Stockholm School of Economics

Master thesis in Finance (4210)

Volatility and Value at Risk modelling using univariate GARCH models

Rishi Thapar (80212)

January 2006

Tutor: Joel Reneby

Acknowledgements

I wish to thank Mr. Joel Reneby for his valuable guidance during the thesis. I am grateful to Mr. Mårten Liljefors of Riksbank for his guidance and support during the thesis work. I wish to thank Mr. Johan Grönquist of Riksbank for inviting me to discuss an opportunity to do thesis with Riksbank. I also wish to thank Mr. David Branto of Riksbank for his support.

Table of Contents

List of Tables

List of Figures

1 Introduction

Value-at-Risk (VaR) has become the most widely used market risk measurement methodology in banks and financial institutions. VaR for a portfolio is a function of volatility of returns of the portfolio. Therefore, the task of VaR estimation can be reduced to forecasting volatility. Forecasting volatility of financial time series has been one of the most active areas of research in finance. In order to capture the empirically observed stylized facts of the financial time series, most important of them being volatility clustering and leptokurtosis, various models have been proposed in literature. Generalised Autoregressive conditional heteroscedasticity (GARCH) models have been very popular to model the time varying variance of the returns as a function of the lagged variance and lagged square returns.

Considering the prominence of VaR in risk management and presence of wide variety of alternative VaR methodologies, evaluation of the predictive accuracy of VaR models is an important issue in risk management. The main challenge in VaR evaluation is that, like in the case of volatility, VaR is a latent (unobserved) variable, therefore, it is not possible to calculate its actual realized value.

This thesis seeks to answer the following two questions of Value-at-Risk Modelling
-*Do volatility forecasting models based on GARCH lead to better performance than a Naïve model that measures volatility by standard deviation of the past returns?*
-*What evaluation framework should be used to test the accuracy of the VaR estimates?*

There has been large volume of literature on VaR modelling issues and approaches. There are many papers on evaluating different GARCH models for estimation and forecasting of volatility of financial assets returns series.

However, regarding evaluation of VaR models, only a few papers have looked at VaR performance in practice, most important of these are Berkowitz and O'Brien (2002) and Jaschke, Stahl, and Stehle (2003), using daily revenues and VaRs for U.S. and German banks respectively. Most of other papers have used simulations or illustrative portfolios to evaluate different VaR models. The returns/P&L series data and internal VaR models of banks and financial institutions are not publicly available. Therefore, there hasn't been substantial empirical work on VaR modelling of the actual portfolio returns/Profit & Loss distributions and on evaluation of VaR models, which are actually in use in the banks.

In this thesis, volatility and VaR modelling using the actual portfolio returns of investment portfolios of the Central Bank of Sweden (Riksbank) is performed using univariate GARCH models. The thesis takes the reduced form approach of Berkowitz and O'Brien (2002) paper. In the reduced form model, volatility or VaR of the portfolio is modelled directly by fitting GARCH on daily returns series rather than the common practice of using GARCH to model risk factors of a portfolio in the context of the popular variance-covariance approach of VaR modelling. The methodology used in the thesis is also in line with the Portfolio aggregation method proposed in Riskmetrics' technical paper (Zangari, P., 1997).The Portfolio Aggregation approach estimates VaR using the volatility of the portfolio returns rather than variance-covariance matrix of the risk factors. Reduced form GARCH models, due to their parsimony and flexibility, offer a simple alternative to the structural models. However, there is a need to empirically evaluate the performance of these models.

This thesis contributes in this direction by empirically studying the performance of reduced form GARCH modelling of the actual investment portfolios of a bank. A comprehensive evaluation framework is employed to test the predictive accuracy of volatility forecasts and VaR estimates. To test the accuracy of volatility forecasts, various error statistics and hypothesis testing are employed and to evaluate VaR estimates, back testing methodology recommended by Basel guidelines as well as advanced tests are performed.

The outline of the thesis is as follows. In section 2, theoretical background about volatility forecasting and VaR is presented. Details about some related previous papers are presented in section 3. Data used in the thesis is presented in section 4. Detailed methodology for model estimation and volatility forecasting, evaluation of volatility and VaR estimates is presented in Section 5. The empirical results and analysis of volatility and VaR estimates from the GARCH models are described in Section 6 and section 7 gives conclusions and also presents some suggestions for further research.

2 Theoretical background

This section introduces various volatility-forecasting models and discusses, in details, GARCH based models. This section also gives a brief overview of Value at Risk concept. The section starts with salient stylized facts about the volatility of financial assets returns, which needs to be considered while specifying volatility forecasting models.

2.1 Stylized facts of market volatility

A typical trading portfolio of a bank or investment house is characterised by variety of different types of financial assets and derivatives and thus has linear as well as non-linear exposure to a set of market factors. Therefore, the portfolio return series usually exhibits non-normal and fat tail characteristics. It is important to consider empirically observed stylized facts of market volatility of a financial asset/portfolio before specifying a volatility estimation model. A model's ability to capture important empirical stylized facts is a desirable feature. The important stylized facts about of financial assets series, which have been documented in numerous studies, are described below.

2.1.1 Leptokurtosis (heavy tails and sharp peaks)

The distribution of the financial assets returns is *leptokurtosis*, i.e., exhibit excess kurtosis (heavy-tails) and sharp peaked. Typical kurtosis estimates for the financial return series are found to be in the range of 4 to 50. A normal distribution has kurtosis value equals 3. Therefore, kurtosis value exceeding 3 indicates heavy-tails. In a heavy-tailed distribution, extreme outcomes are more frequent than what the use of a normal distribution would predict. Even after correcting returns for volatility clustering (e.g. via GARCH-type models and/or fitting fat-tailed distributions), the residual time series still exhibit heavy tails. Therefore, this non-Gaussian and heavy-tailed characteristic of financial time series makes it necessary to use other measures of dispersion than the standard deviation in order to capture the variance of the returns.

2.1.2 Volatility Clustering

Extreme returns show high variability, as evident from the heavy tails and non-negligible probability of occurrence of extreme values. Also extreme values appear in clusters, extreme returns to be followed by other extreme returns, although not necessarily with the same sign. The implication of volatility clustering is that the volatility shocks today influences the expectation of volatilities of many future periods ahead.
Return series are not strictly white noise although they show little autocorrelation especially in liquid markets. The absence of autocorrelations in returns series gives some empirical support for 'random walk' theory in which the returns are considered to be independent random variables. However, it has been shown empirically that this absence of serial correlation does not imply any nonlinear function of returns will also have no autocorrelation. Absolute or squared returns exhibit significant positive autocorrelation or persistence (slow decay in autocorrelations). Therefore, due to this nonlinear dependence, financial time series have auto correlation in volatility of returns but not in the returns themselves. Also, if we increase the time scale i.e. weekly and monthly return series tend to exhibit serial correlation.

2.1.3 Leverage effects

It is found in many studies that there are leverage effects (i.e. volatility of returns are negatively correlated with the returns of the assets.) in financial time series. A negative shock leads to a higher conditional variance in the subsequent period than a positive shock would do.

2.2 Modelling time varying volatility

In forecasting volatility, especially in modelling variance of short-horizon asset returns, usually mean return is assumed equal to zero. This is justified by the argument that mean return of an asset is typically several orders of magnitude lower than its standard deviation.
Therefore, the first moment for the return series is usually defined as below.

$$r_t = \varepsilon_t$$

The most popular class of volatility forecasting models, described below, are discrete-time parametric volatility models, which explicitly model the expected volatility, σ^2_{t+h} (h-step ahead variance) as a non-trivial function of the historical time information set, F_t. Therefore, these models parameterize the first two conditional moments (mean and variance) of the returns time series. These models can be broadly classified into three categories viz. MA/EMWA models, ARCH and Stochastic Volatility (SV) models. MA/EMWA and ARCH models are described below.

2.2.1 MA/EWMA

Moving Average (MA) models are one of the simplest models, where forecasted volatility is calculated as moving average of the historical variance. In the case of exponential weighted moving average (EWMA) models volatility of the next period is forecasted as a MA process of weighted square deviations from the mean and the weights decay exponentially with a decay factor λ. EWMA models are more responsive than the simple moving average to sudden changes in volatility. Risk Metrics, the most commonly used model in practice employs EWMA model to model the variance σ^2_{t+1} with λ = 0.94 and can be represented by the following equation.

$$\sigma^2_{t+1} = a_0 + \lambda\sigma^2_t + (1-\lambda)(r_t - \hat{\mu}_t)^2$$

In fact, the Risk Metrics model is a non-stationary version of GARCH (1, 1), where the persistence parameters sum to 1.

2.2.2 ARCH

In 1982, R F Engle introduced ARCH class of models in which time-varying conditional variance is modelled with the AutoRegressive Conditional Heteroscedasticity (ARCH) processes that use past disturbances to model variance of the series. In other words, today's conditional variance is a weighted average of the past squared disturbances. An ARCH (q) model is specified by

$r_{t+1} = \mu_{t+1} + \varepsilon_{t+1}$ where $\mathrm{E}(r_{t+1} \mid F_t) = \hat{\mu}_{t+1}$ and $\mathrm{E}(\varepsilon^2_{t+1} \mid F_t) = \sigma^2_{t+1}$ and

The variance $\sigma^2_{t+1} = \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon^2_{t-j+1}$. Therefore, conditional on the past, the ARCH model is normal but heteroscedastic.

