a_{t-i} \leq 0, \end{cases}$$

this shows that negative and positive movements contribute asymmetrically to σ_t^2.

Since the model uses the logarithm, the positivity constraints on the parameters are relaxed Tsay (2005, p. 125). To ensure stationarity it is sufficient that $\sum_{j=1}^{m} \beta_j < 1$. More details of the properties of the EGARCH model can be seen in Nelson (1991).

2.3 Statistical Tests for Time Series

Model checking is an important step in any data-fitting assesment. So, in this section, some common statistical tests and validation methods will be presented.

2.3.1 Augmented Dickey-Fuller test

The whole analysis will be done under the assumption that the time series of daily log return is stationary, therefore, it is crucial to check whether this assumption is reasonable. To test if a series $\{x_t\}$ is stationary, we apply the Augmented Dickey-Fuller (ADF) test. The ADF regression can be written as

$$\Delta x_t = \alpha_t + (\beta - 1)\, x_{t-1} + \sum_{i=1}^{p-1} \phi_i\, \Delta x_{t-i} + \epsilon_t, \tag{1}$$

where $\Delta x_j = x_j - x_{j-1}$, ϕ_i are constants, α_t is some function of t (constant or some trend) and ϵ_t is assumed to be white noise (Tsay, 2005, p. 69).

The null hypothesis

$$H_0 \colon \beta = 1,$$

i.e. the process has a unit root. In that case each shock ϵ_t has a permanent effect and the series is non-stationary. Under H_0, and when we set $\alpha_t = 0$ for our daily log return series, Equation 1 can be written as

$$x_t = x_{t-1} + \sum_{i=1}^{p-1} \phi_i\, \Delta x_{t-i} + \epsilon_t$$

This is a random walk where each ϵ_t is the increments (non-stationary).

Under the alternative

$$H_1 : \beta < 1,$$

and by the same reasoning as above,

$$x_t = \beta\, x_{t-1} + \sum_{i=1}^{p-1} \phi_i\, \Delta x_{t-i} + \epsilon_t.$$

Since $|\beta| < 1$, each past value x_{t-1} is scaled down by β each period, so any shock of ϵ_t gets smaller and smaller over time. In other words, the effect of ϵ_t quickly fades away rather than sticking around, making the series stationary.

Moreover, the ADF test statistic is

$$t_\beta \;=\; \frac{\hat{\beta} - 1}{\mathrm{std}(\hat{\beta})},$$

where $\hat{\beta}$ is the least square estimate and $\mathrm{std}(\hat{\beta})$ its standard error. When comparing t_β to critical values from the Dickey-Fuller distribution, a sufficiently negative value leads us to reject H_0, and conclude that the series is stationary.

2.3.2 Ljung-Box

An important property when working with time series data is to assess whether there is any autocorrelation, that is whether there is evidence of any correlation between the lags in the data.

One way of testing this is with the Ljung-box test. We begin by defining the sample autocorrelation at lag l for a sample of returns $\{r_t\}_{t=1}^{T}$:

$$\hat{\rho}_l = \frac{\sum_{t=l+1}^{T} (r_t - \bar{r})\,(r_{t-l} - \bar{r})}{\sum_{t=l+1}^{T} (r_t - \bar{r})^2}, \qquad 0 \le l < T,$$

where $\bar{r} = \sum_{i=1}^{T} r_t$ (Tsay, 2005, p. 26). Using these sample autocorrelations, the Ljung-Box statistic for m lags in a sample of size T is defined by

$$Q(m) = T\,(T+2) \sum_{l=1}^{m} \frac{\hat{\rho}_l^2}{T-l},$$

see Tsay (2005, p. 27). The $Q(m)$ statistic tests the joint null hypothesis, that all autocorrelations up to lag m are zero. A common choice of m is $m \approx \ln(T)$.

The null hypothesis is that

$$\hat{\rho_1} = ... = \hat{\rho_m} = 0.$$

Moreover, under the null hypothesis, $Q(m)$ is asymptotically distributed as a chi-squared random variable, with m degrees of freedom. Therefore, if the p-value corresponding to $Q(m)$ is less than a chosen significance level, often 5%, the null hypothesis is rejected, indicating that there is some evidence of autocorrelation in the data.

When checking whether the standardized residuals, see Section 2.3.5, are serially correlated this test will be useful. This test will also come in handy in the next section when we examine if any time-varying volatility is present.

2.3.3 ARCH Effects

Before specifying a volatility model, we first check if any time-varying variance is present: after fitting an ARMA(p, q) to the mean equation μ_t, by maximum likelihood (see Section 2.7), we form the residuals

$$a_t = r_t - \mu_t,$$

which in practice reduces to $a_t = r_t$, since $\mu_t = 0$. This assumption is motivated in Section 4.2.2.

Even though the series $\{a_t\}$ may behave like white noise, its squared series $\{a_t^2\}$ often exhibits significant serial correlation, indicating that dependence arises in the second moment of a_t. To detect this, we will apply the Ljung-Box test to $\{a_t^2\}$ (Tsay, 2005, p. 101).

2.3.4 ACF to Detect Autocorrelation

It is shown in Brockwell and Davis (1991) that the sample autocorrelation $\hat{\rho}_l$ is asymptotically normal, i.e.

$$\hat{\rho}_l \overset{a}{\sim} N\Big(0, \tfrac{1}{T}\Big).$$

Thus, when plotting the ACF we include horizontal lines at

$$\pm 1.96 \, \frac{1}{\sqrt{T}},$$

as illustrated in Figure 6. Any sample autocorrelation $\hat{\rho}_l$ for $l > 0$ that crosses these bounds is significant at the 5% level.

By examining the sample autocorrelation function (ACF) of $\{a_t\}$, we can detect whether any lags exhibit significant autocorrelation.

2.3.5 Standardized Residuals & Model Checking

For the ARCH, GARCH and EGARCH models we consider

$$a_t = \sigma_t \epsilon_t,$$

with $\{\epsilon_t\}$ being a white noise process with zero mean and unit variance. Recall that we assumed a distribution for ϵ_t, usually normal or Student-t. As seen in Tsay (2005, p. 109), we form the standardized residuals for our data as

$$\tilde{a}_t = \frac{a_t}{\hat{\sigma}_t},$$

where $\hat{\sigma}_t$ is the estimated conditional standard deviation obtained from the fitted model (see Section 2.7). The series $\{\tilde{a}_t\}$ is of interest when assessing whether the model is appropriate. If the model is correctly specified, it should hold that $\tilde{a}_t \approx \epsilon_t$, i.e., $\{\tilde{a}_t\}$ should only exhibit white noise behavior. To see if this holds for the data, we create a QQ-plot of $\{\tilde{a}_t\}$ against the theoretical quantiles of the chosen distribution. In particular, if ϵ_t is assumed to follow a standardized Student-t distribution with ν degrees of freedom, the theoretical quantiles are

$$q_\alpha = \sqrt{\frac{\nu-2}{\nu}}\, t_{\nu,\alpha},$$

where $t_{\nu,\alpha}$ denotes the α quantile of the standard Student-t distribution with ν degrees of freedom. See Section 4.2.5 for an illustration of such a QQ-plot.

Furthermore, to verify that $\{\tilde{a}_t\}$ contains no remaining autocorrelation, we conduct the Ljung-Box test on $\{\tilde{a}_t\}$ and $\{\tilde{a}_t^2\}$, see Section 4.2.5. If there is evident of any serial correlation, this would mean that the standardized residuals at different time points are related, i.e., it remains some structure or "memory" in the residuals that the model has not captured. Finding such dependence suggests that the volatility model is mispecified.

2.4 Model Selection

When selecting a model, one can use the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC). Both criteria balance the model's fit against its complexity (the number of parameters), but they apply different weightings:

AIC: The Akaike Information Criterion is defined as

$$\mathrm{AIC} = -2\ln(L) + 2k.$$

BIC: The Bayesian Information Criterion is defined as

$$\mathrm{BIC} = -2\ln(L) + k\ln(n).$$

Here n denotes the sample size, k the number of estimated parameters, and L the likelihood, some details on likelihood theory are provided in Appendix D (Held and Bové, 2020, Chapter 7). Intuitively, the likelihood quantifies how well the model explains the observed data.

BIC has a larger penalty for complex model when the sample sizes is large, this leads to simpler models being preferred compared to AIC.

These criteria are used to select a model that provides a balance between a good fit and complexity. When deciding on the order of the model for the ARMA(p, q), GARCH(p, q) and EGARCH(m, s), these information criteria will come in handy.

2.5 Backtesting and VAR Validation

Our primary goal is to forecast Value at Risk. Once these VaR forecasts have been generated, it is important to validate them to assess model accuracy.

For datasets without a time index, one might use methods such as cross-validation, where the model is trained on a subset of the data and evaluated on the remaining observations. This is approach is not applicable for time series data. Instead, a commonly used method is backtesting using a rolling window. Before proceeding, we introduce terminology that will be used throughout the following sections.

2.5.1 VaR Violations

A *violation* (or *exception*) occurs when the realized loss exceeds the forecasted Value at Risk threshold. For example, if we compute the 90% VaR, denoted $\text{VaR}_{0.90}$, we expect returns to fall below this threshold in approximately 10% of cases. Tracking the frequency and the distribution of these violations is important to determine if the model is accurately capturing the true risk of the investment, this will be discussed more in Section 2.6.

2.5.2 Rolling Window

The idea of a rolling window is that, given a window size $n < T$, where T is the total sample size, to let it "roll" over the entire dataset. In each step, the window moves forward by discarding the first data point and including the next one.

For each window, we estimate our model on that entire dataset, and forecast VaR one step ahead, then compare the forecasted VaR with the actual log return to determine whether a violation occurs. This ensures that the data used for forecasting are excluded from the estimation sample, thereby avoiding overfitting. This procedure yields $T - n$ rolling windows, and consequently produces $T - n$ one-day-ahead VaR forecasts. We will use this approach to analyze the predictive performance of the model over time. Figure 2 is illustrating the procedure the rolling window technique.

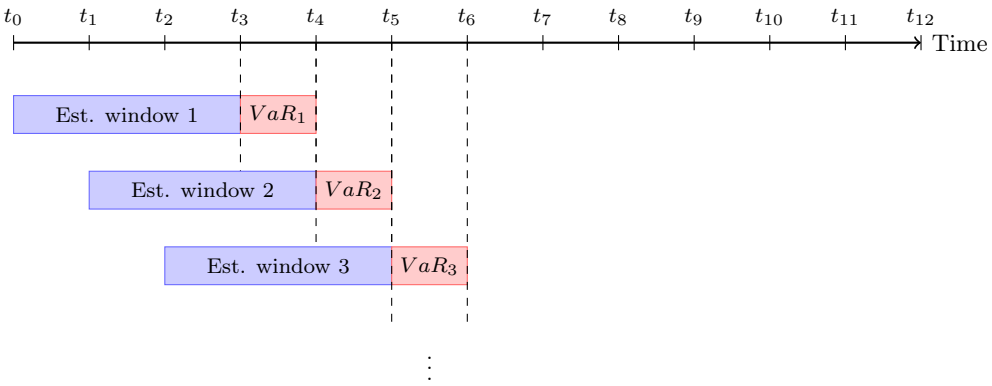


Figure 2: Rolling window illustration - the model is first fitted to the data in window i and is then use to forecast VaR_i forward into the future.

2.5.3 The Root Mean Squared Error (RMSE)

Our goal is to evaluate how well our volatility models capture the true conditional variance σ_t^2. Since the true σ_t^2 is not directly observable, we replace it with a proxy. We assume that $\mathrm{E}[r_t \mid \mathcal{F}_{t-1}] = 0$ (motivated in Section 4.2.2.), under this assumption

$$\mathrm{Var}(r_t \mid \mathcal{F}_{t-1}) = \mathrm{E}[r_t^2 \mid \mathcal{F}_{t-1}],$$

i.e., the conditional variance of r_t is on average the realized value of r_t^2. Hence, we define the realized variance (proxy) as

$$RV_t \;:=\; r_t^2,$$

i.e., the squared log return at time t, which serves as an observable substitute for the unobservable σ_t^2.

Using rolling windows of length $n = 500$ over a total sample of size T, we have $T - n$ one-day-ahead forecasts. Let $\hat{\sigma}_{t+1}^2$ denote the model's forecast of the conditional variance for day $t+1$ based on the rolling window ending at time t, and let RV_{t+1} be the realized variance proxy. Following Hansen and Lunde (2005, Sec 3), we compute the Root Mean Squared Error as

$$\mathrm{RMSE} = \sqrt{\frac{1}{T-n}\sum_{t=n}^{T-1}\left(\hat{\sigma}_{t+1}^2 - RV_{t+1}\right)^2}.$$

A lower RMSE indicates that the one-day-ahead forecasts $\hat{\sigma}_{t+1}^2$ are on average closer to the observed realized variances RV_{t+1}.

