ERASMUS UNIVERSITY ROTTERDAM

Erasmus School of Economics

Bachelor Thesis [International Bachelor Economics and Business Economics]

Value-at-Risk: An Optimal Application

Name: Pranav Muralidhar

Student ID: 526653

Supervisor: Professor Laurens Swinkels

Second Assessor: Professor Sebastian Gryglewicz

Date final version: 28-07-2022

TABLE OF CONTENTS

Abbreviations……………………………………………………………………………………………3

1. Introduction ……………………………………………………………………………………..4

2. Theoretical Framework………………………………………………………………………...6

2.1. Value-at-Risk………………………………………………………………………….6

2.2. Ex-post volatility………………………………………………………….............7

2.2.1. Overview………………………………………………………………….7

2.2.2. Historical mean model……………………………………………………9

2.2.3 Autoregressive model……………………………………………………..9

2.2.4. EWMA model………………………………………..………………………10

2.2.5. GARCH model………………………………………………………….11

2.2.6. GJR-GARCH model…………………………………………………….12

2.3. Implied volatility……………………………………………………………………..12

3. Data……………………………………………………………………………………………..14

4. Methodology…………………………………………………………………………………17

4.1. Optimization………………………………………………………………………….17

4.1.1. EWMA………………………………………………………………………17

4.1.2. GARCH……………………………………………………………………18

4.1.3. GJR-GARCH…………………………………………………...............18

4.2. Forecasting and the construction of the VaR…………………………………………18

4.3. Evaluation of the VaR……………………………………………………………..19

5. Results…………………………………………………………………………………………..20

5.1. Optimized specifications……………………………………………………………..20

5.2. Daily VaR using ex-post volatility………………………………………………………22

5.2.1. Results……………………………………………………………………..22

5.2.2. Comparing the models…………………………………………………….26

5.3. Comparison of daily VaR modelling with ex-ante and ex-post volatility……………………27

6. Discussion……………………………………………………………………………………32

7. Bibliography…………………………………………………………………………………….35

Appendix………………………………………………………………………………………..38

Abbreviations

For ease of reading, a comprehensive list of abbreviations and their extended form is provided below.

1. AR - Autoregressive
2. ARCH – Autoregressive Conditional Heteroskedasticity
3. CBOE – the Chicago Board Options Exchange
4. DJIA – Dow Jones Industrial Average
5. EWMA – Exponential Weighted Moving Average
6. FTSE – Financial Times Stock Exchange Group
7. GARCH – Generalized Autoregressive Conditional Heteroskedasticity
8. GJR-GARCH – Glosten-Jagannathan-Runkle GARCH
9. HSI – Hang Seng Index
10. iGARCH – Integrated GARCH
11. KOSPI – Korean Composite Stock Price Indexes
12. MAE – Mean Absolute Error
13. MAPE – Mean Absolute Percentage Error
14. NASDAQ – National Association of Securities Dealers Automated Quotations
15. NIFTY – National Stock Exchange Fifty
16. OLS – Ordinary Least Squares
17. RMSE – Root Mean Squared Error
18. SMI – Swiss Market Index
19. S&P – Standard & Poor's
20. TSX – Toronto Stock Exchange
21. VaR – Value-at-Risk

1. Introduction

An extensive analysis of any investment is incomplete without consideration of the risk associated with it. An understanding of the risk profile of a given portfolio position is of particular importance for financial institutions as negligence in this regard can result in severe repercussions that threaten the financial health of the institution itself and, possibly, even the national and global economy. This has been evident in the financial crises of the past, where poor risk management practices have threatened global financial stability. Consequently, the attention devoted to risk management in financial institutions has greatly increased following these previous crises.

The introduction of the Basel Regulations in 1988 is also responsible for the increased interest in financial risk management. The financial crisis of 1987 prompted the introduction of the Basel accord, which sought to formalise the process of risk management among financial institutions. This was to be achieved by incorporating measures to improve the capital standards among banks within the G-10 countries. However, following criticism on the regulations that were initially introduced, a modified version of the agreement was eventually implemented in 1995, which formally incorporated the requirement for banks to assess their liquidity risk through the use of Value-at-Risk (VaR) models (van den Goorbergh & Vlaar, 1999).

Following its establishment as part of the Basel Regulations, the VaR has become a key instrument for financial institutions to quantify the market risk associated with their portfolio positions. Although different methods of computing the VaR exist, this paper focuses on the variance-covariance method. Within this framework for the VaR, an integral component is the forecasted volatility used in its calculation. Although an abundance of models and techniques are available to forecast volatility, there is no clear consensus on which method is most suited for the modelling of the VaR. The approach that is used to estimate volatility can have a material impact on the computed VaR, which can in turn significantly influence a financial institution's investment decisions and their ability to adequately manage the risk of their investments.