2.2.3 GARCH

Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model, first suggested by Tim Bollerslev in 1986 is obtained by adding p autoregressive terms for σ^2_t to the ARCH (q) model. Therefore GARCH (p, q) has the following specification.

$$\sigma_{t+1}^2 = \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j+1}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j+1}^2$$

With $\alpha_0 \geq 0$, $\alpha_j \geq 0$ for all *j = 1 to p*, $\beta_j \geq 0$ all *j = 1 to q* and $\sum_{j=1}^{q} \alpha_j + \sum_{j=1}^{p} \beta_j < 1$

Just as an ARMA model often leads to a more parsimonious representation of the dependencies in the conditional mean than an AR model, the GARCH (*p, q*) model provides a similar added flexibility over the linear ARCH (*q*) model when parametrizing the conditional variance. The ARCH (q) model corresponds to a GARCH (0,q) model. GARCH (p,q) models are used in practice because GARCH (p,q) model allows for parsimonious parameterization of an ARCH (∞) model.
In practice, low order GARCH models are widely used. GARCH (1,1) model is given by the following equation.

$$\sigma_{t+1}^2 = \alpha_0 + \alpha\ \varepsilon_t^2 + \beta\ \sigma_t^2$$

The variance σ_{t+1}^2 is a weighted average of

- long term variance α_0 /(1-α - β) ,
- prior variance σ_t^2 with weight β and
- squared disturbance term ε_t^2 with weight α

The restrictions on the parameters α and β are 0 <= α_0 <= 0 , 0 <= α <= 1, 0 <= β <= 1 and $\alpha + \beta$ < 1. These restrictions ensure that the weights are positive and sum to 1.
The magnitude of α and β determines the short-term dynamics of the forecasted volatility series. A large value of β indicates persistence, i.e. Shocks (extreme values) of the conditional variance will take long time to die out. Large value of α indicates that the volatility reacts quite fast to the market movements.
The above formula for GARCH (1,1) nicely demonstrates the essence of the volatility clustering feature in the GARCH model. If the market has been volatile in the current period, next period's variance will be high, which is intensified or offset in accordance with the magnitude of the return deviation of the current period. If, on the other hand, today's volatility has been relatively low, tomorrow's volatility will be low as well, unless today's portfolio return deviates from its mean considerably. The impact of these effects depends on the parameter values. For $\alpha + \beta < 1$, the conditional variance exhibits mean reversion, i.e., after a shock it will eventually return to its unconditional mean. The condition $\alpha + \beta < 1$ also ensure that model is covariance-stationary.
Due to its ability to capture salient features of the return dynamics in very parsimonious and easily estimated specifications, GARCH has become the popular model in financial risk management.

2.2.4 Extensions to the basic GARCH model

The basic GARCH model is usually a good starting point while modelling volatility but various extensions and variants to the basic GARCH(p,q) model have been proposed and used in finance. The development of these extensions and variants aim to capture the stylized facts of the financial assets distribution in a better manner. One of the major restrictions of the basic GARCH model is that fails to capture the asymmetric or leverage effects i.e. asymmetrical response of volatility to the market moves. It is It is often noticed in the financial markets that a negative shock leads to a higher conditional variance in the subsequent period than a positive shock would do.
Exponential GARCH or EGARCH, introduced by Nelson in 1991, captures this asymmetric response by specifying the conditional variance as a function of not only of the magnitudes of the lagged residuals and but also their signs. Many empirical studies have also shown that conditional distribution for the error term in the conditional mean equation often has heavier tails than the Gaussian distribution as assumed in the basic GARCH model. Although in the basic GARCH model,

conditionally normal distributions produce heavy-tailed unconditional distributions, often this is not enough to capture the excess kurtosis in the data. Therefore, Symmetric but fat-tailed distributions like Student-t or generalized error distribution (GED) has been used instead.
In the thesis following variants of GARCH models are used.

- **GARCH with normal distributed errors (GARCH)**

$$\varepsilon_{t+1} = \sigma_{t+1} z_{t+1}\ ,\ z_{t+1} \mid F_t \ \sim N(0,1)$$
$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1 \varepsilon_t^2 + \beta_1 \sigma_t^2$$

- **GARCH with student t distributed errors (GARCH_t)**

$$\varepsilon_{t+1} = \sigma_{t+1} z_{t+1}\ ,\ z_{t+1} \mid F_t \ \sim t(v)$$
$$\sigma_{t+1}^2 = \alpha_0 + \alpha_2 d_t + \alpha_1 \varepsilon_t^2 + \beta_1 \sigma_t^2$$

- **Exponential GARCH with normal errors (EGARCH)**

As explained above, exponential GARCH model can capture the asymmetric response of volatility to the market moves. Taking logarithms of conditional variances allows asymmetry in response of volatility to market moves. With appropriate conditioning of the parameters, the EGARCH specification below captures the stylized fact of leverage effects.

$$\varepsilon_{t+1} = \sigma_{t+1} z_{t+1}\ ,\ z_{t+1} \mid F_t \ \sim N(0,1)$$
$$\log \sigma_{t+1}^2 = \alpha_0 + \alpha_1 \left\{ \upsilon_1 \varepsilon_t + \upsilon_2 \left| \varepsilon_t \right| - E \left| \varepsilon_t \right| \right\} + \beta_1 \log \sigma_t^2$$

- **GARCH in mean (GARCHM)**

The GARCH-M model is defined simply by taking the conditional variance as a regressor in the mean equation.

$$r_{t+1} = \mu + \delta \sigma_t + \varepsilon_{t+1}$$
$$\varepsilon_{t+1} = \sigma_{t+1} z_{t+1}\ ,\ z_{t+1} \mid F_t \ \sim N(0,1)$$
$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1 \varepsilon_t^2 + \beta_1 \sigma_t^2$$

- **Asymmetric Threshold GARCH (ATGARCH)**

Another asymmetric specification for a GARCH model is ATGARCH model. The idea behind this model is that asymmetric behaviour of the negative shocks are sources for additional risk.

$$\varepsilon_{t+1} = \sigma_{t+1} z_{t+1}\ ,\ z_{t+1} \mid F_t \ \sim N(0,1)$$
$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1 (\varepsilon_t - \kappa_1)^2 + \kappa_2 D_t (\varepsilon_t - \kappa_1)^2 + \beta_1 \sigma_t^2$$

Where D_t = 1 if $\varepsilon_t \prec \kappa_1$ otherwise it is zero. κ_1 is asymmetry parameter and κ_2 is threshold parameter.

- **Asymmetric GARCH (AGARCH)**

Asymmetric model captures asymmetrical response of volatility to the market moves. In this specification, κ_2, the threshold parameter is set to zero.

$$\varepsilon_{t+1} = \sigma_{t+1} z_{t+1}\ ,\ z_{t+1} \mid F_t \ \sim N(0,1)$$

$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1(\varepsilon_t - \kappa_1)^2 + \kappa_2 D_t(\varepsilon_t - \kappa_1)^2 + \beta_1\sigma_t^2$$

With $\kappa_2 = 0$

- **Threshold GARCH (TGARCH)**

Another asymmetric specification is the threshold GARCH (TGARCH) model, which adds a dummy variable to the GARCH process.
In this specification, κ_1, the asymmetric parameter is set to zero.

$$\varepsilon_{t+1} = \sigma_{t+1} z_{t+1}\ ,\ z_{t+1} \mid F_t \sim N(0,1)$$
$$\sigma_{t+1}^2 = \alpha_0 + \alpha_1(\varepsilon_t - \kappa_1)^2 + \kappa_2 D_t(\varepsilon_t - \kappa_1)^2 + \beta_1\sigma_t^2$$

2.3 Value at Risk

Value at risk (VaR) approach has emerged as industry standard to measure market risk, both for capital requirements and for internal risk control, during the last few years. In the 1996 amendment of Basel Accord, which outlined market risk capital requirement for the banks, advised the banks to use VaR approach for assessment of their market risks and for calculating regulatory capital requirement. VaR of a portfolio is defined as the maximum loss on the portfolio that can be expected with a certain level of confidence over a certain holding period.
To introduce some notation, consider a portfolio of risky assets and assume V_t as the value of portfolio at time t. Assume that we want to calculate risk for the time period [t, t+1]. We denote the loss distribution of the portfolio by $L_{t+1} = -(V_{t+1} - V_t)$ and the distribution function for the loss series is F_L such that $P(L \le x) = F_L(x)$. Then VaR at α ($\alpha \in (0,1)$) can be defined as the α -quantile of F_L . In other words, VaR is product of standard deviation of distribution of L_{t+1} series and α -quantile of the standardized distribution with unit variance (ϕ) and zero mean.

$$VaR(\alpha)_{t+1} = q_\alpha(F_L) = |q_\alpha(\phi)|\sigma_{t+1}$$

Nominal value of VaR can is then a product of Value of the portfolio at t, α -quantile of the standardized distribution and volatility of return series $r_{t+1} = (V_{t+1} - V_t)/V_t$

$$VaR(\alpha)_{t+1} = V_t|q_\alpha(\phi)|\sigma'_{t+1}$$

Therefore, as evident from the above expressions, VaR is a function of the volatility forecast and is dependant on the assumptions of distribution of loss/return series. Therefore, accuracy of VaR relies on accuracy of volatility forecasts.
A common assumption while calculating VaR is that the return series are normally distributed. However, as we have discussed in the earlier sections, the returns show fat-tailed. Therefore, this assumption of normality under-estimates VaR and introduce substantial model risk.

In conclusion, this section begins with an overview of salient stylized facts of financial time series. The accuracy of a particular volatility model is dependant on the degree to which it is able to captures these characteristics of financial time series. Hence, it is important to consider the above-mentioned stylized facts while modelling volatility. In this section, different discrete-time parametric volatility models were explained. The concept behind GARCH models was presented and various extensions of GARCH models were introduced. GARCH models have become popular in financial risk management due to their ability to capture salient features of the return dynamics in very parsimonious and easily estimated specifications.

3 Related Research

This section presents a brief overview of the various academic papers that have studies volatility forecasting and VaR models. After that, a summary of the methodology and results of research papers, that are closely related to the thesis are presented.

3.1 Overview

Over the years, researchers have approached volatility forecasting from different angles viz. historical time series based models like AR/EWMA, GARCH, stochastic volatility, implied volatility models, non parametric models, genetic and neural networks based models. The research paper by Poon and Granger (2002) does an extensive survey of 93 different papers that studied forecasting performance of volatility models. Out of these about 17 papers studied performance of GARCH models. According to the paper, although, the conclusions from the papers vary a lot, they have significant common characteristics e.g.

- They test a large number of very similar models all designed to capture volatility persistence
- They use a large number of error statistics each of which has a very different loss function
- They forecast and calculate error statistics for variance and not standard deviation, which makes the difference between forecasts of different models even smaller
- They use squared daily, weekly or monthly returns to proxy daily, weekly or monthly "actual volatility", which result in extremely noisy volatility estimates. The noise in the volatility estimates makes the small differences between forecasts of similar models indistinguishable.

Value-at-Risk (VaR) has become a well-known tool for measuring market risk since the implementation of the Basel accord on Capital Requirements (1996). There has been large volume of literature on VaR modelling issues and approaches. The internal VaR models and VaR figures of banks and financial institutions are, however, not publicly available. Many papers have used simulations or illustrative portfolios to evaluate different VaR models. Therefore, there hasn't been substantial empirical work on evaluation of VaR models, which are actually in use in the banks and financial institutions.