2.6 VaR Forecast Diagnostics

In this section, we outline t he theory behind some diagnostic tools used to assess our one-day-ahead VaR forecasts: the count of violations, Kupiec's Proportion of Failures (POF) test, Christoffersen's Conditional Coverage test, and the Tick Loss function. The empirical implementation of these tests is presented later in Chapter 5.

2.6.1 Test for Number of Violations

When forecasting VaR at level $1-\alpha$, we expect $\alpha \cdot T$ violations out of T trails. In other words, if the VaR model is correctly specified, each observation from $\{r_t\}$ should have probability α of not exceeding the forecasted VaR_{1-a}^t. Let

$$I_t = \begin{cases} 1, & \text{if a violation occurs at time } t, \\ 0, & \text{otherwise.} \end{cases}$$

We assume that $\{I_t\}$ are i.i.d. Bernoulli(α). Then, the total number of violations

$$X = \sum_{t=1}^{T} I_t$$

is distributed as

$$X \sim \text{Bin}(T, \alpha),$$

with

$$\text{E}[X] = \alpha T, \quad \text{Var}(X) = \alpha(1-\alpha)\, T.$$

This binomial setting leads us to the Kupiec-Test explained in the next section.

2.6.2 Kupiec's Proportion of Failures (POF) Test

Following Kupiec (1995, p. 79), we compare the observed violation rate with the expected rate α for $VaR_{1-\alpha}$.

Let $X \sim \text{Bin}(T, p)$ be the number of violations in T number of trials, and x denote the observed number of violations.

The two hypotheses are

$$H_0 : p = \alpha \quad \text{and} \quad H_1 : p = x/T,$$

Under null hypothesis, the likelihood is

$$L_0 = (1-\alpha)^{T-x}\, \alpha^x,$$

and under the alternative $\hat{p} = x/T$, i.e. the observed violation proportion, gives

$$L_1 = (1-\hat{p})^{T-x}\, \hat{p}^x,$$

The Kupiec test statistic is

$$LR_{POF} = -2\ln\left(\frac{L_0}{L_1}\right) = -2\ln\left[\frac{(1-\alpha)^{T-x}\,\alpha^x}{(1-\hat{p})^{T-x}\,\hat{p}^x}\right],$$

which under H_0 is asymptotically χ^2_1. For significantly large values of LR_{POF}, we reject the null hypothesis, equivalently, if the corresponding p-value is significantly small.

The POF test provides a useful check on the violation rate, but it does not check whether violations are time-dependent. This is the topic of the next section.

2.6.3 Conditional Coverage Test

The Kupiec test checks if the total number of VaR violations is in line with the expected rate α for $\text{VaR}_{1-\alpha}$. However, as mentioned, it does not tell us if these violations happen independently over time. If violations tend to cluster, it might indicate that the model struggles during periods of high volatility.

The Conditional Coverage Test (Christoffersen, 1998, Chapter 3), also known as the Christoffersen Test, does two things: it verifies if the number of violations is in line with our expectations, and it checks if the violations occur independently over time.

The idea behind this test is that, if there is no clustering of the violations, the probability of transitioning from a period in which a violation has occurred to a period with a new violation should be the same as the probability of transitioning from a period with no violation to a period with a violation.

The Conditional Coverage Test statistic is defined as

$$LR_{CC} = LR_{POF} + LR_{ind},$$

where LR_{ind} is defined as

$$LR_{ind} = -2\log\left(\frac{(1-\pi)^{n_{00}+n_{10}}\pi^{n_{01}+n_{11}}}{(1-\pi_0)^{n_{00}}\pi_0^{n_{01}}(1-\pi_1)^{n_{10}}\pi_1^{n_{11}}}\right),$$

where LR_{POF} as defined above. The counts n_{ij} are as follows:

- n_{00}: Number of days with no violation at $t-1$ and no violation at t.
- n_{10}: Number of days with a violation at $t-1$ and no violation at t.
- n_{01}: Number of days with no violation at $t-1$ and a violation at t.
- n_{11}: Number of days with a violation at $t-1$ and a violation at t.

The probabilities are defined by

$$\pi_0 = \frac{n_{01}}{n_{00} + n_{01}}, \quad \pi_1 = \frac{n_{11}}{n_{10} + n_{11}}, \quad \pi = \frac{n_{01} + n_{11}}{n_{00} + n_{01} + n_{10} + n_{11}}.$$

Under the null hypothesis of the Conditional Coverage Test, the violations form an i.i.d. Bernoulli(α) sequence, i.e., both the violation rate equals α (tested by LR_{POF}) and violations are independent over time (tested by LR_{ind}). The combined statistic

$$LR_{CC} \;=\; LR_{POF} + LR_{ind}$$

tests this joint hypothesis, and under H_0 is asymptotically χ^2_2. A large value of LR_{CC} leads us to reject H_0, indicating either incorrect violation rate, clustering of violations, or both.

A drawback of this test is that it only checks for dependence at a single lag, when dependencies could occur at higher-order lags.

2.6.4 Tick Loss

We define the tick loss function for our forecasted VaR sequence $\{\mathrm{VaR}^{t+1}_{1-\alpha}\}^{T-1}_{t=n}$ following Gneiting and Raftery (2007, p. 370) as

$$L_\alpha\big(r_{t+1}, \mathrm{VaR}^{t+1}_{1-\alpha}\big) = \big(\mathbf{1}\{r_{t+1} < \mathrm{VaR}^{t+1}_{1-\alpha}\} - \alpha\big)\,\big(r_{t+1} - \mathrm{VaR}^{t+1}_{1-\alpha}\big).$$

The average tick loss over the sample is then

$$\overline{L}_\alpha = \frac{1}{T-n}\sum_{t=n}^{T-1} L_\alpha\big(r_{t+1}, \mathrm{VaR}^{t+1}_{1-\alpha}\big).$$

Here $t = n, \ldots, T-1$ indexes the trading days on which we forecast $\mathrm{VaR}^{t+1}_{1-\alpha}$ based on a rolling window of length $n = 500$, and $\mathbf{1}\{r_{t+1} < \mathrm{VaR}^{t+1}_{1-\alpha}\}$ is the indicator that equals 1 if a violation occurs on day $t+1$, and 0 otherwise.

This loss function penalizes prediction errors asymmetrically, assigning a higher penalty when the actual daily log return r_{t+1} falls below the $\mathrm{VaR}^{t+1}_{1-\alpha}$ forecast. With $\alpha \in [0, 1]$, we can equivalently write

$$L^{t+1}_\alpha = \begin{cases} -(1-\alpha)|r_{t+1} - \mathrm{VaR}^{t+1}_{1-\alpha}| & \text{if } r_{t+1} \le \mathrm{VaR}^{t+1}_{1-\alpha}, \\ -\alpha|r_{t+1} - \mathrm{VaR}^{t+1}_{1-\alpha}| & \text{if } r_{t+1} > \mathrm{VaR}^{t+1}_{1-\alpha}. \end{cases}$$

This reflects the higher cost of underestimating risk in financial applications.

A average tick loss closer to zero indicates that the VaR forecasts are more in line with the observed outcomes. Unlike the Kupiec and Christoffersen tests, which assess the frequency and independence of the violations, the average tick loss analyze the magnitude of the forecasting errors.

2.7 Parameter Estimation by Maximum Likelihood

Until now we have mostly covered theoretical modeling. In practice we would like to fit our model to some data, for time series data, this can be done with maximum likelihood. Some basic likelihood theory can be seen in Appendix D.

In Tsay (2013, p. 189), the joint density

$$f(x_t, x_{t-1}, \ldots, x_1)$$

of the sequence $\{X_i\}_{i=1}^t$ can be written recursively as

$$f(x_t, \ldots, x_1) = f(x_t \mid x_{t-1}, \ldots, x_1)\, f(x_{t-1}, \ldots, x_1).$$

Applying this argument repeatedly we eventually obtain

$$f(x_t, \ldots, x_1) = f(x_t \mid x_{t-1}, \ldots, x_1) \cdot f(x_{t-1} \mid x_{t-2}, \ldots, x_1) \cdots f(x_2 \mid x_1) \cdot f(x_1),$$

where $f(x_t \mid x_{t-1}, \ldots, x_1)$ is the conditional density of X_t given all past observations, and $f(x_1)$ is the marginal density of the first observation.

The full log-likelihood for a sample of size T can then be written as

$$\ell(\theta) = \ln f(x_1 \mid \theta) + \sum_{t=2}^{T} \ln f(x_t \mid x_{t-1}, \ldots, x_1; \theta).$$

For example: lets us assume that the process $\{X_t\}$ follows a GARCH(1,1) with Student-t residuals, then $\theta = (\alpha_0, \alpha_1, \beta_1, \nu)$. It follows that

$$X_t \mid \mathcal{F}_{t-1} \overset{d}{=} \sigma_t\, \epsilon_t,$$

with

$$\epsilon_t = \sqrt{\frac{\nu-2}{\nu}}\, Z_t, \quad Z_t \sim t_\nu, \quad \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

where ϵ_t is a standardized Student-t random variable (with zero mean and unit variance) with density function $g(u)$.

The conditional density of X_t can be obtained using the transformation theorem (see Gut (1995, p. 23)). Here, the pdf of ϵ_t is

$$g(u) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\,(\nu-2)}} \left(1 + \tfrac{u^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}.$$

and Jacobian is $\frac{1}{\sigma_t}$, therefore, by the transformation theorem

$$f_{X_t|\mathcal{F}_{t-1}}(x_t) = \frac{1}{\sigma_t}\, g\left(\frac{x_t}{\sigma_t}\right) = \frac{1}{\sigma_t} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\pi\,(\nu-2)}} \left(1 + \tfrac{x_t^2}{(\nu-2)\,\sigma_t^2}\right)^{-\frac{\nu+1}{2}}.$$

The log-likelihood for an observation at time $t > 1$ is therefore

$$\ln f(x_t \mid \mathcal{F}_{t-1}; \theta) = \ln \Gamma\left(\frac{\nu+1}{2}\right) - \ln \Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\ln\left(\pi(\nu-2)\right) - \ln \sigma_t - \frac{\nu+1}{2}\ln\left(1 + \frac{x_t^2}{(\nu-2)\sigma_t^2}\right).$$

The full log-likelihood function for the data $\{x_1, \ldots, x_T\}$ is then

$$\ell(\theta) = \ln f(x_1 \mid \theta) + \sum_{t=2}^{T}\left[\ln \Gamma\left(\frac{\nu+1}{2}\right) - \ln \Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2}\ln\left(\pi(\nu-2)\right) - \ln \sigma_t - \frac{\nu+1}{2}\ln\left(1 + \frac{x_t^2}{(\nu-2)\sigma_t^2}\right)\right].$$

The maximum likelihood estimate $\hat{\theta}_{MLE}$ is obtained by maximizing $l(\theta)$ with respect to θ. If the sample size T is sufficiently large, it is common practice to omit the contribution of $\ln f(x_1|\theta)$ (Tsay, 2013, p. 189).

After obtaining the MLE $\hat{\theta} = (\hat{\alpha}_0, \hat{\alpha}_1, \hat{\beta}_1, \hat{\nu})$, the estimated conditional variance for the GARCH(1,1) is computed recursively as

$$\hat{\sigma}_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1 x_{t-1}^2 + \hat{\beta}_1 \hat{\sigma}_{t-1}^2,$$

with the initial σ_0^2 given by the sample variance of $\{X_t\}$ (Tsay, 2005, p. 116).

3 Data

In this section, we briefly describe the data used in this thesis. Figure 3 shows the full log return series for SEB-A and Swedbank-A from 1 January 2010 to 1 January 2025.