Two fundamental approaches to estimating volatility can be identified: backward-looking (ex-post) volatility forecasts are obtained by estimating statistical models using historical data, while forward-looking (ex-ante) volatility estimates are acquired by determining the implied volatility associated with an asset. In theory, the use of implied volatility offers practical advantages over the use of backward-looking estimates of volatility. For one, financial organisations require a future estimate of volatility upon which to base their financial decisions and the implied volatility of an asset provides such an estimate. Backward-looking volatility estimates are based on prior historical data, which may not necessarily be representative of future financial conditions. However, implied volatility estimates may suffer from measurement error, expiration effects, and speculative noise, all of which can distort its value (Day and Lewis, 1992). Additionally, the presence of a variance risk premium may also affect implied volatility estimates. Hence, despite the underlying theory favouring the use of implied volatility estimates in financial decision-making, its practical efficacy in modelling the VaR must be determined.

Given the uncertainty regarding the best approach and method to forecast volatility for the purposes of VaR modelling, the following research question is considered in this paper:

Which approach to modelling volatility results in the most accurate one-day VaR?

The central research question with which this paper is concerned assumes a holding period of one day for the computation of the VaR. Although the Basel regulations require the VaR to be calculated using a 10-day holding period, banks generally tend to modify their portfolio positions on a daily basis. As such, this paper focuses on modelling the one-day VaR. By determining the answer to the aforementioned research question, financial institutions can be advised on the optimal application of the VaR measure. This may enable these institutions to obtain a more accurate assessment of the risk profile associated with their portfolio positions, which can then prevent excessive risk from being taken on. This could potentially reduce the threat of unforeseen losses leading to bankruptcy and, thus, promote greater financial stability for the company and the global economy. Additionally, a more accurate overview of the organisation's risk exposure can better inform financial decisions and potentially also allow for more profit-maximizing decisions to be undertaken.

In the academic literature, the application of volatility forecasting techniques to historical financial data has been extensively investigated. Despite this, there remains uncertainty on the optimal model to use for forecasting volatility due to the conflicting results found in the literature. In their evaluation of eight separate volatility forecasting techniques, Brailsford and Faff (1996) utilize daily data for an index on the Australian Stock Exchange and do not find conclusive evidence in favour of one particular model. On the other hand, Akgiray (1989) finds support for the GARCH model being superior to the ARCH, EWMA, and historical mean model. Contrary to these findings, Dimson and Marsh (1990) determined that an autoregressive model with a single lag of the prior observed volatility is superior in its ability to predict conditional volatility compared to more sophisticated models.

However, these prior studies have relied on the use of error statistic(s) to evaluate the predictive ability of the statistical models in question. While this approach has its merits, its relevance for financial institutions is questionable. The most accurate volatility forecasting model (determined using a loss function) may not necessarily offer the optimal approach for modelling the VaR. Hence, these statistical models must be assessed in a risk management framework for the results to be applicable for financial institutions.

Additionally, the analysis conducted in each of these previous studies is centred upon a limited number of indices and, so, the general applicability of their results can be questioned.

The topic of comparing forward-looking and backward-looking estimates of volatility has also been previously studied in the academic literature. Day and Lewis as well as Giot (2005) analyse the information content of implied volatility estimates in forecasting future volatility, i.e., they utilize implied volatility as an independent variable in various statistical models in order to predict the conditional volatility. In this approach, the implied volatility is treated as a supplemental measure that may contain useful information in predicting the conditional volatility, rather than being used directly as a measure of conditional volatility.

To overcome these limitations, the analysis conducted in this paper involves modelling the daily VaR for 13 indices using a multitude of approaches to forecast volatility. In this manner, this study provides a practical assessment of different volatility forecasting techniques by examining their performance in a risk management framework. Furthermore, the use of 13 indices, most of which are geographically dispersed, provides a generally applicable framework for modelling the daily VaR. Finally, this study also contributes to the current literature by examining the efficacy of using ex-ante estimates of volatility in comparison to ex-post estimates of volatility when modelling the one-day VaR. In this context, this paper uniquely considers the implied volatility as a measure of conditional volatility, rather than as supplemental information that may be useful in predicting conditional volatility.

The theory underlying ex-post and ex-ante volatility is described in the Theoretical Framework, which also includes a review on relevant previous literature written on the topic of volatility forecasting and the VaR. Following this, the Data section highlights the sources as well as the type of data used while the Methodology section documents the process of using this data to construct the VaR and evaluate its performance. The Results section contains the outcome of this analysis and highlights the optimal approach to modelling the one-day VaR. Finally, the Conclusion section contains an interpretation and examination of the primary insights derived from the Results section.

2. Theoretical Framework

2.1. Value-at-Risk

The VaR measures the potential loss associated with a portfolio position over a given time period with a prespecified probability of the realized loss exceeding this amount (Hendricks, 1996). The probability with which the loss on the investment exceeds the VaR is determined by the choice of the confidence level, which is typically chosen to be either 1% or 5%. The VaR accounts for market or price risk, i.e., the risk that the price (and, hence, the value) of the portfolio declines due to changes in the interest rates, foreign exchange rates, equity prices, or commodity prices (Duffie & Pan, 1997).