3.2 Related Papers

3.2.1 Wong Sham CM et al (2003)

This paper evaluates performance of VaR forecasts of nine univariate time series models (random walk with constant volatility, AR/ARMA models with constant volatility, and AR/ARMA returns model with GARCH (1,1) volatility), using Basel back-testing criteria. The paper has selected a stock portfolio, Australia All Ordinary Index (AOI), as a proxy for the portfolio of a bank. The normality of the distribution of the portfolio returns is assumed. Therefore, VaR is calculated by multiplying square root of the forecasted variance with the Market Value of portfolio and the critical value of the required confidence level. Relative VaRs (which refers to the percentage of the portfolio value which may be lost after h-day holding period with a specified probability), for long and short positions of portfolio on AOI are also calculated. The VaR models are estimated using 4000 observations and one-step ahead forecasts are produced. For evaluating reproductive accuracy of the volatility forecasts, Mean Square Error (MSE) and Mean Absolute Error (MAE) are calculated. In order to test the accuracy of VaR estimates, failure rates and size of the forecast errors are calculated. Dividing the sample period into four sub periods tests the robustness of the results and the performance of the models are evaluated across these sub periods also.
The key conclusion of the paper is that ARCH and GARCH models consistently fail back-testing whereas models of constant volatility pass back-testing for most of the sub periods.

3.2.2 Berkowitz and O'Brien (2002)

This paper uses the actual daily Profit/Loss (P/L) and VaR of the trading books of six US banks to evaluate the internal structural VaR models of the Banks against an ARMA-GARCH (1,,1) model. It employs Likelihood- Ratio tests for comparing unconditional as well as conditional coverage of the models. The internal VaR models under evaluations are parametric VaR models, based on variance and co-variance between the various risk factors affecting the trading portfolios of the bank. Thus, these models take into consideration the effect of the change in the portfolio positions. Whereas, the ARMA-GARCH (1,1) doesn't takes this into consideration and model the VaR based on the conditional volatility of the historical P&L. Therefore, the null hypothesis in the paper is that the internal VaR models perform better than the time series based ARMA-GARCH (1) model. The null hypothesis is rejected for almost all banks and the GARCH model based on daily trading P&L outperforms internal VaR model for all the banks.
The results show that GARCH models generally provide for lower VaRs and are better at predicting changes in volatility. However, the mean violation rate for the GARCH VaRs also is lower than that of the banks' VaRs. The internal VaR models pass test of unconditional coverage (with mean violation rate of 0,5% for 99% VaR), however the magnitude of the "failures" (exceedence of losses over VaR) were high (between 2-4 standard deviations beyond the mean VaR) and the failures tend to be clustered. The clustering of failures indicates that the structural VaR models are not able to capture time-varying volatility adequately.
The average GARCH VaRs are also lower than that of internal VaR models but the striking results are that the violations in the GARCH VaRs are not larger than that in banks' internal VaRs models. Thus, the paper concludes that GARCH models are better because they imply low level of regulatory capital requirement without producing larger violations. Although the GARCH models cannot account for positions' sensitivities to current risk factor shocks or changes in current positions, they are more parsimonious and accurate to model the dynamics of portfolio P&L.

3.2.3 Polasek and Pojarliev (2003)

This paper studies the comparative performance of time series models based on Risk Metrics's EWMA and different GARCH models viz. GARCH with normal errors, GARCH with t-distribution errors, asymmetric GARCH and exponential GARCH and Power GARCH. A hypothetic portfolio of 1 Million USD invested in QQQ (a share that tracks NASDAQ 100 index) is used. The normality of returns of NASDAQ 100 index is assumed. Therefore, one day 95% VaR is given by multiplying the square root of the forecasted variance by 1.65. Various p and q values for GARCH (p,q) model were run and GARCH (1,1) was chosen on the basis of lowest AIC and BIC. Regressing the squared returns on a constant and on the forecasted variance compared volatility-forecasting performance of the models. The performance of different VaR models is evaluated using failure rate. The likelihood –ratio tests of unconditional, independence and conditional coverage for 1%-10% VaR range were done. Further, loss function that incorporates penalties (a function of failure rate) and VaR cost (opportunity cost due to overestimation of VaR) was also used. GARCH model with normal errors performs best in terms of lowest number of failures and loss function. It passes conditional coverage test for 2%-5% VaR range. GARCH models in general performed better than Risk Metrics and constant volatility models.

3.2.4 Sarma et al (2003)

This paper uses a comprehensive VaR model selection framework, with failure rate, likelihood-ratio and regression based tests for conditional coverage, loss functions and one-sided non parametric sign tests. 16 models based on EWMA (for 50,125,250,500 and 1250 days window), Historical Simulation (for 50,125,250,500 and 1250 days window) and AR (1)-GARCH (1,1) for VaR estimation (95% and 99%) of S&P 500 and Nifty (India's NSE stock index) indices.
For both 95% & 99% VaR of S&P 500, no model was able to pass regression based conditional coverage tests. For 95% VaR of Nifty Risk Metrics model performed best and survived all tests. For 99% VaR, AR (1)-GARCH (1) performed best.

In conclusion, volatility and VaR forecasting has been approached from different angles in financial research with GARCH based modelling being one of the important approaches. The four related papers, discussed in details in this section, have studied GARCH models among the other models. These papers differ from each other particularly in terms of how the predictive accuracy of the volatility or VaR forecasts is evaluated. As seen in most of the other related research papers, these papers (except Berkowitz and O'Brien (2002).) use stock market index or hypothetical portfolios. Berkowitz and O'Brien (2002) has analyzed the distribution of historical trading P&L and the daily performance of VaR estimates of six large U.S. banks.

4 Data

The data consist of log returns of the three investment portfolios primarily consisting of government-guaranteed securities denominated in the foreign currencies named as Portfolio A, Portfolio B and Portfolio C in this thesis. The log returns data of the respective benchmark portfolios (named Benchmark A, Benchmark B and Benchmark C) for each investment portfolio are also used. These benchmark portfolios incorporate the Bank's preferences for liquidity, risk and return and performance of the investment portfolios are evaluated against the respective benchmark portfolio. Therefore, the volatility of a benchmark portfolio can be considered as a good proxy for the volatility of the corresponding investment portfolio and thus can be used to estimate VaR of the investment portfolio. The data used consist of daily returns of the three investment portfolios and corresponding Benchmark portfolios. An out of sample data consisting of 250 daily returns are used for evaluating volatility and VaR forecasts.

4.1 Descriptive Statistics

From Figure 1 (in Appendix B), it is evident that all the portfolio returns series exhibit volatility clustering. The plots of squared returns in Figure 2 also corroborate volatility clustering. Similarly, from the density plots and Quantile-Quantile (QQ) plots in Figure 3, it is clear that all the series show non-normal and fat-tailed behaviour. The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots of the return series don't show significant auto-correlation but the squared returns do exhibit auto-correlation, up to lag lengths more than 10.
It can be followed from the Table 1 below, that all the series exhibits heavy tails (excess kurtosis values different from zero). This indicates the necessity of fat-tailed distributions to describe the returns series' conditional distribution. It can be seen from table that we can reject the null hypothesis of normality in all returns because the p-values are lower than 5%. Table 5 in Appendix B, presents the Jarque-Bera test statistic and associated p-value of this test for lags 5, 10, 20, 30 and 50 for both returns and squared returns. For most of the return series, we can't reject the null hypothesis of no autocorrelation as the p values are greater than 0.05 .For all squared return series, except in case of BENCHMARK A and BENCHMARK C, that we can reject the null hypothesis of no autocorrelation.

Table 1: Descriptive statistics (in-sample period)

	PORTFOLIO A	BENCHMARK A	PORTFOLIO B	BENCHMARK B	PORTFOLIO C	BENCHMARK C
N	450	450	435	442	444	444
Mean	0,00021	0,00023	0,00026	0,00027	0,00025	0,00027
Std. Dev.	0,0022	0,0022	0,0021	0,0021	0,0031	0,0031
Skewness	-0,4133	-0,4166	-0,5195	-0,4965	-0,5709	-0,6096
Excess Kurtosis	0,7496	0,6975	0,8969	0,8395	0,8361	0,6214
Jarque-Bera	14,011 (0,0009)	13,630 (0,0011)	18,642 (0,0001)	17,549 (0,0002)	21,796 (0,0000)	26,743 (0,0000)

5 Methodology

This section presents volatility &VaR forecasting and evaluation methodology used in the present thesis. VaR modelling of a portfolio based on time-series models involves specifying a parametric distribution for the portfolio returns and estimating the parameters of the distribution using historical data. The VaR of the portfolio can be then calculated by multiplying the square-root of the conditional variance with the appropriate critical value for the standardized distribution.
In the present thesis, GARCH models are used to forecast volatility for the returns of the investment portfolios of the Riksbank. Following types of GARCH (1,1) models are estimated for all six series.

- GARCH with normal distributed errors (GARCH),
- GARCH with student t distributed errors (GARCH_t),
- Asymmetric GARCH (AGARCH),
- Asymmetric threshold GARCH (ATGARCH),
- Exponential GARCH with normal errors (EGARCH),
- GARCH in mean (GARCHM) and
- Threshold GARCH (TGARCH).

The best model is selected on the basis of log-likelihood value and AIC. The selected model is then used to produce one-day ahead variance forecast. The sample is then rolled one-day ahead and the model is re-estimated and again a one-day ahead forecast is generated and so on. In this way, 250 out-of-sample one-day ahead forecasts are generated. Out of sample period with 250 observations is used to evaluate predictive accuracy of the variance forecasts. The volatility forecasts from the GARCH models are compared with a Naïve model, in which the volatility calculated using rolling standard deviation of the past n observations (n equals the number of observations in the in-sample data of the GARCH model). The volatility evaluation methodology is explained in detail in sections below.
Daily VaR forecasts are estimated for the out-of-sample period. Daily VaR for a portfolio for time t can be calculated on day t-1 by using the volatility forecast for day t done on the day t-1 Three VaR forecasts for an investment portfolio are calculated for each day, one using volatility forecast of the returns of the investment portfolio. Second, using the volatility forecast of the corresponding benchmark portfolio. Third, using the volatility forecast of the Naïve model. VaR forecasts evaluation is done using multiple tests, which are detailed, in section below.