3.1 Software and R packages

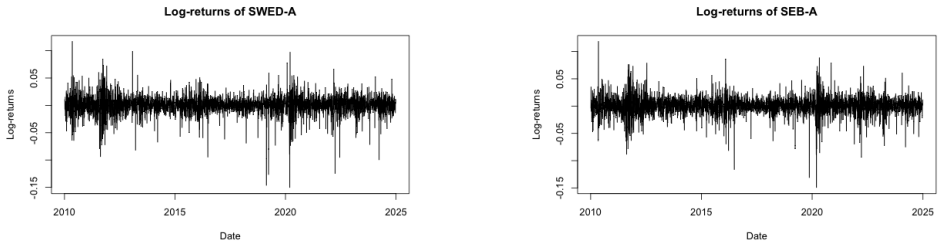
All data retrieval, preprocessing, and time-series handling were performed in R. We used `tidyquant` to fetch daily closing prices for Swedbank-A and SEB-A, then computed log returns as shown in Section 2.1.1. For advanced volatility modeling (GARCH/EGARCH) and backtesting of VaR, we used the `rugarch` package, for model specification, estimation, and forecasting. Because some of the calculations are expensive and time-consuming, we used the R package `doSNOW` to utilize multiple cores and increase efficiency.

3.2 In-sample & out-of-sample

To avoid bias and to assess true forecast performance, we split the full series (2010-01-01 to 2025-01-01) into an in-sample period (2010-01-01 to 2015-01-01) for model calibration and selection, and an out-of-sample period (2015-01-02 to 2025-01-01) for backtesting one-day-ahead VaR forecasts. Based on

the in-sample period, we select and validate our model, i.e., choosing ARMA orders, GARCH specifications and residual diagnostics. The out-of-sample results then show how well those models generalize to unseen data and the quality of the VaR forecasts, this is further explained below.

With the returns and in-/out-of-sample split established, Chapter 4 presents our three VaR methods.



(a) Log returns for Swedbank-A

(b) Log returns for SEB-A

Figure 3: Log returns for Swedbank-A and SEB-A.

4 Modeling

In the following sections we go through, step-by-step, the three different methods used in this thesis for forecasting Value at Risk at a 90% confidence level for a portfolio consisting of the two stocks mentioned above, with equal capital allocation (50% of the total portfolio value invested in each stock). The VaR forecast is compared with the actual realized log return of that day. A correctly specified VaR model at the 90% level ($\alpha = 0.10$) should satisfy

$$\frac{1}{T-n}\sum_{t=n}^{T-1}\mathbf{1}\big(r_{t+1} < \mathrm{VaR}_{0.90}^{t+1}\big) \;\approx\; 0.10,$$

i.e., roughly 10% of observed log returns fall below the forecasted $\mathrm{VaR}_{0.90}^{t+1}$ threshold. Here, T is the total number of observations and $n = 500$ is the length of the rolling window.

First, we begin, in Section 4.1, with a simpler (naive) method, called the Delta-Normal method, where the daily log returns are assumed to be normally distributed. Under this assumption, VaR is computed analytically using the sample mean and standard deviation of the returns.

Next, in Section 4.2, we consider a more sophisticated method by modeling the returns using a GARCH model with Student-t distributed residuals. This model is designed to capture the time-varying volatility in the log returns. In addition, we account for the dependency between the two stocks by using a multivariate t-distribution, which allows for fat tails and jointly extreme events.

Finally, we replace the standard GARCH model with an EGARCH model in order to capture the asymmetric reaction of the market to positive and negative shocks.

When assessing performance for each of the three methods, we use a rolling window of length $n = 500$. That is, for each forecast date $t + 1 = n + 1, \ldots, T$, we:

1. Estimate the model parameters on the subsample $\{r_{t-n+1}, \ldots, r_t\}$.
2. Compute the one-step-ahead VaR forecast $\text{VaR}_{0.90}^{t+1}$.
3. Advance the window by one day (i.e. set $t \leftarrow t + 1$) and repeat.

This procedure produces $T - n$ out-of-sample VaR forecasts, which we then backtest against the realized log returns.

4.1 VaR Modeling - Delta-Normal Method

A simple, parametric (often called Delta-Normal) approach for calculating VaR assumes that log returns are jointly normal (see Jorion 2007, Sec. 10.3). Thus, for each rolling window, we assume the joint distribution of the daily log returns of the two stocks follows a multivariate normal distribution:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_X \\ \mu_Y \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho\,\sigma_X\sigma_Y \\ \rho\,\sigma_X\sigma_Y & \sigma_Y^2 \end{pmatrix} \right).$$

For simplicity of notation, X and Y represent SEB-A and Swedbank-A, respectively.

With equal portfolio weights of 0.5, the portfolios log return is calculated as

$$r_p = 0.5\, r_X + 0.5\, r_Y.$$

Assuming normality, r_p is also normally distributed with mean and variance

$$\mu_p = 0.5\, \mu_X + 0.5\, \mu_Y,$$

$$\sigma_p^2 = (0.5)^2\, \sigma_X^2 + (0.5)^2\, \sigma_Y^2 + 2\,(0.5)(0.5)\,\rho\,\sigma_X\sigma_Y.$$

The one-day-ahead VaR at a 90% confidence level is then computed as the 10th percentile of the distribution of the portfolio's log returns:

$$\text{VaR}_{0.90} = \mu_p + \sigma_p\, z_{0.1},$$

where $z_{0.1}$ is the 0.1 quantile of a standard normal distribution.

The parameters $\mu_X, \mu_Y, \sigma_X^2, \sigma_Y^2$ and ρ are estimated by their sample counterparts. We will use a rolling windows of size $n = 500$, thus, with

$\{x_1, ..., x_{500}\}$ and $\{y_1, ..., y_{500}\}$ as a random samples of X and Y, the sample mean, variance and correlation is estimated as

$$\hat{\mu}_X = \frac{1}{n}\sum_{i=1}^{500} x_i, \qquad \hat{\mu}_Y = \frac{1}{n}\sum_{i=1}^{500} y_i,$$

$$\hat{\sigma}_X^2 = \frac{1}{n-1}\sum_{i=1}^{500}(x_i - \hat{\mu}_X)^2, \qquad \hat{\sigma}_{Y,}^2 = \frac{1}{n-1}\sum_{i=1}^{500}(y_i - \hat{\mu}_Y)^2,$$

$$\hat{\rho}_{XY} = \frac{\sum_{i=1}^{500}(x_i - \hat{\mu}_X)(y_i - \hat{\mu}_Y)}{\sqrt{\sum_{i=1}^{500}(x_i - \hat{\mu}_X)^2}\sqrt{\sum_{i=1}^{500}(y_i - \hat{\mu}_Y)^2}}.$$

These will be re-estimated for each rolling window to ensures that each VaR forecast uses only the most recent $n = 500$ observations.

4.1.1 Model Checking

Based on the in-sample data (i.e. 2010-01-01 to 2015-01-01), we check the normality assumption by comparing each empirical log return distribution to its fitted normal density. The histograms and QQ-plots in Figures 4 and 5 show clear deviations from the straight-line fit in the tails, indicating that the empirical distributions have heavier tails than the normal, which is in line with the discussion in Jorion (2007, p. 262). This suggests the normal model may not fit well, which could lead to inaccurate VaR forecasts.

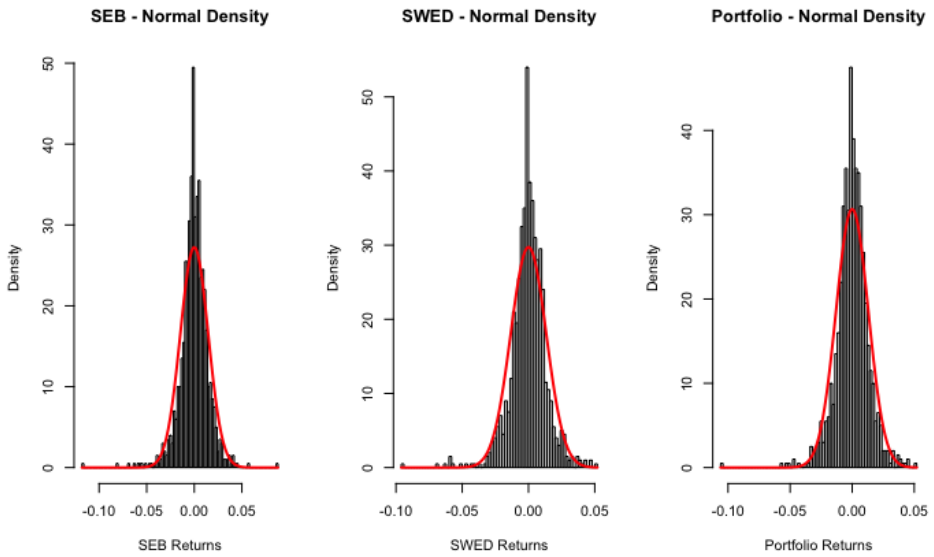


Figure 4: Histograms of empirical log returns vs the theoretical normal density for Swedbank-A, SEB-A, and the 50/50 portfolio.

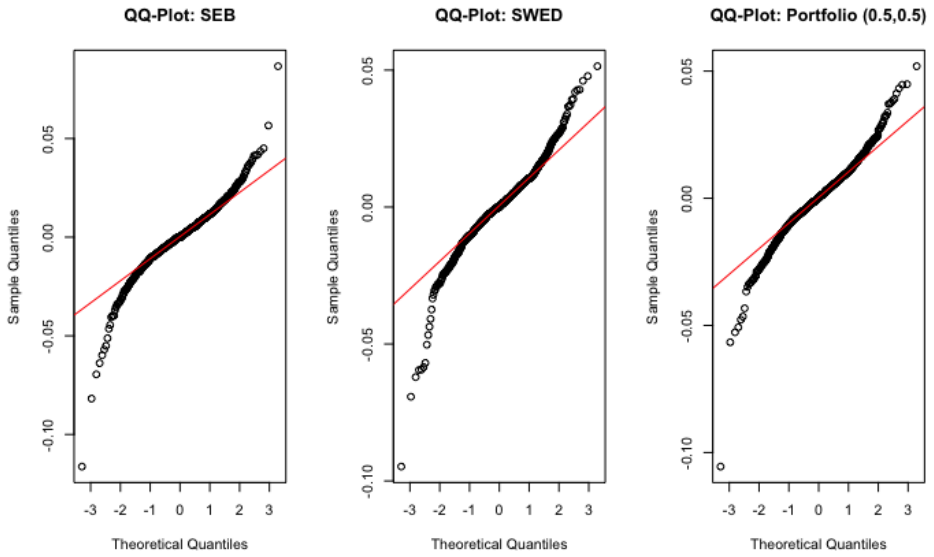


Figure 5: Q-Q plots of empirical log returns against the theoretical normal distribution for Swedbank-A, SEB-A, and the 50/50 portfolio.

4.1.2 Results from Rolling Window Backtesting

The backtesting using a rolling window resulted in approximately 7.1% violations of the VaR forecasts, indicating that the model produced fewer violations than the expected rate of $\alpha = 0.1$. In other words, the model tends to overestimate risk. In later sections, we will conduct statistical tests on the violation sequence to see whether the results are acceptable in terms of both their frequency and clustering of the violations.

We now proceed by building our two more sophisticated VaR models.

4.2 VaR Modelling - GARCH & EGARCH

On the in-sample period (2010-01-01 to 2015-01-01), we first test the series for stationarity and select orders of ARIMA(p, q). We then fit GARCH(p, q) and EGARCH(m, s) models (with Student-t residuals), test for ARCH effects and validate the standardized residuals using QQ-plots and Ljung-Box tests.

With the orders of the models fixed, we use a rolling window on the out-of-sample period (2015-01-02 to 2025-01-01), re-estimating only the model parameters at each step. The one-step-ahead VaR forecasts from each window are then backtested against the realized log returns to assess model performance.

4.2.1 Test for Stationarity

We test for stationarity of the in-sample log return series $\{r_t\}$ for each stock by applying the Augmented Dickey-Fuller (ADF) test directly to $\{r_t\}$. We obtained p-values below 0.01 for both stocks. This provide strong evidence against the null hypothesis of a unit root, indicating that the log return series can be seen as stationary. Stationarity is crucial for further modeling, since both GARCH and EGARCH models assume a stationary time series.

4.2.2 Assessing the Order of the ARMA(p, q) Model

Next we assess the appropriate order of the ARMA model for the mean equation, μ_t. We fit multiple ARMA(p, q) models, using maximum likelihood, and select the one with the lowest Akaike Information Criterion (AIC). For both SEB-A and Swedbank-A, an ARMA(0,0) model provided the lowest score. Thus, the process $\{r_t\}$ can be written as

$$r_t = \mu_t + a_t = \bar{\mu} + a_t$$

where $\bar{\mu}$ is the constant sample mean and $\{a_t\}$ is a white noise process. In other words, the log returns fluctuate randomly around the constant level $\bar{\mu}$, as shown in Figure 3.