Under the variance-covariance method for computing the VaR, the portfolio returns are assumed to be normally distributed. The VaR can then be defined as shown in Equation (1).

$$VaR_t(\alpha) = z_\alpha \sigma_t \tag{1}$$

where α represents the chosen significance level, z_α denotes the critical value of the standard normal distribution associated with the chosen significance level, and σ_t refers to the forecasted volatility for the current period.

2.2. Ex-post volatility

2.2.1. Overview

The estimate for future volatility is a paramount component in the computation of the VaR (as it is specified in the variance-covariance method). One approach to forecasting volatility is to estimate a statistical model using historical data. A common characteristic found in financial data is the phenomenon of volatility clustering, i.e., periods of high volatility tend to be followed by subsequent periods of high volatility, and periods of low volatility tend to be followed by periods of low volatility (Engle, 2001). In other words, there is a level of predictability associated with a stock or portfolio's volatility, which allows a portfolio's volatility to be decomposed into two separate components: a predictable and an unpredictable component (Pagan and Schwert, 1990). The predictable component, known as conditional volatility, can then be forecasted through the use of different statistical models.

One of the difficulties associated with volatility is that it is latent, i.e., it cannot be observed. As such, it is necessary to proxy volatility. To this end, Day and Lewis (1992) proxy *ex-post* volatility by relying on the square of the weekly observed returns. In a similar manner, this paper also makes use of the squared returns to proxy volatility, but utilizes daily returns as opposed to weekly returns for this purpose.

Equation (2) defines the daily return as the sum of the average historical return of the index and an innovation term, where the innovation is the product of σ_t, i.e., the conditional volatility, and z_t, an independent mean zero, stochastic error process distributed independently and identically (hereafter referred to as "standardized residual"). Assuming the average historical return of the asset to equal zero allows for Equation (2) to be simplified to Equation (3). Taking the square and the expectation of both sides of Equation (3) demonstrates that the conditional variance of the asset equals the expectation of the squared daily returns (as shown in Equation (4)).

$$r_t = \mu + \varepsilon_t = \mu + \sigma_t z_t \tag{2}$$

$$r_t = \sigma_t z_t \tag{3}$$

$$E(r_t^2) = E(\sigma_t^2 z_t^2) = \sigma_t^2 \tag{4}$$

Brailsford and Faff apply a similar approach to estimating *ex-post* volatility and use this to assess the accuracy of the volatility estimates obtained from eight different classes of statistical models. For a fair assessment of the predictive ability of each of these models, their forecasts are evaluated out-of-sample. To this end, the authors use a subset of their observations in order to estimate each of the models and then forecast the volatility for the subsequent periods using a rolling window. The accuracy of the forecasted volatility obtained from different models is then assessed using different error statistics, namely the mean error, the MAE, the RMSE, the MAPE, and two other loss functions created by the authors. When using the RMSE statistic, it is found that the historical mean model and an autoregressive model of order one are the best performing models. However, when using the MAE and the MAPE error statistics, they find the GJR-GARCH to be the most preferred model with the traditional GARCH being only marginally less accurate than this model. The conclusions arrived at using the other loss functions are given little credence by the authors and, so, are not discussed here.

Hansen and Lunde (2005) evaluate the predictive ability of 330 ARCH-type models using a multitude of error statistics. Like Brailsford and Faff, the forecasting accuracy of these models are evaluated out-of-sample and a subset of the total observations are used to estimate the model. The scope of this paper includes an analysis of these different ARCH-type models using exchange rate data and data on the stock returns for IBM. When using exchange rate data, no clear evidence could be found that indicated the GARCH (1,1) model could be outperformed by other models. However, in the case of the IBM stock returns, the results indicated that models that accommodated an asymmetric response to positive and negative innovations outperformed the GARCH (1,1). This finding aligns with Brailsford and Faff's conclusion that the GJR-GARCH model outperforms the GARCH (1,1) model.

Unlike the above two papers, a study by Walsh and Yu-Gen Tsou (1998) finds evidence in favor of the EWMA model being the most accurate at forecasting volatility. In their analysis, the authors use data from three Australian value-weighted indices and compare the performance of four different volatility forecasting models: naïve forecasting, improved extreme-value, EWMA and GARCH. Similar to Brailsford and Faff and Hansen and Lunde, these models are estimated on a rolling-window basis and are evaluated out-of-sample. Using this methodology, the authors find support for the EWMA model being the most accurate model for forecasting volatility, with the GARCH model following closely. Additionally, it is also found that the EWMA and GARCH models dominate the naïve forecasting and extreme-value models for all loss functions used.

Based on the results from prior literature, the following models are included for analysis in this paper: historical mean, a modified version of the autoregressive model, exponential weighted moving average, GARCH, and GJR-GARCH. The results of the aforementioned studies can be seen to present conflicting evidence to some extent on the optimal statistical model to use when predicting volatility. However, Brailsford and Faff as well as Hansen and Lunde both find evidence in favor of the GJR-GARCH model being the most accurate statistical model for predicting volatility. The conclusions of these paper can be extended to the modelling of the daily VaR given that there is a direct relationship between the forecasted volatility and the computed VaR. Hence, the following hypothesis is derived: *the GJR-GARCH model results in the most accurate modelling of the daily VaR in comparison to the alternate statistical models*.