5.1 Evaluation of predictive accuracy of Volatility models

In order to evaluate VaR models, it is important to compare the accuracy of the volatility forecasting process underlying in the VaR model. Evaluating volatility forecasts however, poses a challenge. Since, volatility is a latent (unobserved) variable, therefore, it is not possible to calculate the actual true volatility. Therefore, the ex-post evaluation of volatility forecast accuracy must content with the fundamental error-in-variable problem due to this issue. In most of the empirical studies of volatility forecasting, daily squared returns are used as a proxy for actual volatility. This thesis also takes same approach, as explained below.

5.1.1 Proxy for Actual Volatility

Consider the following specifications for the returns
$r_t = \mu + \varepsilon_t$ and $\varepsilon_t = \sigma_t z_t$
Where z_t is i.i.d.
$\mathrm{E}\left[r_t^2 \mid F_{t-1}\right] = \mathrm{E}\left[\varepsilon_t^2 \mid F_{t-1}\right]$ because $\mathrm{E}\left[\mu \mid F_{t-1}\right] = 0$

Therefore,
$$\mathrm{E}\left[\varepsilon_t^2 \mid F_{t-1}\right] = \mathrm{E}\left[z_t^2 \sigma_t^2 \mid F_{t-1}\right] = \sigma_t^2 \mathrm{E}\left[z_t^2 \mid F_{t-1}\right] = \sigma_t^2$$

Because $\mathrm{E}\left[z_t^2 \mid F_{t-1}\right] = 1 \quad z_t \sim \mathrm{N}(0,1)$ and $z_t^2 \sim \chi_1^2$

Although ε_t^2 is an unbiased estimator for σ_t^2, it is a noisy estimator due to its asymmetric distribution. Therefore, some studies have used intra-day high frequency returns to construct a better proxy for the true realized volatility. True volatility is estimated by the sum of intraday squared returns at short intervals such as fifteen minutes. Such a volatility estimator has been shown to provide an accurate estimate of the latent process that defines volatility. In this thesis, squared daily returns are used as a proxy for the true volatility because the intraday returns data was not available.

5.1.2 Error Statistics

Different error statistics and hypothesis tests based on regression and quadratic loss functions are used to assess the predictive accuracy of the models.. To evaluate performance of the different models in forecasting conditional variance, models, error statistics used are as follows
- Mean Square Error (MSE)
- Median Square Error (MedSE)
- Mean Absolute Error (MAE)
- Adjusted Mean Absolute Error (AMAPE)
- Mean Mixed Error for under-predictions (MME(O))
- Mean Mixed Error for over-predictions (MME(U))
- Theil- Inequality Coefficient (TIC)

These computations of AMAPE, MME and TIC are explained in section A1 of Appendix A.

5.1.3 Non-Parametric tests on error statistics

It is usually not sufficient to compare two or more competing models by taking intro considering the average error statistics like MSE, MAE etc. In order to test the superiority of one model over other, it is also important to see if the specified error loss functions (e.g. MSE etc.) are statistically significantly better in one model than in other. One of the ways to do is to employ non- parametric sign and/or rank tests. In this thesis, one-sided non parametric sign test is used, as used in Sarma et al (2003). The null hypothesis of this test is that both models under consideration have same forecasting accuracy against a one-sided alternative hypothesis of superiority of one model over the other.

$$H_0 : \delta = 0$$
$$H_1 : \delta \prec 0$$

Where δ is defined as the median of the differential loss function distribution, $\{dl_t\}$, where $dl_t = l_{it} - l_{jt}$, with l_{it} and l_{jt} are loss function for model i and j respectively for day t.

Further, define an process, $\{\mathrm{s}_t\}$, where

$s_t = 1$ *if* $dl_t \geq 0$ *0 otherwise*
0 otherwise

The sign statistic is then given by

$$S_{ij} = \sum_{t=1}^{250} s_t$$

Under null hypothesis, the standardized S_{ij}, as given below is asymptotically standard normal.

$$S_{ij}^{\ a} = (S_{ij} - 0.5 * 250) / \sqrt{0.25 * 250} \sim N(0,1)$$

If $S_{ij}^{a} \prec -1.66$, we can reject the null hypothesis at 5% confidence level. Rejection of null hypothesis means that that model i is significantly better than model j in terms of the given loss function.
The advantage of this sign test is that the distribution of the sign statistic is agnostic to the loss function distribution under consideration.

5.1.4 Mincer- Zarnowtiz regression

In addition to the above error statistics a regression based performance measure, , known as Mincer-Zarnowtiz regression, is used to evaluate conditional bias in the volatility forecasting models. It has been largely used for evaluating economics forecasts. However, many studies have also used it and its variants for the conditional variance evaluation.
The forecasted conditional variance is regressed on a constant and on the ex-post true variances(proxied by squared returns) for the out-of-sample period and

$\sigma_t^2 = \alpha + \beta \hat{\sigma}_t^2 + \varepsilon_t$ where $\sigma_t^2 = r_t^2$

The necessary condition for $\hat{\sigma}_t^2$ to be conditionally unbiased is $\alpha = 0$ and $\beta = 1$. The forecasting performance of a model can be measured using R^2 of the regression.

5.2 VaR evaluation

Considering the prominence of VaR in risk management and presence of wide variety of alternative VaR models, assessment of VaR estimate is an important issue in risk management. The risk arising from the faulty forecasts i.e. model risk, is an important issue in risk management in financial institutions. The financial regulatory organizations need to make sure that the VaR models used by the banks are not systematically biased. Alike evaluation of volatility models, evaluation of VaR forecast is not straightforward because actual VaR is unobservable. Various methodologies for VaR evaluation, which are used in the thesis, are discussed below.

5.2.1 Basel Back-Testing

Back-testing of a VaR model, as recommended by on Basel guidelines for market risk capital requirements, requires the model to be accurate(a model is accurate if the actual loss is smaller than the VaR forecast) at least at least on 99%(for VaR with 1% significance level) and 95% (for VaR with 5% significance level) of the time. There should be at least 250 days(around 1-year data) for back testing the daily VaR.
However, this simple approach of evaluating a VaR model is neither powerful nor accurate because to pass the back-test, a VaR model needs only to be correct “on average” and also this test doesn’t take into account the magnitude of failures and independence of failures.

5.2.2 Kupiec Test

More sophisticated tests have been proposed in literature to test the statistical accuracy of the VaR forecasts. The Kupiec test is based on a likelihood-ratio test- statistic. To fix notation, consider a series of one day ahead VaR forecasts which are estimated at confidence level $1-p$ (e.g. for 95% VaR, p = 0,05).
We can define a “failure process”, $\{f_t\}$ with $t \in [1,T]$ and

f_t = 1 if $VaR_t \leq$ Actual $loss_t$
0 otherwise

The process $\{f_t\}$ is a binomial process with independent draws of *1s and 0s*. Under the null hypothesis, VaR estimates are accurate and thus probability of occurrence of “failures” (when $f_t = 1$) on each draw equals *p*.
Since the probability of occurrence of *n* number of failures is $(1-p)^{T-n} p^n$, the *LR* statistic is given by

$$LR_{uc} = -2\ln\left[(1-p)^{T-n} p^n\right] + 2\ln\left[(1-n/T)^{T-n}(n/T)^n\right]$$

Under null hypothesis, the above test statistic has a χ^2 distribution with one degree of freedom.

This test has some limitations. Firstly, since, the failures occur rarely (by design), Kupiec test has poor power characteristics, which become worse as the confidence interval being tested increases. We need a large sample size for the test to have significant power.
Secondly and more importantly, this test assumes that the occurrences of failures are unconditional because this test provides average and unconditional (i.e. without reference to the information available at each time point) coverage by simply counts the failures over the entire period and this test lacks power against the dependence between the failures i.e. the zeros and ones come clustered together in a time-dependent fashion.

5.2.3 Christoffersen’s Likelihood-Ratio tests

Since, VaR are interval forecasts (i.e. one-sided interval forecasts of the portfolio returns), there is more information available in the failure process rather than just average coverage. Also, due to the presence of persistence and conditional heteroscedastic volatility of portfolio returns, conditional probabilities of failures should also be tested. For example, in the periods of high volatility of portfolio returns, the VaR forecasts should be larger than over-all average value and vice-verse. A VaR model that ignores the time-varying dynamics of the returns, might produce correct unconditional coverage, but it may fail to account for persistence and time-varying attributes.
Christoffersen developed Likelihood-Ratio tests for evaluating unconditional coverage, independence and correct conditional coverage. These tests are described in details in Appendix A.

5.2.4 Loss Function tests

Apart from the hypothesis based tests of Kupiec and Christoffersen, VaR models can also be evaluated using loss functions, that test “economic” significance rather than “statistical” significance and take consideration the specific interests (in other others utility function) of the risk managers. Lopez(1999) introduced regulatory loss functions that assign a numerical score, which reflects specific regulatory concerns, to VaR estimates. A model that has minimum value of the loss function is the better one.

One example of such a loss function, which is used in the thesis, is

$L_t = 1+(Actual\ Loss_t - VaR_t)^2$ *if* $VaR_t \leq Actual\ loss_t$
0 otherwise

The above loss function takes into consideration the magnitude of the failure, i.e. by how much the actual loss exceeds VaR estimate and thus penalises the model that produces higher magnitude.
The ability to introduce extra information, i.e. about the magnitude of the failure and flexibility to define specification of the loss function are two main advantages of this loss function.
Two different VaR models can be easily compared by designing a simple hypothesis test based on the above mention loss function and by performing a one-sided sign test, which is described in detail in the previous section.

In conclusion, in this thesis, seven candidate GARCH models are examined to forecast volatility and a final model is chosen according to of highest log-likelihood values. The one-day ahead volatility forecasts are then obtained from the final chosen model and accuracy of this forecast is determined by utilizing comprehensive criteria that include error statistics calculation, Mincer- Zarnowtiz regression and sign tests on the error statistics. Daily VaR forecasts are using volatility forecast of the portfolio, of the benchmark portfolio and by using naïve forecast.estimated for the out-of-sample period and these forecasts are evaluated using a comprehensive VaR evaluation framework.