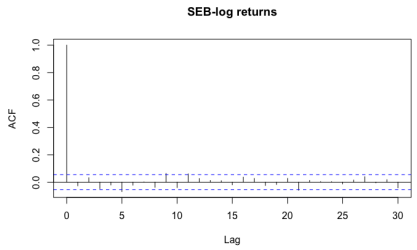
Furthermore, a t-test indicated that the sample means are not significantly different from zero, leading us to assume $\mu_t = 0$ for both stocks. The model thus simplifies to

$$r_t = a_t.$$

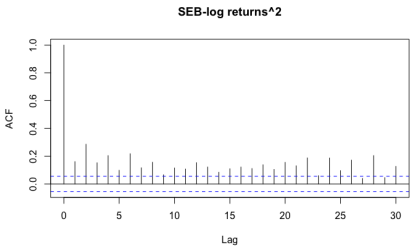
We proceed by examining the series $\{a_t\}$, to see it exhibits any serial dependence.

4.2.3 Checking for ARCH Effects

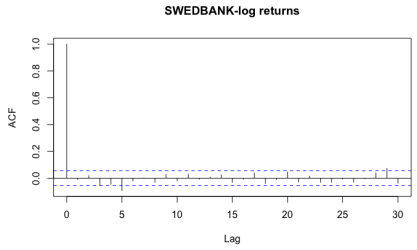
Next, we examine the residual series $\{a_t\}$ for ARCH effects. This is done by plotting the sample autocorrelation function (ACF) of $\{a_t^2\}$ to identify any significant spikes, and by conducting a Ljung-Box test on $\{a_t^2\}$. In Figure 6, we present the ACFs of the original series $\{a_t\}$ and of the squared series $\{a_t^2\}$. The ACFs of $\{a_t^2\}$ show clear evidence of serial autocorrelation for both stocks, indicating that a_t must be modeled using a more sophisticated volatility model. To confirm this, we apply the Ljung-Box directly to the series $\{a_t^2\}$ for each stock. This yields $Q(10) = 20.4$ for Swedbank-A and $Q(10) = 21.03$ for SEB-A. At the 5% significance level, both test statistics exceed the threshold, and we therefore reject the null hypothesis of no serial autocorrelation.



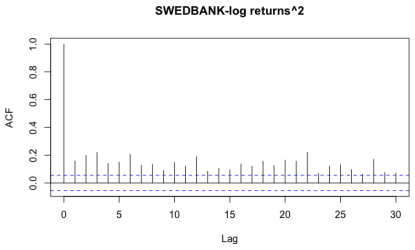
(a) SEB-A: ACF of daily log returns.



(b) SEB-A: ACF of sqr log returns.



(c) Swedbank-A: ACF of daily log returns.



(d) Swedbank-A: ACF of sqr log returns.

Figure 6: Sample autocorrelation functions of the daily log returns and their squares for SEB-A and Swedbank-A.

4.2.4 Assessing the Orders of the GARCH & EGARCH Models

We proceed by fitting GARCH(p, q) and EGARCH(m, s) models to each stock's daily log return series separately using maximum likelihood, were we assume Student-t distributed residuals. Table 1 illustrates the AIC and BIC values for models from order (1,1) to (3,3). For both the GARCH and the EGARCH models, the lowest AIC and BIC values were generally obtained at order (1,1), which is in line with we discussed in Section 2.2.10. It is worth noting that for Swedbank-A, an EGARCH(1,3) model had a lower AIC compared to EGARCH(1,1), but the BIC was lower for the EGARCH(1,1). We therefore choose for the simpler EGARCH(1,1) over the more complex alternative.

As mentioned, we assume for all models that the residuals follow a t-distribution. The validity of this assumption is checked in the next section.

Model	AIC	BIC
GARCH(1,1)	-5.499945	-5.466228
GARCH(1,2)	-5.496244	-5.454098
GARCH(1,3)	-5.492926	-5.442351
GARCH(2,1)	-5.497649	-5.455503
GARCH(2,2)	-5.494372	-5.443796
GARCH(2,3)	-5.493628	-5.434623
GARCH(3,1)	-5.494094	-5.443519
GARCH(3,2)	-5.490400	-5.431395
GARCH(3,3)	-5.489628	-5.422194

(a) SEB-A, GARCH

Model	AIC	BIC
GARCH(1,1)	-5.685633	-5.651916
GARCH(1,2)	-5.681937	-5.639791
GARCH(1,3)	-5.678258	-5.627683
GARCH(2,1)	-5.682995	-5.640849
GARCH(2,2)	-5.681047	-5.630472
GARCH(2,3)	-5.678529	-5.619524
GARCH(3,1)	-5.679207	-5.628631
GARCH(3,2)	-5.677069	-5.618065
GARCH(3,3)	-5.674529	-5.607095

(b) Swedbank-A, GARCH

Model	AIC	BIC
EGARCH(1,1)	-5.518198	-5.476052
EGARCH(1,2)	-5.514612	-5.464037
EGARCH(1,3)	-5.511729	-5.452724
EGARCH(2,1)	-5.514129	-5.455124
EGARCH(2,2)	-5.510862	-5.443428
EGARCH(2,3)	-5.507849	-5.431986
EGARCH(3,1)	-5.508919	-5.433056
EGARCH(3,2)	-5.506004	-5.421712
EGARCH(3,3)	-5.503536	-5.410814

(c) SEB-A, EGARCH

Model	AIC	BIC
EGARCH(1,1)	-5.738633	-5.696487
EGARCH(1,2)	-5.737116	-5.686541
EGARCH(1,3)	-5.743576	-5.684571
EGARCH(2,1)	-5.736433	-5.677429
EGARCH(2,2)	-5.732518	-5.665084
EGARCH(2,3)	-5.736539	-5.660676
EGARCH(3,1)	-5.756135	-5.680272
EGARCH(3,2)	-5.755073	-5.670780
EGARCH(3,3)	-5.750272	-5.657551

(d) Swedbank-A, EGARCH

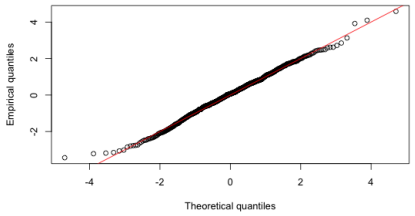
Table 1: Information criteria (AIC and BIC) for GARCH and EGARCH models with Student-t residuals for SEB-A and Swedbank-A.

4.2.5 Standardized Residual Diagnostics

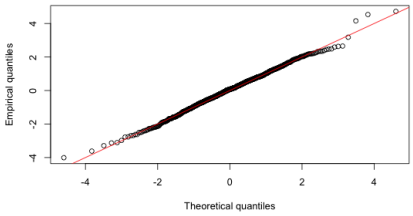
Since both the GARCH(1,1) and EGARCH(1,1) models assume that the standardized residuals $\{\epsilon_t\}$ are i.i.d. Student-t random variables with mean

zero and unit variance, it is important to verify this assumption.

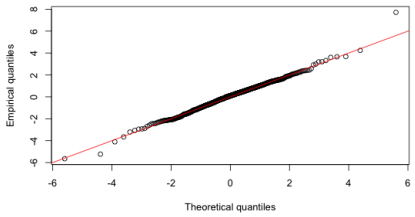
For each fitted model we extract the standardized residuals and compare the empirical quantiles with the theoretical quantiles of a standardized Student-t distribution using QQ-plots, see Figure 7. As we can see, the overall fit is good, except for some extreme observations in the far tails. It's a relatively small sample, so this result could reflect limited data rather than a model misspecification. Also, this does not have to be a major issue, especially since we are interested in forecasting VaR at 90% confidence level. Exploring alternative heavy-tailed distributions could be the done in further research.



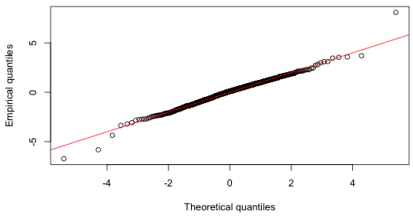
(a) SEB GARCH(1,1)



(b) SEB EGARCH(1,1)



(c) SWEDBANK GARCH(1,1)



(d) SWEDBANK EGARCH(1,1)

Figure 7: QQ-plots: Comparing the empirical quantiles of the standardized residuals from the GARCH$(1,1)$ and EGARCH$(1,1)$ models to a standardized Student-t distribution.

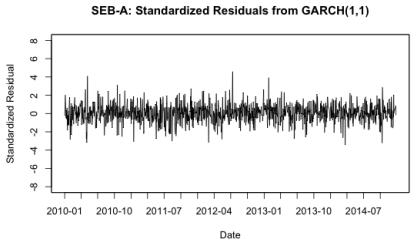
To check for any remaining ARCH effects we first plot the standardized residual series, see Figure 8. They appear to behave like white noise, aside from a few extreme shocks.

In Appendix E (Figure 12), we show the ACFs of both the standardized and squared standardized residuals for each model. There are no extreme spikes, indicating little remaining autocorrelation or volatility clustering. Although, the Ljung-Box test at lag $m = 10$ resulted in low p-values for both the standardized and squared standardized residuals for the SEB-A EGARCH(1,1) model (see Table 2). These rejections are driven by a few small spikes, and the ACF beyond those lags shows no meaningful auto-

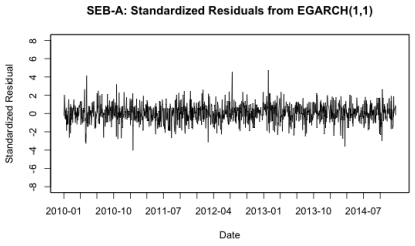
correlations. With this in mind, we treat the residuals as white noise and proceed with our analysis under this assumption.

Model	Std. Residuals (Ljung–Box p-value, lag 10)	Squared Residuals (Ljung–Box p-value, lag 10)
SEB-A GARCH	0.10697	0.06854
SEB-A EGARCH	0.01948	0.02703
Swedbank-A GARCH	0.36316	0.68059
Swedbank-A EGARCH	0.29455	0.70068

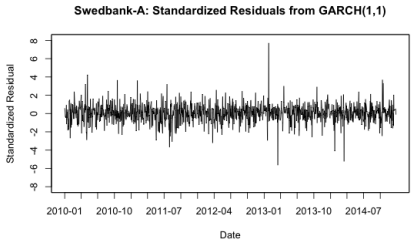
Table 2: Ljung–Box test p-values (lag 10) for standardized residuals and their squares, by model and stock.



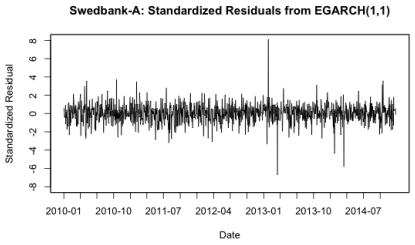
(a) SEB-A: Standardized residuals (GARCH(1,1)).



(b) SEB-A: Standardized residuals (EGARCH(1,1)).



(c) Swedbank-A: Standardized residuals (GARCH(1,1)).



(d) Swedbank-A: Standardized residuals (EGARCH(1,1)).

Figure 8: Standardized residuals series from the GARCH(1,1) and EGARCH(1,1) models for SEB-A and Swedbank-A.

4.2.6 Parameter Estimates for GARCH(1,1) & EGARCH(1,1)

The parameter estimates are showcased is Table 3. Both GARCH(1,1) specifications satisfy the stationarity condition

$$\alpha_1 + \beta_1 < 1$$

(SEB: $0.0653 + 0.9210 = 0.9863$ and Swedbank: $0.0467 + 0.9481 = 0.9948$), ensuring that volatility eventually goes back to a long-run level.

When $\alpha_1 + \beta_1$ is very close to one, nearly all unconditional variance contributions comes from the lag terms, driving α_0 toward zero. As shown in Appendix B, as long as $\alpha_1 + \beta_1 < 1$, the unconditional variance

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}$$

remains well-defined, so a α_0 near zero does not compromise stationarity or model validity.

Moreover, for the EGARCH(1,1) models, all parameters are significant, and the estimated β values are

$$\hat{\beta}_1 = 0.988573 \quad \text{for SEB,} \qquad \hat{\beta}_1 = 0.990335 \quad \text{for Swedbank,}$$

both of which are below 1 and therefore ensure stationarity.