An extensive theoretical exploration of the various statistical models included for analysis follows.

2.2.2. Historical mean model

The historical mean model offers a nonparametric approach to modelling volatility. The average of the previously observed volatilities over some prior *k* periods provides the estimate for the current conditional volatility. This average value then represents the forecasted volatility for the following period or for the future periods under consideration. Following Yu-Gen and Tsou, a representation of this model can be found in Equation 5.

$$\sigma_t^2 = \frac{1}{t-1} \sum_{i=t-k-1}^{t-1} (r_i - \bar{r})^2 \tag{5}$$

2.2.3. Autoregressive (AR) model

The autoregressive model involves a regression of the observed volatility on its immediate lag, i.e., the volatility observed in the preceding period. Equation (6) depicts a modified version of this model, where the current conditional volatility is a function of the previous days squared return. In effect, this model represents an ARCH (1) specification. The coefficients in this model can be estimated using OLS.

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

Following the estimation of the coefficients in Equation (6), the current level of the volatility can be predicted.

2.2.4. EWMA Model

A simplistic approach to modelling volatility would involve the computation of an equally-weighted average of all prior observations of the realized volatility. However, the presence of volatility clustering in financial data suggests that recent information is likely to be more relevant when forecasting volatility as compared to observations further in the past. The EWMA model incorporates this notion by weighing recent observations more heavily than older observations. As its name suggests, the weights applied to the observations decline exponentially over time.

$$\sigma_t^2 = (1-\lambda)\Sigma_{j=0}^{\infty}\lambda^j(r_{t-j}-\bar{r})^2 \tag{7}$$

The model is represented in Equation (7) and can be shown to simplify to the specification in Equation (8) as shown by Bollersev (2008).

$$\sigma_t^2 = \lambda\sigma_{t-1}^2 + (1-\lambda)r_{t-1}^2 \tag{8}$$

Equation (8) depicts a limiting case of the iGARCH (1,1) model (see section A1 in the Appendix), where restrictions are imposed on the parameters of the model. Specifically, if $\omega = 0$, $\alpha = 1 - \lambda$, and $\beta = \lambda$, then the EWMA model can be shown to be equivalent to an iGARCH (1,1) specification. The weights are represented by λ, which can also be referred to as the decay factor or the smoothing factor (from here on, the terms “decay factor”, “smoothing factor”, “smoothing exponent”, and “lambda” will be used interchangeably to refer to the weight given by λ). The value of this decay factor must lie between 0 and 1, with a larger decay factor indicating that the current volatility is influenced to a greater extent by the prior realized volatility.

The optimal value for the decay factor can be determined by estimating an iGARCH (1,1) model and setting the value of the constant term to 0. The remaining coefficients are then estimated using Maximum Likelihood, i.e., by finding the values of the coefficients that maximize a log-likelihood function. The value of lambda found through this process can then be considered to be its optimal value. However, in practice, the decay factor is commonly set to 0.94 based on the methodology applied by *RiskMetrics,* a risk management tool introduced by J.P. Morgan (1996). Despite the popularity of this approach, a smoothing factor of 0.94 may not always be appropriate for data with different variance structures.

2.2.5. GARCH Model

The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model was introduced by Bollerslev (1986) as an extension of the ARCH model proposed by Engle (1982). In an ARCH specification, the conditional variance is modelled as a linear function of the square of the previously observed innovation terms. In addition to this, the GARCH model incorporates lagged conditional variance terms, making the current forecast of volatility a function of the previously observed values of volatility and the lagged error terms. The GARCH model allows for a more parsimonious model to be estimated as an ARCH model typically requires several lags in order to capture the variance structure present in the data. Like the ARCH model, the GARCH model is mean reverting and has constant *unconditional* variance.

$$r_t = \sigma_t z_t = \varepsilon_t \tag{9}$$

$$\sigma_t^2 = \omega + \Sigma_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \Sigma_{k=1}^{q} \alpha_k r_{t-k}^2 \tag{10}$$

As shown in Equations (9) and (10), the GARCH model requires the joint estimation of a conditional mean and conditional variance equation. The coefficients for these equations are estimated using Maximum Likelihood. Prior to the estimation of these equations, their specification, i.e., the number of lagged terms, must be determined. Prior academic literature typically utilizes a GARCH (1,1) model as it is assumed that it can sufficiently capture the variance structure in the data. However, the GARCH (1,1) may not be the optimal specification in all circumstances and, as such, the optimal lag order for the GARCH model must be determined.

A possible approach that can be used to determine the optimal specification of the GARCH model is to use Information Criterion, such as Akaike Information Criteria, Bayensian (Schwarz) Information Criterion, etc. However, Brooks and Burke (2003) as well as Javed and Mantalos (2013) show that using Information Criteria to identify the optimal number of lags in conditional variance models would lead to biased results. In such cases, the Information Criteria will always prefer the most parsimonious model, regardless of the true variance structure present in the data. Although modified versions of the Information Criteria that are suitable for conditional variance models exist, their use has not been widely accepted and the results from these papers are conflicting.