6 Empirical Results

6.1 Volatility forecasting Models

GARCH (1, 1) models were evaluated on all of the six log-returns series and log-likelihood values are tabulated in table 6 in Appendix B. In some cases, no convergence was reached while fitting GARCH models. In those cases, no value of log-likelihood value was reported in this table.
Table-6 in appendix B shows GARCH_t (1, 1) i.e. GARCH models with student t distribution errors performed best in term of highest log-likelihood values for all series. Also, it performed best in terms of lowest AIC values for almost all series. Therefore, GARCH_t (1, 1) model is used to perform one-day ahead volatility forecasts. Table 2 gives the parameters of the GARCH_t (1, 1) models fitted to the different series.

Table 2: Parameters for the estimated GARCH_t (1, 1) Models

	PORTFOLIO A	BENCHMARK A	PORTFOLIO B	BENCHMARK B	PORTFOLIO C	BENCHMARK C
μ	0,00025	0,00027	0,00032	0,00032	0,00042	0,00041
α0	3,2E-07	4,9E-06	2,4E-07	2,4E-07	8E-07	6,9E-06
α1	0,03868	-0,0152	0,04608	0,0479	0,07828	-5E-05
β1	0,89209	0,01522	0,90142	0,8984	0,83998	0,27042
df	11,2737	10,7789	9,99573	10,2762	8,51358	9,33711

As inferred from the estimated values of the coefficients of the fitted GARCH models, in all portfolio series, except for BENCHMARK A and BENCHMARK C, the value of $\alpha_1 + \beta_1$, is more than 0,9. This implies that these series exhibit high volatility persistence and that the response function of volatility of shocks decays at a relatively slow rate. As the sum tends to 1 the higher is the instability in the variance and shocks tend to persist instead of dying out. For example, in case of PORTFOLIO A series, $\alpha_1 + \beta_1 = 0{,}93$, meaning that 93% of a variance shock remains the next day.

The long-term steady state variance in a GARCH (1,1) model is given by $\sigma^2 = \alpha_0 /(1 - \alpha_1 - \beta_1)$. If we compare the σ^2 value implied by the each of the models above, it comes out to be approx. equal to the square of the standard deviation of the sample series.
The degrees of freedom for the student-t error term, *df*, can be used to infer about the degree of "heavy-tails" in the series. The df value can be used to calculate the fourth moment of the implied by the model ($\left(\mathrm{E}\left[\varepsilon_t^4\right] = 3(df - 2)/(df - 4)\right)$. The calculated value is an inference about the kurtosis of the sample series. For example, in case of the PORTFOLIO A model with estimated df value of approx. 11, the value of implied kurtosis is 3,825 and thus the implied excess kurtosis equals 0,825 , which is sufficiently close to the kurtosis value of the sample, 0,75 as given in the table -1.

6.2 Volatility Forecasts Evaluation

Based on the GARCH_t (1,1) models, as specified in Table 2, one-day ahead forecasts were calculated using rolling sample. In total 250 out of sample forecasts were obtained. The volatility forecasts from the GARCH models are compared with a Naïve model, where the volatility calculated using rolling standard deviation of the past n observations (n equals the number of observations in the in-sample data of the GARCH model). The square of the log returns was used as a proxy for the "true volatility".

In order to compare the predictive accuracy of GARCH models vis-à-vis Naïve models, different error statistics were calculated. One-sided sign tests were also performed for the error statistics, in order to test the statistical significance of the difference in the performance of GARCH and Naïve models. The results are presented in Table 3.

Table 3: Volatility Evaluation

	PORTFOLIO A		BENCHMARK A	
	GARCH	NAIVE	GARCH	NAIVE
MSE	1.80E-08	1.99E-08	1.79E-08	2.02E-08
	-8.348		*-8.095*	
MedSE	3,67E-09	7,09E-09	3,018E-09	6,10E-09
MPE	2,97E-03	3,61E-03	2,99E-06	0,003721
	-8,348		*-8,095*	
AMAPE	0,6139	0,6429	0,6186	0,6508
	-8,222		-7,463	
MME(U)	0,000454	0,000354	0,000420	0,000323
	-14,167		*-12,376*	
MME(O)	0,005624	0,001454	0,005856	0,001525
	-6,831		*-6,451*	
TIC	0,001302	0,001256	0,001296	0,001258
R^2	0,272%	0,759%	0,110%	0,801%

	PORTFOLIO B		BENCHMARK B	
	GARCH	NAIVE	GARCH	NAIVE
MSE	1.54E-08	1.81E-08	1,57E-11	1.82E-08
	-9,360		*-9,866*	
MedSE	2.34E-09	4.57E-09	2,47E-12	4.72E-09
MPE	2.51E-03	3.43E-03	2.54E-03	3,4E-06
	-9,360		*-9,866*	
AMAPE	0,6007	0,6545	0,6060	0,6564
	-8,981		*-8,854*	
MME(U)	0,000398	0,000276	0,000400	0,000284
	-13,914		*-14,546*	
MME(O)	0,005173	0,001492	0,005214	0,001475
	-6,831		*-7,463*	
TIC	0,001307	0,001254	0,001321	0,001263
R^2	0,081%	0,619%	0,036%	0,598%

	PORTFOLIO C		BENCHMARK C	
	GARCH	NAIVE	GARCH	NAIVE
MSE	1,03E-10	1.05E-07	1.15E-07	1.16E-07
	-2,783		*-1,518*	
MedSE	2.29E-08	4.83E-08	2.568E-08	5.51E-08
MPE	6.44E-03	7.01E-03	6.93E-03	0.007367
	-2,783		*-1,518*	
AMAPE	0,5805	0,5931	0,5841	0,5925
	-2,783		-1,012	
MME(U)	0,000567	0,000488	0,000600	0,000536
	-5,060		*-3,668*	
MME(O)	0,008782	0,001974	0,009099	0,001987
	-2,403		*-1,265*	
TIC	0,002062	0,002006	0,002104	0,002053
R^2	0,523%	0,000%	0,831%	0,000%

GARCH model for all series have lower MSE, MedSE, MAE, AMAPE, MME (O) and MME (U) values compared to Naïve model. All one-sided sign test values, except in case of BENCHMARK C are less than -1.65. Therefore, we can reject the null hypothesis at 5% confidence level and may conclude that GARCH models are better than Naïve model on these error loss functions. We observe that in all portfolios, Naïve models have lower TIC values, therefore, indicating that Naïve models are better than GARCH on the basis of this error function. However, one-sided sign tests for this error function shows the opposite (except in BENCHMARK C). We may conclude that although the mean TIC values for Naïve models are lower compared to GARCH models, the GARCH models are better than Naïve models statistically. It is difficult to make a consistent conclusion based on Mincer-Zarnowtiz regression regarding the relative performance of GARCH and Naïve models. Naïve models seem to perform better for PORTFOLIO A, BENCHMARK A, PORTFOLIO B and BENCHMARK B.

In conclusion, GARCH models for almost all portfolios seems to have better predictive accuracy as compare to Naïve models on the basis of most of error measures.

6.3 Value-at- Risk Evaluation

5% Value at Risk estimates using GARCH and GARCH_B models are calculated as product of the volatility forecast, portfolio return and the critical value (which is obtained from the t distribution table corresponding to the degrees of freedom of the errors of the fitted GARCH model). To calculate 5% Value at Risk based on the naïve model, critical value is taken as 1.65. Please refer to figure 4 (in Appendix B) for plots of actual losses and forecasted VaR from the GARCH and Naïve models. Table-4 summarizes results from VaR accuracy tests for the different portfolios.

Table 4: VaR Evaluation

	PORTFOLIO A			***PORTFOLIO B***			***PORTFOLIO C***		
	GARCH	***GARCH-B***	***NAIVE***	***GARCH***	***GARCH-B***	***NAIVE***	***GARCH***	***GARCH-B***	***NAIVE***
Nr of failures	10	9	4	8	9	5	3	3	3
fail. Rate	4,00%	3,60%	1,60%	3,20%	3,20%	2,00%	1,20%	1,20%	1,20%
kupiec test	0,563	1,138	8,185	1,944	1,944	6,071	10,812	10,812	10,812
Uncoditional Coverage(LRu)	0,563	1,138	8,185	1,944	1,944	6,071	10,812	10,812	10,812
Independence (LRind)	7,136	7,297	0,130	7,422	7,422	8,173	0,073	0,073	0,073
Uncoditional Coverage(LRcc)	7,699	8,435	8,315	9,366	9,366	14,244	10,885	10,885	10,885
Loss Fn	10,048	9,044	4,011	8,244	8,215	5,105	3,623	3,537	3,533
$S_{garch,garch-b}$	*15,558*			*15,811*			*15,811*		
$S_{garch,naive}$	*15,685*			*15,685*			*15,558*		

The critical value for LR_{uc} (Likelihood Ratio for unconditional coverage) and LR_{ind} (Likelihood Ratio for independence) is 3, 8415 and for LR_{cc} (Likelihood Ratio for conditional coverage) is 5, 9915. We may conclude from the above results that all the three VaR models (GARCH, GARCH-B and Naïve) in the case of all three portfolios pass Basel back test because the failure rate is less than 5%.
In the case of PORTFOLIO A portfolio, both GARCH models (GARCH and GARCH-B) pass Kupiec/ unconditional coverage test whereas the Naïve model fails. However, opposite is true for the test of independence. For test of correct conditional coverage, both GARCH and Naïve models fail the test.
In the case of PORTFOLIO B portfolio, both GARCH models (GARCH and GARCH-B) pass Kupiec/ unconditional coverage test whereas the Naïve model fails. Test of independence and correct conditional coverage is failed on all three models.

In the case of PORTFOLIO C portfolio, all three models fail Kupiec/conditional coverage test, but pass test of independence. Test of correct conditional coverage is failed on all models.

Due to the lower failure rate for Naïve models for all the portfolios, the corresponding loss function are also lower for Naïve models than that for the GARCH models.
The test statistic, $S_{garch,garch-b}$, for one-sided sign tests comparing GARCH and GARCH-B models, value is greater than -1.65 for all portfolios. Therefore, we can’t reject the null hypothesis at 5% confidence level that the loss function is same for GARCH and GARCH-B (in other words, we can say that GARCH-B is better than GARCH model).
Similarly, The test statistic, $S_{garch,naive}$, for one-sided sign tests comparing GARCH and naive models, value is greater than -1.65 for all portfolios. Therefore, we can reject the null hypothesis at 5% confidence level that the loss function is same for GARCH and GARCH-B (in other words, we can say that Naive model is better than GARCH model).