As mentioned in Section 2.1.3, negative movements usually have a larger impact on volatility than positive shocks of the same magnitude. Recall that for an EGARCH(1,1) the conditional volatility σ_t is modeled as

$$\begin{aligned} \ln(\sigma_t^2) &= \omega + \alpha_1 \frac{|a_{t-1}| + \gamma_1 a_{t-1}}{\sigma_{t-1}} + \beta_1 \ln(\sigma_{t-1}^2) \\ &= \omega + \alpha_1 |\epsilon_{t-1}| + \alpha_1 \gamma_1 \epsilon_{t-1} + \beta_1 \ln(\sigma_{t-1}^2), \end{aligned}$$

where $\epsilon_{t-1} = \frac{a_{t-1}}{\sigma_{t-1}}$. The asymmetric effect comes from the term

$$\alpha_1 \gamma_1 \epsilon_{t-1}.$$

Since we obtained $\alpha_1 < 0$ and $\gamma_1 > 0$ for both stocks, for negative shocks ($\epsilon_{t-1} < 0$) it is a positive contribution to $\ln(\sigma_t^2)$, and for positive shocks ($\epsilon_{t-1} > 0$) it contributes a negative value. Therefore, negative shocks inflate volatility more than an equally sized positive shocks. This is in line with what we discussed in Section 2.1.3.

Model	Parameter	Estimate	p-value
GARCH(1,1) SEB	α_0	0.000004	0.28
	α_1	0.065322	<0.01
	β_1	0.920991	<0.01
EGARCH(1,1) SEB	ω	-0.091900	<0.01
	α_1	-0.077542	<0.01
	β_1	0.988573	<0.01
	γ_1	0.077048	<0.01
GARCH(1,1) SWED	α_0	0.000002	0.57
	α_1	0.046699	0.02
	β_1	0.948141	<0.01
EGARCH(1,1) SWED	ω	-0.076927	<0.01
	α_1	-0.079951	<0.01
	β_1	0.990335	<0.01
	γ_1	0.106434	<0.01

Table 3: Parameter Estimates for GARCH(1,1) and EGARCH(1,1) models for SEB-A and Swedbank-A.

4.2.7 Forecasting Volatility

Before forecasting VaR, we compute the one-step-ahead conditional variance. In particular, for each $t = n, \ldots, T-1$, we re-estimate the model parameters by maximum likelihood using the most recent $n = 500$ observations $\{x_{t-n+1}, \ldots, x_t\}$: GARCH$(1,1)$: $\quad \hat{\theta}_t = (\hat{\omega}_t,\ \hat{\alpha}_{1,t},\ \hat{\beta}_{1,t})$,

$$\text{EGARCH}(1,1): \quad \hat{\theta}_t = (\hat{\omega}_t,\ \hat{\alpha}_{1,t},\ \hat{\beta}_{1,t},\ \hat{\gamma}_{1,t}).$$

Then, the one-day-ahead forecast of the conditional variance is computed as in Tsay (2005, p. 115 & 129):

$$\widehat{\sigma}^2_{t+1} = \hat{\alpha}_{0,t} \ + \ \hat{\alpha}_{1,t}\, a_t^2 \ + \ \hat{\beta}_{1,t}\, \hat{\sigma}_t^2 \quad \text{(GARCH(1,1))},$$

$$\ln\big(\widehat{\sigma}^2_{t+1}\big) = \hat{\omega}_t \ + \ \hat{\alpha}_{1,t}\, \frac{|a_t| + \hat{\gamma}_{1,t} a_t}{\hat{\sigma}_t} \ + \ \hat{\beta}_{1,t}\, \ln(\hat{\sigma}_t^2) \quad \text{(EGARCH(1,1))}.$$

Here, $a_t = r_t$ (since we assume that $\mu_t = 0$) and $\hat{\sigma}_t^2$ denotes the estimated conditional variance at time t, and is calculated recursively. After obtaining $\widehat{\sigma}^2_{t+1}$, we roll the window forward by one day and repeat the estimation and forecasting steps. This process yields four sequences of forecasted $\hat{\sigma}^2_{t+1}$, one for each combination of stock (SEB, Swedbank) and method (GARCH, EGARCH).

4.2.8 Forecasting VaR

Having obtained the one-day-ahead conditional variance forecasts, we are finally ready to forecast VaR for the equally weighted portfolio of Swedbank-A and SEB-A.

For both GARCH(1,1) and EGARCH(1,1), we model the daily log returns as

$$r_t = \mu_t + a_t,$$

with $\mu_t = 0$, so $r_t = a_t$. In each rolling window the one-day-ahead log return is

$$r_{t+1} = a_{t+1} = \epsilon_{t+1}\,\hat{\sigma}_{t+1},$$

where ϵ_{t+1} is standardized Student-t distributed and $\hat{\sigma}_{t+1}$ is the forecasted volatility. Hence, conditional on $\mathcal{F}_t$,

$$r_{t+1} \mid \mathcal{F}_t \overset{d}{=} \hat{\sigma}_{t+1}\sqrt{\tfrac{\nu-2}{\nu}}\,Z_{t+1}, \quad Z_{t+1} \sim t_\nu,$$

i.e., a Student-t random variable with ν degrees of freedom, zero mean, and variance $\hat{\sigma}^2_{t+1}$. For the equally weighted portfolio, the portfolios daily log return is written as

$$r_{p,t+1} = 0.5\,r_{X,t+1} + 0.5\,r_{Y,t+1} = 0.5\,\epsilon_{X,t+1}\,\hat{\sigma}_{X,t+1} + 0.5\,\epsilon_{Y,t+1}\,\hat{\sigma}_{Y,t+1},$$

here X and Y represent Swedbank-A and SEB-A, respectively.

We assume that the standardized residuals form a bivariate Student-t distribution, i.e.

$$(\epsilon_{X,t+1}, \epsilon_{Y,t+1})^T \;\sim\; t_\nu\big(\mathbf{0}, \Sigma\big), \qquad \Sigma = \frac{\nu-2}{\nu}\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

with ν degrees of freedom. Under this assumption, any linear combination of $\epsilon_{X,t+1}$ and $\epsilon_{Y,t+1}$ is again t-distribution with ν degrees of freedoms.

For each rolling window, we pragmatically estimate the common ν by

$$\hat{\nu}_t = \frac{\hat{\nu}_{X,t} + \hat{\nu}_{Y,t}}{2},$$

where $\hat{\nu}_{X,t}$ and $\hat{\nu}_{Y,t}$ are the maximum likelihood estimates of the degrees of freedom obtained from the series of the standardized residuals $\{\epsilon_{X,t}\}$ and $\{\epsilon_{Y,t}\}$, respectively. Following the steps in Appendix C, $r_{p,t+1}$ is a scaled standard Student-t. The estimated 0.1 quantile of the one-day-ahead portfolio can be computed as

$$\hat{q}_{0.1}(r_{p,t+1}) = \sqrt{\frac{\hat{\nu}-2}{\hat{\nu}}(C_1^2 + C_2^2 + 2\,C_1\,C_2\,\hat{\rho}_t)}\cdot t^{-1}(0.1, \hat{\nu}_t),$$

where

$$C_1 = 0.5\,\hat{\sigma}_{X,t+1}, \quad C_2 = 0.5\,\hat{\sigma}_{Y,t+1}.$$

Here, $\hat{\rho}_t$ is the sample correlation of the two series of standardized residuals in the rolling windows ending at day t and $t^{-1}(0.1, \hat{\nu}_t)$ is the 0.1 quantile of the standard t-distribution with $\hat{\nu}_t$ degrees of freedom. This quantile represents the VaR forecast for each estimation period, i.e.

$$\text{VaR}_{0.90}^{t+1} = \hat{q}_{0.1}(r_{p,t+1}).$$

This procedure is repeated for each of the rolling windows of length $n = 500$, yielding the sequence of VaR forcasts:

$$\left\{\text{VaR}_{0.90}^{t+1}\right\}_{i=n}^{T-1} = \{\text{VaR}_{0.90}^{501}, \ldots, \text{VaR}_{0.90}^{T}\},$$

where $t+1$ is the day on which the forecast in made and T the size of the sample. This sequence of VaR forecasts is then compared to the realized portfolio log returns by examining the resulting sequence of violations:

$$I_{t+1} = \mathbf{1}\{r_{p,t+1} \leq \text{VaR}_{0.9}^{t+1}\}, \quad t = n, \ldots, T-1.$$

In the following section, we analyze these backtesting results in detail.

5 Results

In this section we present the empirical performance of our three VaR methods. We begin with the accuracy of the volatility forecasts, measured by the Root Mean Squared Error (RMSE). We then compare the VaR forecasts with the realized log returns using the tick loss average. Finally, we present the outcomes of our VaR backtests, including Kupiec and Christoffersen tests. All performance metrics below are computed over $T - n$ number of rolling windows of length $n = 500$ trading days covering the out-of-sample period 2015-01-01 to 2025-01-01, yielding $T - n = 2015$ forecasting days.

5.1 RMSE

As stated above, we use rolling windows of length $n = 500$ over a total sample of size T, yielding $T - n$ one-day-ahead forecasts. For each window ending at time t ($t = n, \ldots, T-1$), we compute the forecasted portfolio variance $\hat{\sigma}_{p,t+1}^2$ in two ways:

1. **Delta-Normal:** Forecast the variances $\hat{\sigma}_{X,t+1}^2$ and $\hat{\sigma}_{Y,t+1}^2$ of Swedbank-A and SEB-A log returns, as the sample variance on the windows $\{r_{t-n+1}, \ldots, r_t\}$, and their sample correlation $\hat{\rho}_t$. Then

$$\hat{\sigma}_{p,t+1}^2 = 0.25\,\hat{\sigma}_{X,t+1}^2 + 0.25\,\hat{\sigma}_{Y,t+1}^2 + 2 \cdot 0.5 \cdot 0.5\,\hat{\rho}_t\,\hat{\sigma}_{X,t+1}\,\hat{\sigma}_{Y,t+1}.$$

2. **GARCH/EGARCH:** With the one-day-ahead model-based volatility forecasts $\hat{\sigma}_{X,t+1}$ and $\hat{\sigma}_{Yt+1}$, and the sample correlations $\hat{\rho}_t$ of the two series of standardized residuals for each window, calculate

$$\hat{\sigma}_{p,t+1}^2 = 0.25\,\hat{\sigma}_{X,t+1}^2 + 0.25\,\hat{\sigma}_{Y,t+1}^2 + 2 \cdot 0.5 \cdot 0.5\,\hat{\rho}_t\,\hat{\sigma}_{X,t+1}\,\hat{\sigma}_{Y,t+1}.$$

Here, X and Y represent Swedbank-A and SEB-A, respectively. We then compare each $\{\hat{\sigma}^2_{p,t+1}\}_{t=n}^{T-1}$ to the realized proxy variances $\{r^2_{p,t+1}\}_{t=n}^{T-1}$, where

$$r_{p,t+1} = 0.5\, r_{X,t+1} + 0.5\, r_{Y,t+1}.$$

The RMSE across all $T-n$ windows is

$$\text{RMSE} = \sqrt{\frac{1}{T-n}\sum_{t=n}^{T-1}\left(\hat{\sigma}^2_{p,t+1} - r^2_{p,t+1}\right)^2}.$$

The results are presented in Table 4. The EGARCH(1,1) model obtained the lowest RMSE, indicating the best performance in capturing time-varying volatility, while the Delta-Normal method which assumes constant volatility, performed worst. However, the differences in RMSE are very small, making it difficult to draw definitive conclusions about relative forecasting accuracy from these values alone.

Method	RMSE
Delta-Normal	0.0008072
GARCH(1,1)	0.0007877
EGARCH(1,1)	0.0007814

Table 4: RMSE values for the Delta-Normal, GARCH, and EGARCH model.

5.2 Tick Loss

We now proceed to evaluate model performance using the average tick loss.

Let VaR^m_{t+1} denote the one-day-ahead VaR forecast from method $m \in \{\text{Delta, GARCH, EGARCH}\}$ based on the window ending at t (for $t = n, \ldots, T-1$). We compute the tick loss at each step as

$$L^m_{t+1} = \left(\mathbf{1}\{r_{t+1} < \text{VaR}^{t+1}_{1-\alpha}\} - \alpha\right)\left(r_{t+1} - \text{VaR}^{t+1}_{1-\alpha}\right), \quad t = n, \ldots, T-1.$$

With the sequence $\{L^m_{t+1}\}_{t=n}^{T-1}$, we calculate the average tick loss for each method as

$$\overline{L}^m = \frac{1}{T-n}\sum_{t=n}^{T-1} L^m_{t+1}.$$

Table 5 reports the average tick loss of the one-day-ahead VaR forecasts for each method over the backtesting period. An average tick loss closer to zero indicates better alignment between VaR forcasts and realized log returns.