Instead, a general portmanteau test can be used to determine if the GARCH model is appropriately specified. A test of this form is the Li-Mak test (Li & Mak, 1994), which tests for the presence of ARCH effects in the standardized residuals of the model. If the null hypothesis of no autocorrelation in the squared and standardized residuals is not rejected, then it can be concluded that the GARCH model has been appropriately specified (Lundbergh & Teräsvirta, 2002).

2.2.6. GJR-GARCH Model

The traditional GARCH model has a symmetric response to both positive and negative innovations as it only accounts for the magnitude of the residuals and not their direction. However, typically, the reaction to negative innovations is more severe than to positive innovations, i.e., the increase in volatility following a negative innovation is greater than for a positive innovation. This empirical regularity was first observed by Black (1976) and is referred to as the leverage effect. To accommodate this phenomenon, the GJR-GARCH model introduced by Glosten, Jagannathan, and Runkle (1993) employs an additional term in its specification that enables it to capture the leverage effects of negative innovations. Thus, by allowing for an asymmetric response to positive and negative innovations, the model can (theoretically) better fit the data.

$$r_t = \sigma_t z_t = \varepsilon_t \tag{11}$$

$$\sigma_t^2 = \omega + \Sigma_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \Sigma_{k=1}^{q} \alpha_k r_{t-k}^2 + \Sigma_{k=1}^{q} \gamma_k r_{t-k}^2 I_{r_{t-k<0}} \tag{12}$$

Similar to the GARCH model, the coefficients of the model are estimated using Maximum Likelihood and the optimal lag order can be determined using the Li-Mak test.

2.3. Implied volatility

Implied volatility represents the market's assessment of the expected future volatility of a given underlying asset (a stock market index in this context). This estimate is derived by relating the price of financial options (namely, call and put options) on the underlying stock market index and a given option pricing formula, typically the Cox-Ross-Rubinstein binomial model or the Black-Scholes model (Mayhew, 1995) (see Section A2 in the Appendix). The values of all the parameters (except the volatility) used to price an option in the Black-Scholes model are readily available including the option price as well, thus allowing the formula to be inverted and the implied volatility to be calculated. If the option pricing formula is assumed to be correct, then the implied volatility calculated in this manner represents the market's expectation of the future volatility of the underlying index.

As the implied volatility of an index represents a forward-looking estimate of volatility, it is expected to overcome the difficulties of using historical data to forecast volatility, namely the potential issue of historical data not being representative of future financial conditions. These two approaches to estimating volatility were examined by Day and Lewis in their 1992 paper, where they examined the information content of implied volatility estimates obtained from call options on the S&P 100 and compared this with the predictive ability of GARCH and EGARCH models. In order to do this, the authors estimated GARCH (1,1) and EGARCH (1,1) specifications along with an extension of these specifications in which an additional exogenous term that represents the derived implied volatility is included. Given that the GARCH (1,1) and EGARCH (1,1) specifications are nested within the extended models that include the implied volatility term, a likelihood ratio test was then used to assess the incremental information content offered by including the implied volatility in the model. Using this methodology, it is found that implied volatility offers additional predictive ability when modelling conditional volatility beyond that which can be obtained by simply modelling it with GARCH and EGARCH specifications. However, the authors find that implied volatility on its own is not sufficient to model the conditional volatility appropriately. The out-of-sample predictive ability of these models is also assessed by regressing proxies for *ex-post* volatility on the forecasts obtained from alternative models, but no conclusive results were obtained from this.

The primary findings of the above paper strongly align with those found by Giot, who applies a similar methodology to assess the information content of the implied volatility obtained from the VXO, VXN, and VIX indices in the context of daily VaR application. The current volatility is forecasted using an EWMA model, a GJR-GARCH model, as well as an extension of these models that includes the addition of a lagged implied volatility term. Using the estimated volatility from each of these approaches, the daily VaR is modelled for three distinct subperiods that collectively range from 01/08/1994 to 31/01/2003. By applying this methodology, the GJR-GARCH model with the inclusion of the lagged implied volatility term is found to perform the best in its ability to model the daily VaR. Much like Day and Lewis, Giot concludes that using implied volatility can improve the modelling of the VaR, but is not sufficient on its own to model it.

It is difficult to predict the efficacy of directly using implied volatility estimates in modelling the VaR since the aforementioned papers utilise implied volatility in an alternate methodology. From a theoretical standpoint, implied volatility would be expected to outperform backward-looking estimates of volatility in modelling the VaR since it represents the entire market’s expectation of future volatility. This would suggest that the computed VaR is more likely to accurately align with future market conditions as opposed to when backward-looking estimates of volatility are used. Furthermore, the results from Day and Lewis as well as Giot fail to provide conclusive evidence on whether including only an implied volatility term in a regression specification can outperform traditional statistical models. However, given the evidence suggesting that implied volatility on its own is not optimal for modelling the conditional volatility or the daily VaR, a second hypothesis is constructed as follows: *using ex-post measures of volatility results in a more accurate modelling of the daily VaR as compared to using implied volatility*.