7 Conclusions & Further Suggestions

In this thesis, GARCH models for volatility and VaR forecasting were analyzed and compared with Naïve model that estimates by moving standard deviation of the past values. The empirical study was done on the investment portfolios of Riksbank. Volatility and VaR estimates for an out-of-sample period of 250 days were forecasted and evaluated.

Based on the empirical results and by employing various statistical tests, GARCH models consistently performed better than Naïve models in forecasting volatility. The portfolio returns distributions showed fat-tails and volatility clustering, which were captured by GARCH models.

Out of the all candidate GARCH models, GARCH_t (GARCH with student t error term) came out to be the best model for all data series based on likelihood Ratio criteria. The results are consistent with conclusion of many other papers. GARCH_t captured the fat tails of the distribution and time varying volatility of the distribution in better way than the other GARCH model that assumes normal error distributions.

The error statistics that were used to evaluate the volatility forecasts by GARCH and Naïve model are statistically robust and objective and thus serve well for the statistical evaluation of the forecasts. However, one needs to consider that these error statistics are limited in their capabilities to give evaluation from economic point of view. The parameters of the fitted GARCH model also depend on the specific characteristics of the portfolio returns series e.g. length of the sample data, time interval between the consecutive data points etc. Therefore, in order to evaluate the models from both statistical and economic point of view, a comprehensive VaR evaluation methodology was employed in the thesis in addition to evaluation based on error statistics.

Although VaR is a function of volatility, the results of VaR evaluation tests were not entirely identical to volatility evaluation tests, as discussed later in this section. The one-day VaR forecasts using Naïve model were relatively higher than those forecasted using GARCH models for 2 out of the 3 portfolios. Based on this, we may say that Naïve model is relatively more conservative than GARCH models. All models passed Basel back-testing for all three portfolios, with Naïve model resulting in lower number of failures (failure being when actual portfolio loss is greater than VaR). If we evaluate VaR forecasts solely on Basel back-testing, we may conclude that Naïve models have relatively over-estimated VaR. Although VaR forecasts by GARCH models were less than VaR forecasts by Naïve models, GARCH based VaR models also passed Basel back-testing.

If we look at the results of the VaR evaluation based on Loss functions, we may conclude that Naïve models have performed better than GARCH models. Over estimation of VaR in Naïve models can easily explain this observation. However, one important point to consider here is the loss function equation. The loss function equation used in the thesis penalizes under-estimation of VaR, not the over-estimation. However, for a bank, both under-estimation and over-estimation of VaR is undesirable whereas a financial regulator will be more worried about the under-estimation of VaR. Hence, in case a different loss function was used that penalized over-estimation, the results would have been different.

The Basel back-testing criterion doesn’t take into consideration the frequency and dependence of the failures. Hence, we need to look at the results of Christoffersen’s Likelihood Ratio tests. The GARCH models have performed relatively better than Naïve models in these tests. Although, for 2 out of 3 portfolios, both GARCH and Naïve models failed on conditional coverage tests, GARCH models passed likelihood tests for unconditional coverage (or Kupiec) and independence.

In conclusion, if the evaluation tests are considered independently, there is definite lack of agreement among the results of the different tests. A practitioner needs to design a specific evaluation framework selecting various relevant tests and designing them according to his or her utility/cost function.

One proposed suggestion for further research is to compare the GARCH based time series models with parametric variance-covariance based methods. Various risk factors, which the given portfolio is sensitive to, are identified and VaR is calculated from the variance and co-variances of these risk factors.

Although GARCH models are parsimonious, and are able to capture the time varying nature of volatility, they fail to capture structural shifts in the time series data that are caused by extreme events. Therefore, one possible extension of GARCH modelling will be to extend it with regime-switching in the models, as done in some studies.Also, the volatility evaluation methodology can be extended further by incorporating better non-parametric tests such as Diebold-Mariano tests.
Regarding VaR evaluation framework, in addition to the statistical significance tests of the models, it is suggested to extend the evaluation with comprehensive economic significance tests by using relevant loss functions according to the utility function of the users.

8 References

- Basel committee in banking supervision, amendment to the capital accord to incorporate market risks, August 1996.
- Berkowitz, Jeremy, and James O'Brien, 2002, How accurate are value-at-risk models at commercial banks? Journal of Finance 57, 1093-1111.
- Berkowitz, J. (2001), "Testing the Accuracy of Density Forecasts," Journal of Business and Economic Statistics
- Burns P., 2002, Quality of Value at Risk via Univariate GARCH, working paper October 2002.
- Christoffersen PF, 1998. Evaluating Interval Forecasts. International Economic Review 39: 841-862.
- Christoffersen PF, Diebold FX, 2000. How Relevant is Volatility Forecasting for Financial Risk Management. Review of Economics and Statistics 82:1-11.
- Cont. R., Statistical Properties of Financial Time Series, August 1999.
- Dave R. D. and Stahl G., On The Accuracy Of VaR Estimates Based On The Variance-Covariance Approach.
- Diebold FX, Mariano R, 1995 July. Comparing Predictive Accuracy. Journal of Business and Economic Statistics 13(3):253-263.
- Engle R, 1982 July. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. Econometrica 50:987-1007.
- Guermat C., Harris D.F Richard, Forecasting Value At Risk Allowing For Time Variation In The Variance And Kurtosis Of Portfolio Returns.
- Jorion, P. (2000), "How Informative are Value-at-Risk Disclosures," manuscript, University of California, Irvine.
- Jorion P, 1997. Value at Risk: The New Benchmark for Controlling Market Risk . Irwin Professional Publishing.
- J.P.Morgan/Reuters, 1996 December. RiskMetrics Technical Document. New York, 4th edn.
- Kupiec PH, 1995 May. Techniques for Verifying the Accuracy of Risk Management Models. Tech. Rep. 95-24, Board of Governors of the Federal Reserve Systems, Federal Reserve Board, Washington, D. C. Finance and Economic Discussion Series.
- Laurent, S. & Peters, J. (2002), A tutorial for G@RCH 2.3, a complete Ox Package for Estimating and Forecasting ARCH Models.
- Lopez JA, 1999. Methods for Evaluating Value-at-Risk Estimates. Federal Reserve Bank of San Francisco Economic Review (2):3-17.
- Manganelli S., Engle R. F. , 2001 Value At Risk Models, Portfolio Bopean Central Bank, Working paper 75.
- Polasek W. and Pojarliev, VaR Evaluations Based On Volatility Forecasts Of GARCH Models.
- Poon and Granger, Forecasting Volatility In Financial Markets; A Review 2002.
- Sarma et al , 2003.Selection of Value-at-Risk models," Journal of Forecasting,John Wiley & Sons, Ltd., vol. 22(4):337-358.
- Vlaar J G Peter, Value At Risk Models For Dutch Bond Portfolios.

- Vlaar J G Peter, Value At Risk Analysis Of Stock Returns Historical Simulation, Variance Techniques Or Tail Index Estimation.
- Wong Sham C M et al , 2003, Market Risk Management In Banks, Implications From The Accuracy Of Value At Risk Forecasts, Journal of forecasting 22 , 2003.
- Zangari, P., 1997. Streamlining the Market Risk Measurement Process, RiskMetrics Monitor, 1, 29-35.

9 Appendix

9.1 Appendix A

9.1.1 Error Statistics for Volatility Forecasts Evaluation

$AMAPE = \frac{1}{250}\sum_{1}^{250}\left|\frac{\hat{\sigma}_t^2 - \sigma_t^2}{\hat{\sigma}_t^2 + \sigma_t^2}\right|$ Where $\hat{\sigma}_t^2$ is the forecasted variance and σ_t^2 is the true variance.

$$MME(O) = \frac{1}{250}\left(\sum_{1}^{O}\left|\hat{\sigma}_t^2 - \sigma_t^2\right|^{0.5} + \sum_{1}^{U}\left|\hat{\sigma}_t^2 - \sigma_t^2\right|\right)$$

$$MME(U) = \frac{1}{250}\left(\sum_{1}^{O}\left|\hat{\sigma}_t^2 - \sigma_t^2\right| + \sum_{1}^{U}\left|\hat{\sigma}_t^2 - \sigma_t^2\right|^{O.5}\right)$$

$MME(O)$ and $MME(U)$ penalizes under-predictions and over-predictions more severely respectively.

$$TIC = \frac{\sqrt{\sum_{1}^{250}\left(\hat{\sigma}_t^2 - \sigma_t^2\right)^2}}{\sqrt{\sum_{1}^{250}\left(\hat{\sigma}_t^2\right)^2} + \sqrt{\sum_{1}^{250}\left(\sigma_t^2\right)^2}}$$

TIC is a scale invariant measure, with value between 0 and lower the value of TIC ism better is the forecasting ability of the model.

9.1.2 Christoffersen's Likelihood-Ratio tests

The test statistic for unconditional coverage, LR_{uc}, is equivalent to that of Kupiec test as described above. The test for independence evaluates whether the failures are independently distributed over time and hence are unpredictable.
Based on the failure process, $\{f_t\}$, as described above, we can define the following transition probability matrix for the failure process.

$$\pi = \begin{bmatrix} 1-\pi_{01} & \pi_{01} \\ 1-\pi_{11} & \pi_{11} \end{bmatrix}$$

Where $\pi_{ij} = \Pr(f_t = j \mid f_{t-1} = i) \;\; i, j = 0,1$
i.e. conditional probability of state i being followed by state j. The likelihood function for the process is then $(1-\pi_{01})^{n_{00}}\pi_{01}^{\;n_{01}}(1-\pi_{11})^{n_{10}}\pi_{11}^{\;n_{01}}$
where n_{ij} = number of times state i being followed by state j.The LR statistic, LR_{ind} is given by
$LR_{ind} = 2[\ln L_A - \ln L_O]$

where

$$L_A = (1-\pi_{01})^{n_{00}} \pi_{01}^{\ n_{01}} (1-\pi_{11})^{n_{10}} \pi_{11}^{\ n_{01}}$$

$$L_O = \left(1-\hat{\pi}\right)^{n_{00}} \hat{\pi}^{\,n_{01}} \left(1-\hat{\pi}\right)^{n_{10}} \hat{\pi}^{\,n_{01}}$$

$$\pi_{ij} = n_{ij} / (n_{i0} + n_{i1})$$

$$\hat{\pi} = n_{01} + n_{11} / (n_{00} + n_{01} + n_{10} + n_{11})$$

Under null hypothesis, the above test statistic has a χ^2 distribution with one degree of freedom. The test for correct conditional coverage is a joint test of the above two tests. Therefore, the LR test statistic, LR_{cc} is equal to $LR_{uc} + LR_{ind}$. Under null hypothesis, the above test statistic has a χ^2 distribution with two degrees of freedom.