Overall, the three methods yield very similar tick loss values, the ranking is: EGARCH performs best, closely followed by GARCH, while Delta-Normal slightly behind. The differences are small in absolute terms, but they

confirm that modeling time-varying volatility yields more accurate VaR forecasts than a static, and the not as flexible, normal approach.

Metod	Tick Loss
Delta-Normal	-0.00307
GARCH(1,1)	-0.00288
EGARCH(1,1)	-0.00286

Table 5: Tick Loss for each VaR model.

5.3 Violation Rate

Each of the three methods resulted in different violation rates. Table 6 shows these rates for all three models. Figure 9 shows the evolution of the cumulative violation rates over the rolling windows. As the figure illustrates, the rates appears to have converged.

The EGARCH model produces a violation rate closest to the target 10% level, indicating the most accurate risk estimation. The GARCH model had intermediate performance, whereas the Delta-Normal method significantly overestimates risk, resulting in a violation rate that deviates substantially from the expected 10% target.

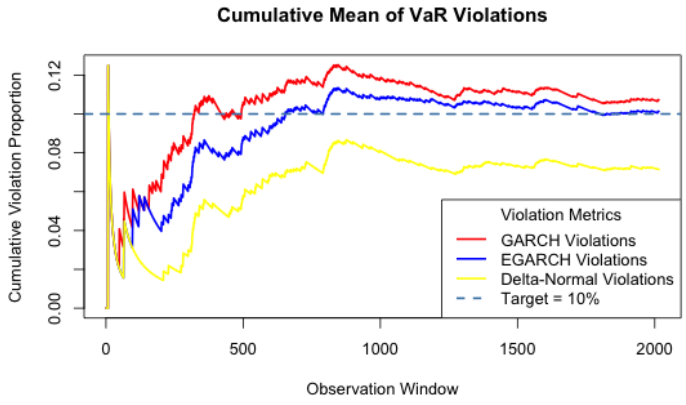


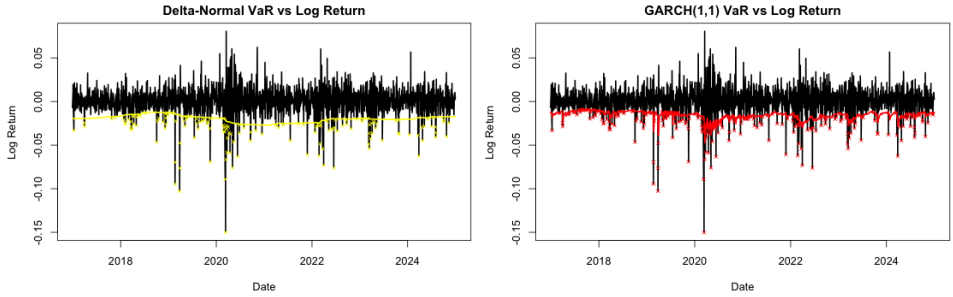
Figure 9: The Cumulative Violation Rate for each Model.

5.4 Evaluating Violation Sequences

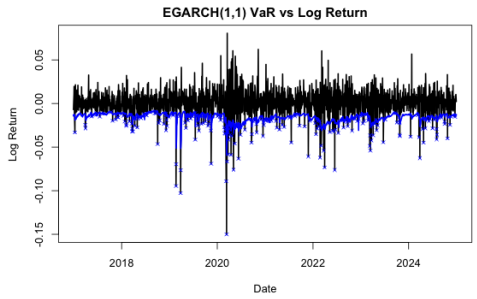
We apply two tests to the VaR violation sequences, the Kupiec-test and the conditional coverage test (or Christoffersen test). The Kupiec-test evaluates whether the total number of violations matches the expected rate, while the conditional coverage test jointly tests both the violation frequency and the independence of violations over time. The p-values for each test are presented in Table 6.

For the Kupiec test, the GARCH and EGARCH models had p-values of 0.287 and 0.853, respectively, hence, we do not reject the null hypothesis of a correct violation rate. The Delta-Normal model, however, fails the test with a p-value below 0.001, indicating significant overestimation of risk.

For the conditional coverage test, both the GARCH and EGARCH models pass with p-values of 0.387 and 0.713, respectively, indicating that these models capture both the frequency and independence of violations. The Delta-Normal model, however, is again rejected, confirming its inability to account for time-varying volatility. This pattern is also evident in Figure 10, where EGARCH and GARCH adjust effectively to periods of high and low volatility, while the Delta-Normal model struggles to do so.



(a) Delta-Normal VaR (b) GARCH(1,1) VaR



(c) EGARCH(1,1) VaR

Figure 10: One-day-ahead VaR forecasts vs actual log returns. Crosses mark violations.

Model	**Violation Rate**	**Kupiec p-value**	**CC p-value**
GARCH(1,1)	0.107	0.287	0.387
EGARCH(1,1)	0.101	0.853	0.713
Delta-Normal	0.0715	<0.001	<0.001

Table 6: Violation rate, Kupiec and Christoffersen p-values.

6 Discussion & Conclusions

6.1 Key Findings

This thesis compare three VaR forecasting methods on a portfolio of SEB-A and Swedbank-A over 2015–2025. Our backtests of one-day-ahead VaR forecasts using a rolling window reveal a clear performance ranking: EGARCH(1,1) at the top, followed by GARCH(1,1), and lastly Delta-Normal. The Delta-Normal method produced a 7% violation rate, less than the target of 10%. By contrast, both GARCH and EGARCH maintained violation rates within the acceptable boundaries in terms of the Kuipec and Christoffersen tests. These findings demonstrate the importance of modeling time-varying and asymmetric volatility for accurate tail-risk estimation.

6.2 Why Delta-Normal Underperform

A 7% violation rate may appear conservative, it could reflects a misfit of the normal tail and the true log return distribution. Under the normality assumption, the one-day-ahead 90% VaR is given by

$$\mathrm{VaR}_{0.90} = \mu_p + z_{0.10}\,\sigma_p, \quad z_{0.10} \approx -1.2816.$$

However, in Section 4.1.1 we saw that the empirical log return distributions had fatter tails than a normal. Fat tails inflate the sample variance under the normal assumption, pushing $\mu_p - 1.2816\,\sigma_p$ too far into the left tail and thereby reducing the observed violation rate to only 7 %.

6.3 Limitations and Improvements

Our VaR modeling provides valuable insights, some limitations lead to potential improvements, we will go through some of the here.

The choice of a 500-day window involves a trade-off between reactivity and stability. In particular, because the Delta-Normal method uses the sample variance within each window, this quantity updates more slowly when the window is larger. Testing alternative lengths (e.g. 250 or 750 days) would reveal how sensitive our results are to this parameter. Future work should compare these alternatives.

Furthermore, our focus on two large-cap Swedish bank stocks may limits generalizability to broader portfolios, such as small-cap or more volatile tech-stocks. Extending the analysis, including additional asset types, would test the generality.

In this thesis we use a 90% confidence level for VaR. If we instead applied a 99% level, we would expect about

$$0.01 \cdot 2015 \approx 20$$

number of violations, and a deviation of one or two violations could be from sampling variation alone, making the test statistics very sensitive. By choosing 90%, we ensure a sufficient number of violations for a more reliable testing. For future work, extending the historical data sample (e.g. using more years of log returns) would allow us to use of higher confidence levels (such as 99%) without sacrificing statistical power.

Finally, our backtest uses Christoffersen's one-day independence test. More complex alternatives, such as duration-based tests could be included. These go beyond Christoffersen's test by looking at a longer dependence patterns than one day.

7 Extension

In the previous analysis we assumed that the two standardized residuals $\epsilon_{X,t}$ and $\epsilon_{Y,t}$ from each stocks returns series shared a common degrees of freedom parameter ν. The maximum likelihood estimates of the degrees of freedom ν over the in-sample period (2010-01-01 to 2015-01-01) are

$$\begin{aligned} \text{GARCH(1,1):} \quad & \hat{\nu}_{\text{Swedbank-A}} = 4.9, \quad \hat{\nu}_{\text{SEB-A}} = 7.2, \\ \text{EGARCH(1,1):} \quad & \hat{\nu}_{\text{Swedbank-A}} = 5.1, \quad \hat{\nu}_{\text{SEB-A}} = 7.6. \end{aligned}$$

These are noticeably different indicate that the assumption of a common ν for both series based on in-sample fitting is quite strong and not fully supported by the data. While convenient, this "common-ν" assumption is somewhat restrictive.

We will now extend the method-framework from the previous sections, by using a copula-based VaR method to capture joint tail dependence more flexibly than a multivariate-t distribution (which forces both residuals to share a common degree of freedom). This flexible approach would be to model the joint distribution of $(\epsilon_{X,t}, \epsilon_{Y,t})$ via a copula. In particular, we could:

1. Estimate GARCH/EGARCH models for each stock and extract the standardized residuals.
2. Apply an inverse transform to those residuals and fit a copula (e.g. Student-t) to model their dependence structure.
3. Jointly simulate from the fitted copula, back-transform through the marginal t-distributions, aggregate into portfolio log returns, and extract the 90 % VaR.
4. Conduct test as described in Chapter 5.

This extension trades the bivariate-t approach in Section 4.2.8 for greater flexibility modeling. Full theoretical background is not included in this thesis,

instead, this should be viewed as a blueprint for potential future work. We will first provide a theoretical overview of the extension and then implement it in practice, with empirical results presented thereafter.

7.1 Theoretical Background on Copulas

A copula is a function that "couples" together the marginal distributions of two (or more) random variables into a joint distribution, by working on their transformed ranks rather than their original scales. This lets us model dependence separately from the choice of marginals.

This theoretical section is based on the lecture notes Lindskog (2025).

Let X and Y be random variables with continuous marginal CDFs $F_X(x)$ and $F_Y(y)$. Define

$$U = F_X(X), \qquad V = F_Y(Y).$$

By the inverse transform, U and V each uniformly distributed on $[0, 1]$. A copula C is then the joint CDF of (U, V):

$$C(u, v) = P\big(U \leq u,\ V \leq v\big), \quad (u, v) \in [0, 1]^2.$$

Sklar's theorem (Nelsen, 2006, p. 18) ensures that the original joint distribution $H(x, y) = P(X \leq x, Y \leq y)$ can be written uniquely as

$$H(x, y) = C\big(F_X(x),\ F_Y(y)\big),$$

if both F_X and F_Y is continuous.

7.1.1 Proof of the inverse transform theorem (sketch)

Since F_X is continuous and strictly increasing, for any $u \in [0, 1]$:

$$P(U \leq u) = P\big(F_X(X) \leq u\big) = P\big(X \leq F_X^{-1}(u)\big) = F_X\big(F_X^{-1}(u)\big) = u.$$

Hence $U \sim \text{Uniform}(0, 1)$, and similarly $V \sim \text{Uniform}(0, 1)$.

This motivates using (U, V) as the building blocks for the copula C.

7.1.2 Sampling from a Copula

Given two samples $\{x_i\}_{i=1}^n$ and $\{y_i\}_{i=1}^n$, transform them to "pseudo-observations" on $[0, 1]$ by

$$u_i = F_X(x_i), \qquad v_i = F_Y(y_i),$$

Because the CDF transform preserves ranks, the pair (u_i, v_i) have the dependence structure of (x_i, y_i) in terms of their ranks. We then fit a parametric copula, for example a Student-t copula, to the (u_i, v_i) data by maximum likelihood.

To simulate new joint pair, we

1. sample (u^*, v^*) from the fitted copula,

2. convert back to the original scale via the inverse marginals:

$$x^* = F_X^{-1}(u^*), \quad y^* = F_Y^{-1}(v^*).$$

This method let us sample "synthetic" (x^*, y^*) pairs that preserve the marginal distributions and the ranked dependence captured by the copula.

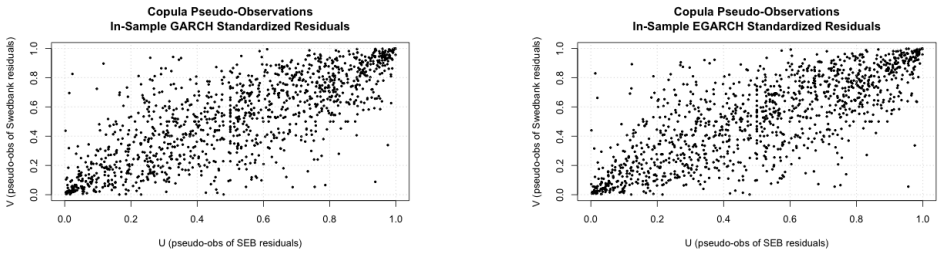
With that said, we are know ready to implement this theoretical framework into our VaR modeling.