3. Data

This paper utilizes historical data for a set of stock market indices that are selected based on the availability of a volatility index which measures market volatility using the given index as its underlying asset. Following this approach, data on thirteen indices and their corresponding volatility index is retrieved from Bloomberg and where this is not possible, from Refinitiv Eikon. Both Bloomberg and Refinitiv Eikon are financial software companies that provide comprehensive financial data on a range of securities. The extensive coverage and duration of data available makes these sources suitable for the purposes of this study.

The extent to which historical data is available for the chosen indices differs and, as such, the number of observations also varies between indices. The first observation corresponds to the earliest date for which data is available for all indices except the S&P 500 and the Dow Jones Industrial Average (DJIA). For the S&P 500, the data is truncated such that the first observation corresponds to March 4th, 1957. This is done because the S&P 500, as it is currently known, was introduced on this date. Prior to this, the index only contained 233 companies rather than 500 companies as it currently does (S&P Dow Jones Indices, 2022). For similar reasons, the first observation for the DJIA corresponds to the 1st of October, 1928 as this was the date on which the index was expanded to include 30 companies as it does today (S&P Dow Jones Indices, 2013). Table 3.1 lists the selected indices, their corresponding volatility index, the date from which data is available for the stock market indices and the volatility indices, and the number of observations for each index.

For each stock market index listed in Column (1) of Table 3.1, daily data on the adjusted close price of the stock market index is obtained. The adjusted close price corrects the closing price of the index for any corporate decisions such as stock splits, dividends, and so on, that could affect its value and potentially distort results. The daily return is then computed using the following formula:

$$r_t = \ln\left(\frac{P_t}{P_{t-1}}\right) * 100 \tag{13}$$

where r_t and P_t denote the daily return and closing price on day *t* respectively. The daily logarithmic returns are then subsequently used to proxy *ex-post* volatility for each index.

Table 3.1: Availability of data for each selected index and its corresponding volatility index

Index (1)	Start Date (2)	No. of Observations (3)	Volatility Index (4)	Start Date (5)	No. of Observations (6)
DAX	12/08/1959	15,720	**VDAX**	10/03/1992	6,174
DJIA	01/10/1928	23,629	**CBOE VXD**	09/12/1997	6,163
FTSE 100	06/06/2002	5,054	**VFTSE**	10/03/2000	3,344
HSI	30/09/1964	12,999	**VHSI**	09/03/2001	5,233
KOSPI	14/03/1980	11,288	**KSVKOSPI**	10/04/2009	3,248
NASDAQ 100	06/06/2002	5,032	**CBOE VXN**	28/06/2002	5,017
NIFTY 50	05/09/1990	7,701	**India VIX**	09/11/2007	3,601
Nikkei 225	11/03/1970	12,860	**JNIV**	03/06/2002	4,898
Russell 2000	02/01/1979	11,053	**CBOE RVX**	02/01/2004	4,633
SMI	06/06/2002	5,026	**VSMI**	28/06/1999	5,821
S&P 100	11/03/1976	11,653	**CBOE VXO**	07/03/1986	8,940
S&P 500	04/03/1957	16,410	**CBOE VIX**	02/01/1990	8,165
S&P/TSX 60	06/06/2002	5,017	**VIXI**	27/04/2021	275

Notes: Column (1) lists the indices chosen based on the availability of a corresponding volatility index, which can be found in Column (4). The date from which the daily data is available for the given index is listed in Column (2) while Column (5) indicates the date from which daily data for the corresponding volatility index is available. Columns (3) and (6) indicate the number of observations for the given index and volatility index respectively. All data is gathered from Bloomberg or Refinitiv Eikon. The daily data is gathered until its latest available date as of 31-05-2022. Note that the dates in Column (2) and Column (5) are presented in the format of DD-MM-YYYY.

For the volatility indices in Column (4) of Table 3.1, daily data on the closing value is gathered. The closing value represents the market's estimate of the volatility for the following 30 days in annualized terms at that point in time. In order to construct the daily VaR, the closing value has to be standardized to represent the daily volatility. Following Giot (2005), the following transformation is applied to the closing price of the volatility index to obtain the daily implied volatility:

$$\sigma_{imp,t} = \frac{P_t}{\sqrt{365}} \tag{14}$$

Where $\sigma_{imp,t}$ refers to the implied volatility on day *t* and P_t denotes the closing value of the given volatility index on day *t*.