9.1.3 Hypothesis tests

9.1.4 Jarque-Bera test of normality

The Jarque-Bera test is done to test whether the skewness and kurtosis of the distribution corresponds to that of a normal distribution and the test statistic is given by

$$\frac{(T-k)}{6}\left(S^2 + \frac{(K-3)^2}{4}\right)$$

where *T* is the number of observations, *k* is the number of estimated parameters, *S* is the skewness and *K* is the kurtosis. The larger the value of test statistic, the lower the probability is that the given series is drawn from a normal distribution. The test statistic of the Jarque-Bera test is χ^2 distributed with 2 degrees of freedom under the null hypothesis, that the series is normally distributed.

9.1.5 Ljung-Box-Pierce Q-test

This test is performed to test whether the series has significant autocorrelation or not. The test statistic is given by

$$Q_k = T(T+2)\sum_{i=1}^{k} \frac{r_i^2}{T-i}$$

Where T is the number of samples, k is the number of lags and r_i is the i th autocorrelation. The larger the value of test statistic, the lower the probability is that the given series has autocorrelation. The test statistic Q_k is χ^2 distributed with *k* degrees of freedom under the null hypothesis of no autocorrelation.

9.2 Appendix B-Tables and Figures

Figure 1: Log-returns series

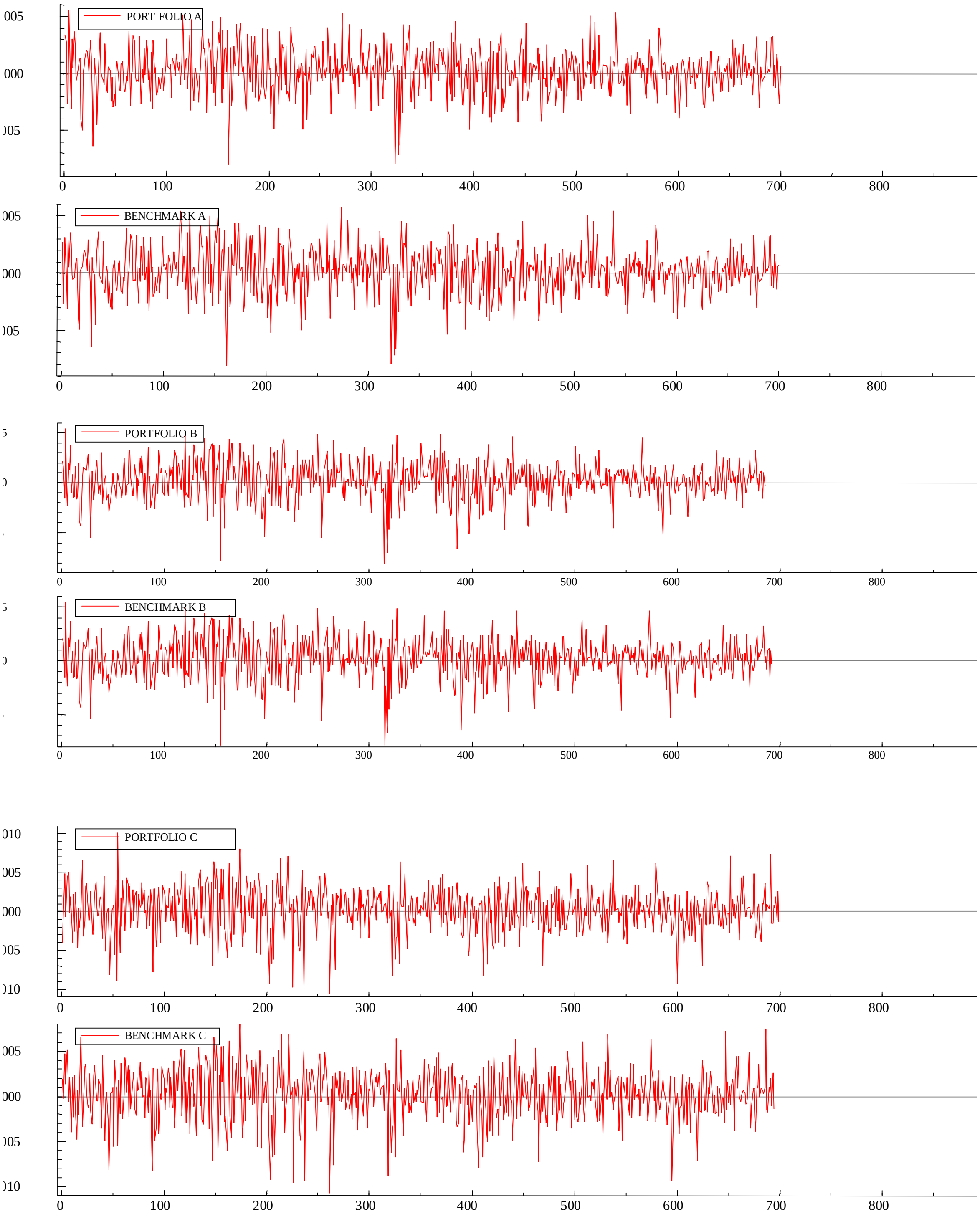


Figure 2: Squared log-returns

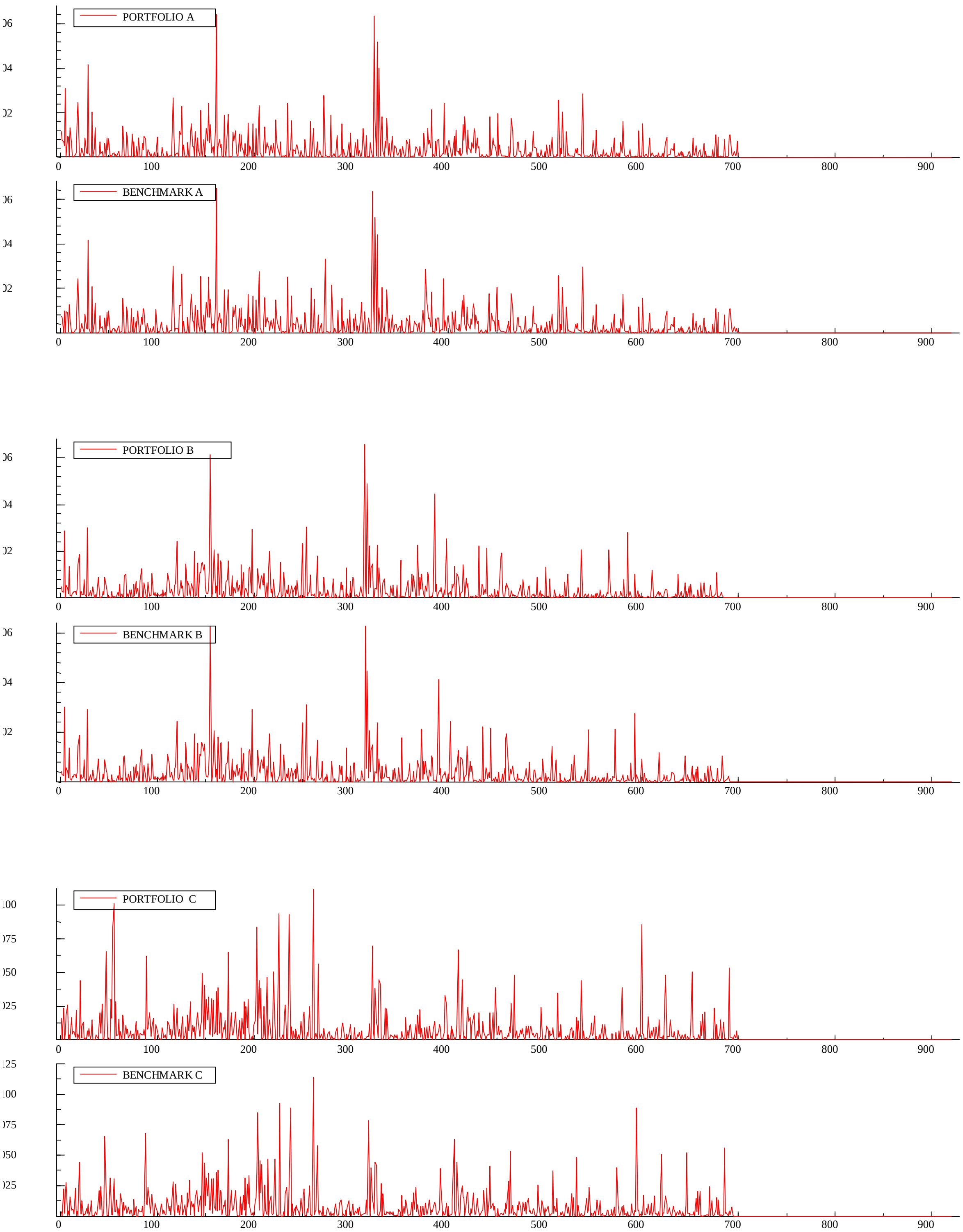


Figure 3: Density and QQ plots

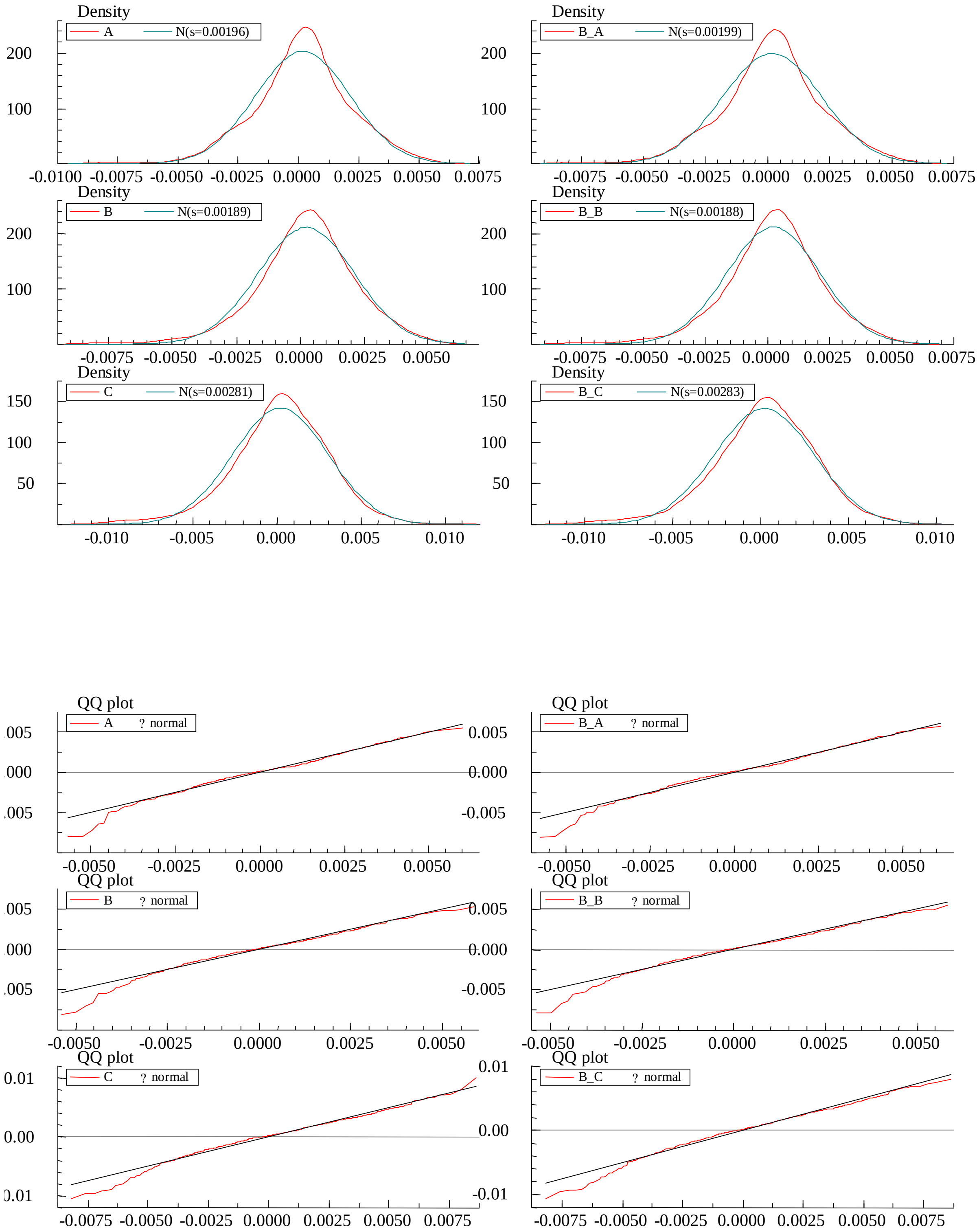
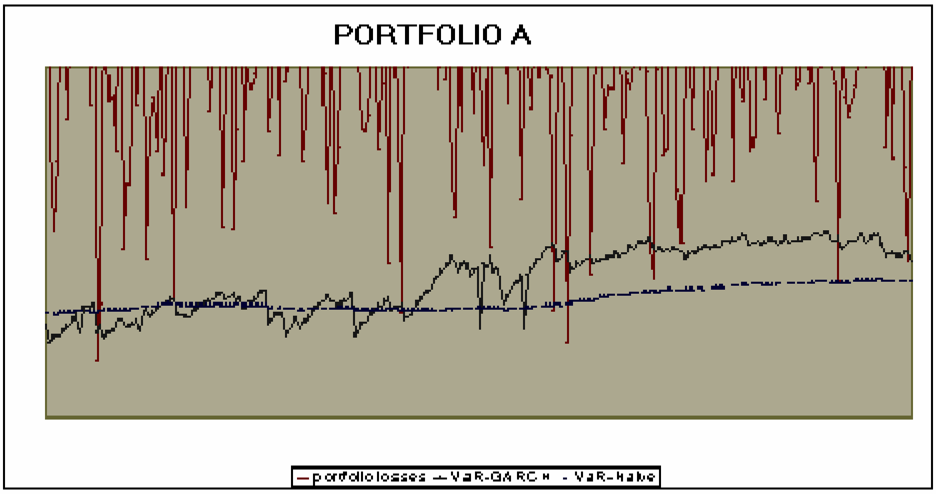
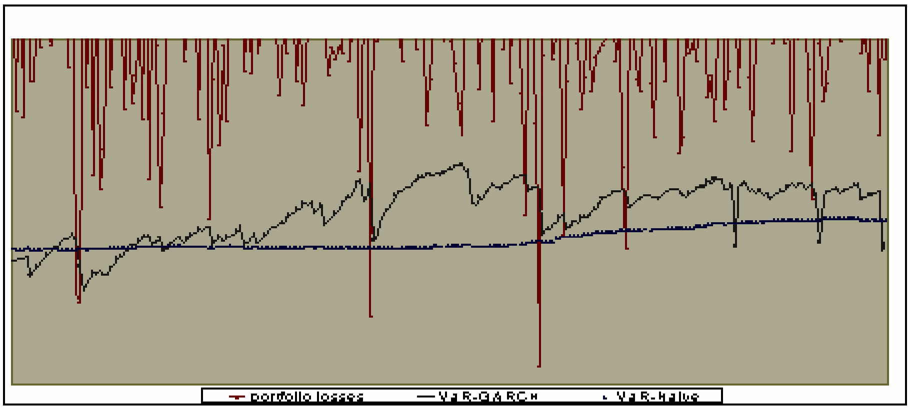


Figure 4: Portfolio Losses and VaR

PORTFOLIO A



PORTFOLIO B



PORTFLIO C

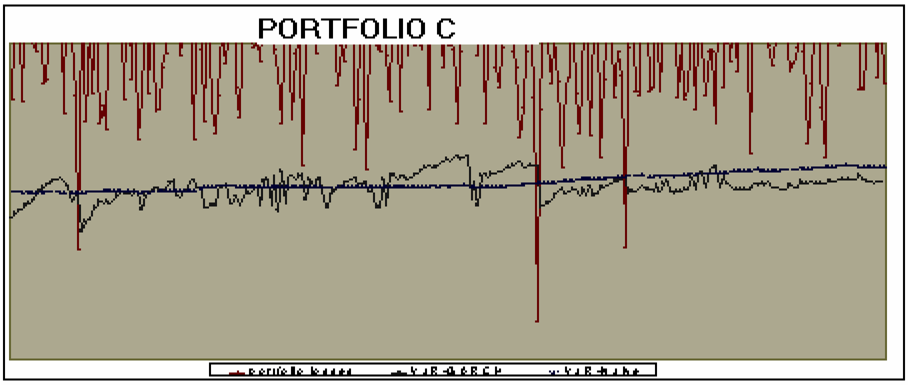


Table 5: Ljung-Box-Pierce Q-test

Series #1/1: PORTFOLIO A
Q-Statistics on Raw data
Q (10) = 9,43715 [0,491179]
Q (20) = 23,7364 [0,254097]
Q (50) = 52,3696 [0,382168]

Q-Statistics on Squared data
Q (10) = 25,3724 [0,00468251]
Q (20) = 37,413 [0,0104362]
Q (50) = 56,9089 [0,23349]