7.2 VaR Modeling - GARCH/EGARCH-Copula

This method replaces the multivariate Student-t distribution used in Section 4.2.8 while leaving all earlier model steps unchanged. For each rolling window, we first fit a standardized Student-t distribution to each stock's series of standardized residuals, estimating each series degrees of freedom by maximum likelihood.

Next, we transform each residual series into pseudo-observations with its fitted Student-t CDF, as described in the previous section, and then fit a copula to these pseudo-observations to capture the dependence structure.

This procedure is done on the in-sample data (2010-01-01 to 2015-01-01). In Figure 11, we see the scatterplot of the pseudo-observation. During normal periods, when fluctuations are small, the two stocks has only weak dependence. However, in more turbulent market conditions they tend to move together. This behavior is intuitive, on an average day each stock follows its own drivers, but when a major shock occurs it impacts both stocks simultaneously. This explains the clustering in the corners, and sparsity in the middle section.



(a) GARCH(1,1) (b) EGARCH(1,1)

Figure 11: Copula pseudo-observations for in-sample standardized residuals of the two volatility models. Each point (u, v) represents the uniform transformed residuals of SEB-A and Swedbank-A.

7.2.1 Model Selection - Copula

Based on the in-sample data (2010-01-01 to 2015-01-01) we fit four different copulas by maximum likelihood, Student-t, Gaussian, Frank, and Gumbel, and select the one with the lowest AIC. This resulted in selecting the t-copula, since it yield the lowest AIC for both GARCH(1,1) and EGARCH(1,1), see Table 7.

Copula	GARCH AIC	EGARCH AIC
Normal	-1209.265	-1192.387
Student-t	-1278.090	-1256.807
Frank	-1159.567	-1148.934
Gumbel	-1200.708	-1183.636

Table 7: AIC values for copulas fitted to GARCH(1,1) and EGARCH(1,1) standardized residuals.

7.2.2 Forecasting VaR - GARCH/EGARCH-Copula

The one-day-ahead VaR forecast for each rolling window is obtained as follows:

1. Fit the t-copula to the pseudo-observations of $\epsilon_{X,t}$ and $\epsilon_{Y,t}$.
2. Simulate $N = 10\ 000$ pairs $(\epsilon_{X,i}, \epsilon_{Y,i})$ from this copula and convert back to the original scale.
3. Forecast one-day-ahead volatilities $\hat{\sigma}_{X,t+1}$ and $\hat{\sigma}_{Y,t+1}$ from the GARCH/EGARCH model.
4. Generate simulated returns:
$$r^{(i)}_{X,t+1} = \hat{\sigma}_{X,t+1}\,\epsilon_{X,i}, \quad r^{(i)}_{Y,t+1} = \hat{\sigma}_{Y,t+1}\,\epsilon_{Y,i}, \quad i = 1, \ldots, 10\ 000.$$
5. Compute portfolio returns with equal weights:
$$r^{(i)}_{p,t+1} = 0.5\, r^{(i)}_{X,t+1} + 0.5\, r^{(i)}_{Y,t+1}, \quad i = 1, \ldots, 10\ 000.$$
6. Estimate the 0.1 quantile of $\{r^{(i)}_{p,t+1}\}$ as the $VaR^{t+1}_{0.90}$ forecast, and compare to the realized $r_{p,t+1}$ to keep track of the violations.

This procedure yields a violation sequence just as before, allowing us to apply the same diagnostic tests (Kupeic and Christoffersen's test) to assess the VaR forecasts.

7.3 Results - Extension

In Table 8, we summarize the diagnostic results for the violation sequences when using the EGARCH(1,1) and GARCH(1,1)-copula model. Overall, both models perform well, with significant p-values for both the Kupiec and Christoffersen tests and the violation rates are very close to the target level of 10%.

The GARCH(1,1)-copula model outperforms the GARCH(1,1)-t (we denote EGARCH/GARCH-t to mean EGARCH/GARCH(1,1) with Student-t residuals) model from the previous chapters, its Kupiec p-value increases from 0.287 to 0.484, and its Christoffersen p-value increases from 0.387 to 0.622, indicating more significant violation rate and less clustering in the violations. The EGARCH(1,1)-copula model shows similar results as for the EGARCH(1,1)-t model, its Kupiec p-value remains virtually unchanged, as does its Christoffersen p-value.

We also compared RMSE (computed by replacing the one-day-ahead conditional variance $\hat{\sigma}^2_{p,t+1}$ with the sample variance of each set of simulated portfolio returns), and average tick loss for the copula-based models against their GARCH(1,1)-t and EGARCH(1,1)-t counterparts. We obtained

$$\begin{aligned} \text{GARCH(1,1)-copula:} \quad & \text{tick loss} = 0.0028861, \quad \text{RMSE} = 0.0007917, \\ \text{EGARCH(1,1)-copula:} \quad & \text{tick loss} = 0.0028596, \quad \text{RMSE} = 0.0007815. \end{aligned}$$

These results are almost identical to those from the GARCH(1,1)-t and EGARCH(1,1)-t models. This shows that adding a copula for the standardized residuals has a minor impact on both volatility and VaR forecast accuracy.

Model	**Violation rate**	**Kupiec p-value**	**Christoffersen p-value**
GARCH(1,1)-t	0.107	0.287	0.387
EGARCH(1,1)-t	0.101	0.853	0.713
GARCH(1,1)-Copula	0.105	0.484	0.622
EGARCH(1,1)-Copula	0.0988	0.853	0.704

Table 8: Violation rate and Kupiec and Christoffersen p-values for each VaR model.

As shown in Table 9, the observed number of VaR violations (out of 2015 trails) for the GARCH(1,1)-t, EGARCH(1,1)-t, GARCH(1,1)-copula, and EGARCH(1,1)-copula models are 216, 204, 212, and 199, respectively. These numbers lie relatively close to the expected value of 201.5, indicating that switching from the t-model to a copula-based approach has only a marginal impact on the violation rate. Because each test is sensitive to even a single additional exception, these small differences do not justify strong conclusions

about one model's superiority over the others. A more rubost analysis would be to include more data, do reduce the sensitivity in these tests.

In conclusion, the simpler EGARCH(1,1)-t models with bivariate Student-t residuals works just as well, for this data, as the more complex models and is easier to implement. If we extend this to a portfolio with more stocks, the copula model becomes more flexible, since it does not require assuming the same degrees of freedom for each standardized residual series.

Model	**Observed Violations** (expected = 201.5)
Delta-Normal	144
GARCH(1,1)-t	216
EGARCH(1,1)-t	204
GARCH(1,1)-copula	202
EGARCH(1,1)-copula	199

Table 9: Observed number of violations (out of 2015 observations) for each VaR models, where the expected number is $2015 \cdot 0.1 = 201.5$.

References

T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327, 1986.

T. Britton and S. E. Alm. *Stokastik: Sannolikhetsteori och Statistikteori med Tillämpningar*. Liber, 2014.

P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer, 1991.

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

R. F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007, 1982.

T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.

A. Gut. *An Intermediate Course in Probability*. Springer, 1995.

P. R. Hansen and A. Lunde. A forecast comparison of volatility models: Does anything beat a garch(1,1)? *Journal of Applied Econometrics*, 20 (7):873–889, 2005.

L. Held and D. Bové. *Likelihood and Bayesian Inference*. Springer, 2nd edition, 2020.

A. H. Joarder and B. M. G. Kibria. A short review of multivariate t-distribution. *Journal of Statistical Research*, 40(1):59–72, 2006.

P. Jorion. *Value at Risk: The New Benchmark for Managing Financial Risk*. McGraw-Hill, 3rd edition, 2007.

P. Kupiec. Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives*, 3(2):73–84, 1995.

F. Lindskog. Lecture notes: Risk models, claims reserving and solvency in non-life insurance, 2025. Unpublished lecture notes.

R. Nelsen. *An Introduction to Copulas*. Springer, 2006.

D. B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370, 1991.

R. S. Tsay. *Analysis of Financial Time Series*. Wiley, 2nd edition, 2005.

R. S. Tsay. *An Introduction to Analysis of Financial Data with R.* Wiley, 2013.

A One-Sample t-Test

Let $X_1, \ldots, X_n$ be an independent and identically distributed sample from the random variable X, with true mean $E[X]$ and variance $\mathrm{Var}(X)$. A one-sample t-test can be used to assess whether the sample mean differs significantly from a specified value (e.g. zero), as discussed in Section 2.2.6. Following Britton and Alm (2014, p. 334), we compute the test statistic

$$T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}},$$

which under the null hypothesis

$$H_0 \colon E[X] = \mu_0$$

follows a Student's t-distribution with $n - 1$ degrees of freedom. The alternative hypothesis is two-sided:

$$H_1 \colon E[X] \neq \mu_0.$$

Here, $\bar{x}$ is the sample mean, n denotes the sample size and s the sample standard deviation.

B Properties of ARCH(1) and GARCH(1,1)

Here we outline some basic properties for the ARCH(1) and GARCH(1,1) models. Similar calculations can be seen in Tsay (2005, p. 105).

Properties of the ARCH(1)

Recall that in the ARCH(1) model we have

$$a_t = \sigma_t \, \epsilon_t, \qquad \sigma_t^2 = \alpha_0 + \alpha_1 \, a_{t-1}^2,$$

where $\{\epsilon_t\}$ is an i.i.d. sequence with $E[\epsilon_t] = 0$, $E[\epsilon_t^2] = 1$, and parameters satisfy $\alpha_0 > 0$, $\alpha_1 \geq 0$ and $\alpha_1 < 1$ for stationarity. We will now explore some of its properties.

Unconditional Mean:

$$\mathrm{E}[a_t] = \mathrm{E}\left[\epsilon_t \sqrt{\alpha_0 + \alpha_1 a_{t-1}^2}\right] = \sqrt{\alpha_0 + \alpha_1 a_{t-1}^2} \cdot \mathrm{E}[\epsilon_t] = 0,$$

since $\mathrm{E}[\epsilon_t] = 0$.

Conditional Mean:

$$\mathrm{E}[a_t \mid \mathcal{F}_{t-1}] = \mathrm{E}\left[\epsilon_t \sqrt{\alpha_0 + \alpha_1 a_{t-1}^2} \mid \mathcal{F}_{t-1}\right] = \sqrt{\alpha_0 + \alpha_1 a_{t-1}^2} \cdot \mathrm{E}[\epsilon_t \mid \mathcal{F}_{t-1}] = 0,$$

where $\mathcal{F}_{t-1}$ represents all the observations up to time $t-1$, and the last equality follows from the independence of ϵ_t from $\mathcal{F}_{t-1}$.

Conditional Variance:

The conditional variance is defined as

$$\sigma_t^2 = \mathrm{Var}[a_t \mid \mathcal{F}_{t-1}] = \mathrm{E}\left[(a_t - \mathrm{E}[a_t \mid \mathcal{F}_{t-1}])^2 \mid \mathcal{F}_{t-1}\right].$$

Given that $\mathrm{E}[a_t \mid \mathcal{F}_{t-1}] = 0$, we have

$$\sigma_t^2 = \mathrm{E}[a_t^2 \mid \mathcal{F}_{t-1}] = \mathrm{E}\left[\epsilon_t^2(\alpha_0 + \alpha_1 a_{t-1}^2) \mid \mathcal{F}_{t-1}\right] = (\alpha_0 + \alpha_1 a_{t-1}^2) \cdot \mathrm{E}\left[\epsilon_t^2 \mid \mathcal{F}_{t-1}\right].$$

Since $\mathrm{E}\left[\epsilon_t^2 \mid \mathcal{F}_{t-1}\right] = 1$, it follows that

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

This shows that the conditional variance is time-varying.

Unconditional Variance:

Taking expectations on both sides of the equation for σ_t^2, we get

$$\mathrm{E}[\sigma_t^2] = \mathrm{E}[\alpha_0 + \alpha_1 a_{t-1}^2] = \alpha_0 + \alpha_1 \mathrm{E}[a_{t-1}^2].$$

Since the process is stationary, $\mathrm{E}[a_{t-1}^2] = \mathrm{E}[\sigma_{t-1}^2] = \sigma^2$. We have

$$\sigma^2 = \alpha_0 + \alpha_1 \sigma^2.$$

Solving for σ^2:

$$\sigma^2 = \frac{\alpha_0}{1 - \alpha_1}.$$

While the conditional variance σ_t^2 changes over time, the unconditional variance is constant.