Table 3.2: Descriptive statistics of the daily logarithmic returns and the daily implied volatility

	Mean of daily returns (1)	**Standard deviation (2)**	**Mean of implied volatility (3)**
DAX	0.030	1.449	1.132
DJIA	0.023	1.188	1.019
FTSE 100	0.011	1.151	0.978
HSI	0.008	1.437	1.188
KOSPI	0.022	1.051	0.940
NASDAQ 100	0.050	1.451	1.212
NIFTY 50	0.029	1.387	1.130
Nikkei 225	0.017	1.456	1.249
Russell 2000	0.0256	1.563	1.116
SMI	0.012	1.117	0.957
S&P 100	0.034	1.185	1.056
S&P 500	0.030	1.146	1.024
S&P/TSX 60	0.035	0.741	0.888

Notes: Column (1) reports the mean of the daily logarithmic returns in percentages for the given stock market index, while Column (2) reports the percentage standard deviation. Column (3) reports the mean of the daily implied volatility computed from the closing values of the corresponding volatility indices. These statistics are computed using a common sample period for which data for the stock market index as well as its corresponding volatility index is available; this common sample period can be inferred from the dates presented in Column (2) and Column (5) in Table 3.1.

A preliminary analysis of the data is presented in Table 3.2. The mean daily return for all indices is close to 0 with the standard deviation being substantially larger in comparison. This suggests that assuming the expected returns to equal zero in order to proxy the *ex-post* volatility is a fair assumption to make in the case of daily returns. The mean of the standardized daily implied volatility shows slight variation around the intervals of 0.888 to1.249. A comparison of Column (2) and Column (3) in Table 3.2 indicate that the standard deviation of the daily logarithmic returns displays some degree of similarity to the mean of the standardized daily implied volatility. However, for the DAX and Russell 2000 indices, a greater degree of divergence between the standard deviation of the daily logarithmic returns and the mean of the daily implied volatility can be observed. This is also true of the HSI, NASDAQ 100, NIFTY 50, and Nikkei 225 indices although only to a lesser extent.

4. Methodology

4.1. Optimization

The optimal specification for the GARCH and GJR-GARCH models are determined in-sample using the first 1000 observations of data for each index as it comprises the first estimation window used for the rolling-window estimation. The optimal lambda for the EWMA model over the first 1000 observations is also reported. The historical average and autoregressive model do not need to be "optimized" and, as such, are not included in this section.

4.1.1. EWMA

The optimal value of the decay factor in the EWMA model is determined by estimating an iGARCH (1,1) specification and fitting it to the data of the given stock market index (see Section 2.2.3). The β coefficient in this model is then considered to be the optimal value for the smoothing factor. In order to compare this value of lambda with that suggested by *RiskMetrics* (0.94), a one-sample t-test is conducted to determine if the optimal lambda is significantly different from 0.94. An examination of the deviation between the optimal lambda and the value of 0.94 for several different indices will provide insight into the general applicability of the *RiskMetrics* methodology for financial institutions.

4.1.2. GARCH

To determine the optimal specification for the GARCH model, multiple GARCH specifications will be estimated with a varying number of lags of the previous squared innovations and the previous realized variances. This iterative process will begin with the estimation of a GARCH (1,1) model, following which subsequent GARCH specifications will be estimated with an increasing number of lags for the previous squared innovations and the previous realized variances. After the estimation of each specification, a Li-Mak test will be applied to ascertain if the estimated specification appropriately captures the variance structure of the data. The most parsimonious specification that returns an insignificant result for the Li-Mak test will be considered the "optimized specification".

An additional point of deliberation in the estimation of a GARCH model is the assumption of the distribution of the standardized residuals in the model. Given the general tendency for financial data to display a skewed distribution, it is possible that assuming the standardized residuals to follow a normal distribution limits the fit of the model to the data. In such instances, it may be more appropriate to posit that the standardized residuals follow a Student's t-distribution. Hence, the aforementioned optimization process will be repeated twice: once for the estimation of GARCH specifications under the assumption that the standardized residuals follow a Gaussian distribution, and once under the assumption that they follow a Student's t-distribution. Both optimized specifications determined through this process will then be estimated and their performance will be compared.

4.1.3. GJR-GARCH

The process for estimating the GJR-GARCH specifications will exactly follow the procedure described for the GARCH models. A Li-Mak test is applied to determine the optimal specification when the standardized residuals are assumed to follow a Gaussian distribution and when they are assumed to follow a Student's t-distribution. For both distributions considered, the most parsimonious specifications that results in an insignificant result for the Li-Mak test are considered the optimised specifications and used for further analysis.

4.2. Forecasting and the construction of the VaR

Each of the aforementioned models (as well as the historical mean and autoregressive models) will be estimated using a subset of data, i.e., in-sample. The performance of the models is then adjudged out-of-sample in order to arrive at a true determination of their predictive ability. To this end, the estimation window used will consist of 1000 observations following the advice of Brownlees et al. (2011), who find that a four-year estimation window performs best for rolling-window forecasting when applying their methodology. The use of 1000 observations roughly approximates the number of trading days in four years and, hence, is used as the estimation window. A rolling window approach is used to estimate the parameters of the given model, where the start and end date of the estimation period is rolled forward by one day between each subsequent re-estimation of the model's parameters. The estimated parameters are then used to obtain one-step ahead forecasts of the conditional volatility from each of these models.