Series #1/1:IBENCHMARK A
Q-Statistics on Raw data
Q (10) = 6,84464 [0,740026]
Q (20) = 30,2381 [0,0660862]
Q (50) = 68,9269 [0,0391497]

Q-Statistics on Squared data
Q (10) = 12,1007 [0,278372]
Q (20) = 19,0197 [0,520547]
Q (50) = 38,8704 [0,872901]

Series #1/1: PORTFOLIO B

Q-Statistics on Raw data
Q (5) = 1,24099 [0,940885]
Q (10) = 5,90924 [0,82283]
Q (20) = 18,8945 [0,528694]
Q 50) = 48,841 [0,519919]

Q-Statistics on Squared data
Q (5) = 7,5567 [0,182418]
Q (10) = 16,2504 [0,0926863]
Q (20) = 23,1894 [0,279591]
Q (50) = 37,9309 [0,894776]

Series #1/1: BENCHMARK B
Q-Statistics on Raw data
Q (5) = 3,63516 [0,603041]
Q (10) = 10,1387 [0,428413]
Q (20) = 21,8573 [0,348302]
Q (50) = 54,9354 [0,29309]

Q-Statistics on Squared data
Q (5) = 10,0176 [0,0747382]
Q (10) = 18,7785 [0,0431675]
Q (20) = 27,5731 [0,119897]
Q (50) = 44,7084 [0,685003]

Series #1/1: PORTFOLIO C
Q-Statistics on Raw data
Q (5) = 2,56043 [0,767368]
Q (10) = 12,5456 [0,2502]
Q (20) = 25,3432 [0,188617]
Q (50) = 62,026 [0,118349]

Q-Statistics on Squared data
Q (5) = 9,82142 [0,0804562]
Q (10) = 15,7789 [0,106138]
Q (20) = 22,6798 [0,30479]
Q (50) = 44,1006 [0,707974]

Series #1/1: BENCHMARK C
Q-Statistics on Raw data
Q (5) = 4,33893 [0,501719]
Q (10) = 15,7855 [0,105939]
Q (20) = 27,9033 [0,111709]
Q (50) = 50,5343 [0,452278]

Q-Statistics on Squared data
Q (5) = 15,5836 [0,00813929]
Q (10) = 19,5632 [0,033666]
Q (20) = 23,5363 [0,263234]
Q (50) = 48,8368 [0,520089]

Table 6: Log-Likelihood Values for Different GARCH models

	GARCH	GARCH_t	AGARCH	ATGARCH	EGARCH	GARCHM	TGARCH
PORTFOLIO A	2126,44	2129,42	2127,42	2123,77	2125,48	2126,44	2126,57
BENCHMARK A	2112,42	2115,45	2112,43	2112,46	-	2112,42	2113,63
PORTFOLIO B	2060,57	2068,38	2065,87	2067,31	-	2063,89	2064,42
BENCHMARK B	2100,27	2104,49	2101,28	2103,09	2100,6	2100,27	2100,52
PORTFOLIO C	1965,35	1970,19	1965,71	1965,95	1965,6	1965,35	1965,49
BENCHMARK C	1942,84	1943,42	1941,43	1943,28	1944,03	1942,84	-