Properties of the GARCH(1,1)

Recall that in the GARCH(1,1) model we write

$$a_t = \sigma_t \, \epsilon_t, \qquad \sigma_t^2 = \alpha_0 + \alpha_1 \, a_{t-1}^2 + \beta_1 \, \sigma_{t-1}^2,$$

where $\{\epsilon_t\}$ is i.i.d. with $E[\epsilon_t] = 0$, $E[\epsilon_t^2] = 1$, and $\alpha_0 > 0$, $\alpha_1, \beta_1 \geq 0$ are such that $\alpha_1 + \beta_1 < 1$ for stationarity. Similar for the ARCH(1), we will explore some of its properties.

Unconditional Mean:

$$\mathrm{E}[a_t] = \mathrm{E}\left[\sigma_t \epsilon_t\right] = \mathrm{E}[\sigma_t] \cdot \mathrm{E}[\epsilon_t] = \mathrm{E}[\sigma_t] \cdot 0 = 0,$$

since $\mathrm{E}[\epsilon_t] = 0$.

Conditional Mean:

$$\mathrm{E}[a_t \mid \mathcal{F}_{t-1}] = \mathrm{E}\left[\sigma_t \epsilon_t \mid \mathcal{F}_{t-1}\right] = \sigma_t \cdot \mathrm{E}[\epsilon_t \mid \mathcal{F}_{t-1}] = \sigma_t \cdot 0 = 0,$$

Conditional Variance

The variance of a_t given the information, $\mathcal{F}_{t-1}$, is defined as

$$\mathrm{Var}[a_t \mid \mathcal{F}_{t-1}] = \mathrm{E}\left[(a_t - \mathrm{E}[a_t \mid \mathcal{F}_{t-1}])^2 \mid \mathcal{F}_{t-1}\right].$$

Since $\mathrm{E}[a_t \mid \mathcal{F}_{t-1}] = 0$, we have

$$\mathrm{Var}[a_t \mid \mathcal{F}_{t-1}] = \mathrm{E}\left[a_t^2 \mid \mathcal{F}_{t-1}\right].$$

Substituting $a_t = \sigma_t \epsilon_t$, and noting that σ_t is known given $\mathcal{F}_{t-1}$ since

$$\alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

only depends on the values at time $t-1$, thus, we obtain

$$\mathrm{Var}[a_t \mid \mathcal{F}_{t-1}] = \mathrm{E}\left[\sigma_t^2 \epsilon_t^2 \mid \mathcal{F}_{t-1}\right] = \sigma_t^2 \,\mathrm{E}\left[\epsilon_t^2 \mid \mathcal{F}_{t-1}\right].$$

Since $\mathrm{E}\left[\epsilon_t^2 \mid \mathcal{F}_{t-1}\right] = 1$, it follows that

$$\mathrm{Var}[a_t \mid \mathcal{F}_{t-1}] = \sigma_t^2 = \alpha_0 + \alpha_1 a_{t-1}^2 + \beta_1 \sigma_{t-1}^2.$$

This proves that the conditional variance of the GARCH(1,1) is time dependent.

Unconditional Variance

We begin by noting that the unconditional variance of a_t is defined as

$$\mathrm{Var}(a_t) = \mathrm{E}[a_t^2].$$

we have

$$\mathrm{E}[a_t^2] = \mathrm{E}[\sigma_t^2 \epsilon_t^2] = \mathrm{E}[\sigma_t^2].$$

Taking expectations on both sides of the conditional variance equation yields:

$$\mathrm{E}[\sigma_t^2] = \alpha_0 + \alpha_1 \,\mathrm{E}[a_{t-1}^2] + \beta_1 \,\mathrm{E}[\sigma_{t-1}^2].$$

Under the stationarity assumption, we have:

$$\mathrm{E}[a_{t-1}^2] = \mathrm{E}[a_t^2] \quad \text{and} \quad \mathrm{E}[\sigma_{t-1}^2] = \mathrm{E}[\sigma_t^2].$$

Noting that $\mathrm{E}[a_t^2] = \mathrm{E}[\sigma_t^2]$, since

$$\mathrm{E}[\sigma_t^2] = \mathrm{E}[\mathrm{E}[\mathrm{a}_\mathrm{t}^2|\mathcal{F}_{\text{t-1}}]] = \mathrm{E}[a_t^2],$$

setting $\mathrm{Var}(a_t) = \sigma^2$ this yields:

$$\sigma^2 = \alpha_0 + (\alpha_1 + \beta_1)\,\sigma^2.$$

Solving for σ^2 gives:

$$\sigma^2(1 - \alpha_1 - \beta_1) = \alpha_0 \quad \Longrightarrow \quad \sigma^2 = \frac{\alpha_0}{1 - \alpha_1 - \beta_1}.$$

Thus, the unconditional variance is:

$$\mathrm{Var}(a_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta_1},$$

which is not dependent of time.

C Multivariate t-Distribution

In this section, we review the theory of the multivariate Student-t distribution, present its probability density function, and discuss some of its key properties. This section follows Joarder and Kibria (2006, p. 59-72).

Probability Density Function

The pdf of the multivariate t-distribution for a p-dimensional random vector X can be written as

$$f(x) = \frac{\Gamma\left(\frac{\nu+p}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{(\nu\pi)^{\frac{p}{2}}} \frac{1}{\sqrt{\det(\Sigma)}} \left(1 + \frac{1}{\nu}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right)^{-\frac{\nu+p}{2}},$$

where $\mu \in \mathbb{R}^p$ is the location vector, Σ is a positive definite $p \times p$ scale matrix, and $\nu > 0$ denotes the degrees of freedom.

Marginal Distributions

An important property of the multivariate t-distribution is that its marginal distributions are also t-distributed. That is, if X follows a multivariate t-distribution, then each component of X is univariate t-distributed (with the same degrees of freedom, but with adjusted location and scale parameters).

Linear combination of a bivariate Student-t

Another useful property of the multivariate Student-t distribution is that any linear combination of its components is again Student-t with the same degrees of freedom. In particular, let

$$(X, Y)^T \sim t_\nu\big(\mathbf{0},\, \Sigma\big)$$

be a bivariate Student-t random vector with ν degrees of freedom. We choose

$$\Sigma = \frac{\nu - 2}{\nu} \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix},$$

so that each marginal has mean zero and $\mathrm{Var}(X) = \mathrm{Var}(Y) = 1$ and $\mathrm{Cov}(X, Y) = \rho$.

For any constants C_1 and C_2, define

$$L = C_1 X + C_2 Y.$$

Then L is t_ν-distributed with variance

$$\mathrm{Var}(L) = C_1^2 \,\mathrm{Var}(X) + C_2^2 \,\mathrm{Var}(Y) + 2C_1 C_2 \,\mathrm{Cov}(X, Y) = C_1^2 + C_2^2 + 2C_1 C_2 \,\rho.$$

Thus, we can write

$$L \overset{d}{=} \sqrt{\frac{\nu - 2}{\nu}\big(C_1^2 + C_2^2 + 2C_1 C_2 \,\rho\big)}\; Z, \quad Z \sim t_\nu.$$

Therefore, the quantiles of L follow by scaling the standard t_ν quantiles by $\sqrt{\frac{\nu-2}{\nu}(C_1^2 + C_2^2 + 2C_1 C_2 \rho)}$.

D Maximum Likelihood Estimation

Our primary goal is to fit some models to data. A standard approach is maximum likelihood estimation (MLE), which seeks for the parameter vector θ that maximizes the probability of the observed data under the assumed model. The theory in this section is from Held and Bové (2020, Chapter. 2).

Likelihood Function

Let $X = (X_1, \ldots, X_n)$ be a random sample with joint density $f(x_1, \ldots, x_n; \theta)$, where θ is the vector of parameters. The likelihood function is then defined as

$$L(\theta | x_1, \ldots, x_n) = f(x_1, \ldots, x_n; \theta).$$

In the special case of independent observations $X_i \sim f(x_i; \theta)$, this factorizes to

$$L(\theta | x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i; \theta).$$

Equivalently, one may write the contribution of a single observation x_i as

$$L_i(\theta|x_i) = f(x_i;\theta),$$

so that $L(\theta|x_1,\ldots,x_n) = \prod_{i=1}^{n} L_i(\theta|x_i)$.

Maximum Likelihood Estimation

The maximum likelihood estimator (MLE) of θ is the parameter value that maximizes the likelihood function:

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} L(\theta \mid x_1,\ldots,x_n).$$

This estimator is the choice of θ that makes the observed data most probable under the assumed model.

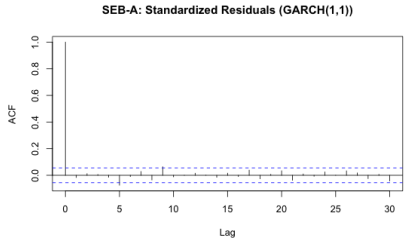
Since the logarithm is strictly increasing, maximizing the likelihood is equivalent to maximizing the log-likelihood

$$\ell(\theta|x_1,\ldots,x_n) = \log L(\theta \mid x_1,\ldots,x_n).$$

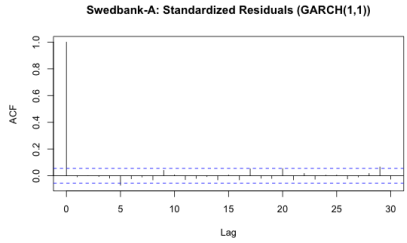
Therefore, the MLE can be defined as

$$\hat{\theta}_{\text{MLE}} = \arg\max_{\theta} L(\theta|x_1,\ldots,x_n) = \arg\max_{\theta} \ell(\theta|x_1,\ldots,x_n).$$

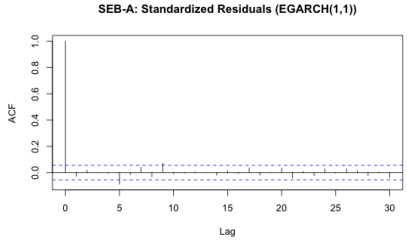
E Residual Diagnostics - ACF Plots



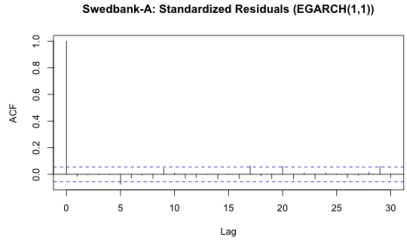
(a) SEB-A: Std. Residuals (GARCH(1,1))



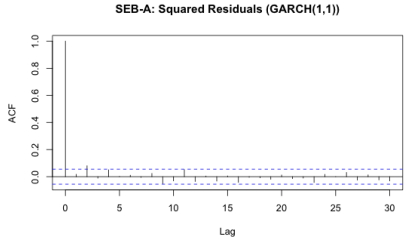
(b) Swedbank-A: Std. Residuals (GARCH(1,1))



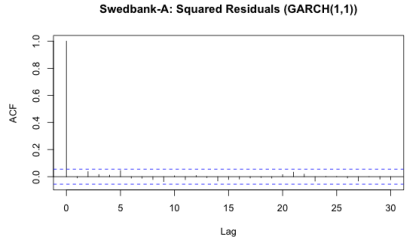
(c) SEB-A: Std. Residuals (EGARCH(1,1))



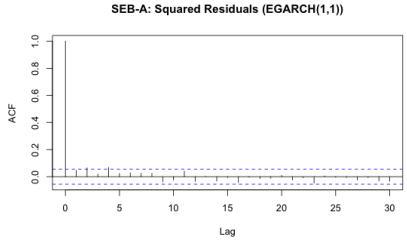
(d) Swedbank-A: Std. Residuals (EGARCH(1,1))



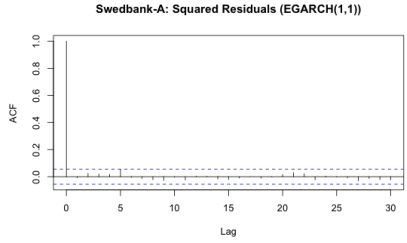
(e) SEB-A: Sq. Residuals (GARCH(1,1))



(f) Swedbank-A: Sq. Residuals (GARCH(1,1))



(g) SEB-A: Sq. Residuals (EGARCH(1,1))



(h) Swedbank-A: Sq. Residuals (EGARCH(1,1))

Figure 12: ACF plots for standardized and squared standardized residuals from GARCH(1,1) and EGARCH(1,1) on SEB-A and Swedbank-A.