The conditional volatility obtained from the rolling window forecasting is then used to construct the daily VaR using the formula displayed in Equation (1). Similarly, the daily implied volatility obtained through the standardization of the closing value of the volatility indices are also used to calculate the VaR. It should be noted that the VaR is calculated using the 5% significance level with the corresponding critical value for the standard normal distribution equalling 1.645.

4.3. Evaluation of the VaR

The computed daily VaR is then compared to the daily logarithmic returns for the given index. Following this, the number of observations for which the daily returns are lower than the estimated daily VaR is noted. In these instances, the realized downside loss is seen to exceed the maximum expected loss and, hence, is considered to be a violation of the VaR principle. The divergence between the expected number of violations of the daily VaR and the realized number of violations for each model will be formally evaluated by means of a backtest, namely the backtest proposed by Kupiec (1995). For this test, the failure rate of the VaR, i.e., the number of realized violations as a proportion of the overall number of observations, is compared with the failure rate that would be expected if the VaR is correctly specified, i.e., the significance level used to compute the VaR. Under the null hypothesis of the failure rate equalling the significance level, the computed test statistic (see section A3 in the Appendix for formula) follows a chi-squared distribution asymptotically. A significant result for this test indicates that the realized failure rate significantly differs the expected failure rate, i.e., the significance level of 5% in this case. Kupiec's backtest is particularly useful as it is a two-sided test, i.e., it returns a significant result if the proportion of the realized number of violations exceeds the proportion of the expected number of violations, or if they the realized number of violations are significantly lower than expected. This process is then repeated for each index and approach used to estimate the VaR.

While Kupiec's backtest is useful in determining the degree of equality between the realized failure rate and the significance level, it does not provide insight into whether the VaR violations are correlated over time. This is also of interest to financial institutions and, hence, should also be considered when determining the optimal approach to modelling the VaR. To this end, the conditional coverage test proposed by Christoffersen (1988) is also used to evaluate the performance of the various approaches used to model the VaR. Christoffersen's conditional coverage test (hereafter referred to as "CCC test") assesses the independence in the violations of the VaR over time, i.e., it examines if the probability of a VaR violation on a given day is dependent on the outcome that is realized on the previous day. Under the null hypothesis, the test statistic, i.e., a likelihood ratio (see section A4 in the Appendix), asymptotically follows a chi-squared distribution. For this test, a significant result would indicate that the violations of the VaR are not independent of each other and that consecutive violations are seen with excessive frequency. This test is repeated for each index and approach used to calculate the VaR.

5. Results

5.1. Optimised Specifications

Column (1) in Table 5.1.1 reports the optimal value of the smoothing exponent for the EWMA model as derived from the estimation of an iGARCH (1,1) model. The optimal value can be seen to display variation among indices, with the value ranging from 0.860 to 0.993. The smoothing factor value of 0.94 advocated by *RiskMetrics* falls within this range, with the optimal lambdas showing slight variation on either side of this value. Furthermore, none of the indices returned a significant result for the one-sample t-test, indicating that the optimal lambda value is not significantly different from 0.94 for any of the indices. Hence, it can be concluded that the lambda value advocated by *RiskMetrics* offers a generally applicable approach for estimating the EWMA model for financial institutions. The optimal value of lambda for each index can also be seen to lie close to the maximum possible value of one. This high value suggests that there is predictability in volatility among financial data as the current value of the forecasted volatility is seen to be influenced by the prior realized volatility.

For all indices, the GARCH (1,1) and GJR-GARCH (1,1) were the most parsimonious specifications that returned an insignificant result for the Li-Mak test. This was found to be the case regardless of whether the standardized residuals were assumed to follow a Gaussian or Student's t-distribution. The test statistic and the corresponding p-value for the Li-Mak test conducted on the GARCH (1,1) and GJR-GARCH (1,1) models is reported in Columns (2) to (5) in Table 5.1.1. These results suggest that the GARCH (1,1) and GJR-GARCH (1,1) specifications contain a sufficient number of lags to capture the variance structure of the data. These results strongly support the common academic notion that including a single lag of the innovations from the model as well as only the immediately preceding realized variance is appropriate for specifying a GARCH model. The results of the optimization process suggests that this belief can be extended to the GJR-GARCH model as well.

Table 5.1.1: Results of the optimisation process for the EWMA, GARCH, and GJR-GARCH models

	EWMA (1)	**GARCH (Gaussian distribution) (2)**	**GARCH (Student's t-distribution) (3)**	**GJR-GARCH (Gaussian distribution) (4)**	**GJR-GARCH (Student's t-distribution) (5)**
S&P 500	0.942	11.890	11.345	12.323	12.010
		(0.220)	(0.253)	(0.196)	(0.213)
Russell 2000	0.915	11.747	8.092	8.681	7.230
		(0.228)	(0.525)	(0.467)	(0.613)
NASDAQ 100	0.963	6.127	6.100	4.973	5.004
		(0.727)	(0.730)	(0.837)	(0.834)
DJIA	0.890	6.314	6.117	6.713	6.700
		(0.708)	(0.728)	(0.667)	(0.668)
S&P100	0.969	6.316	5.940	5.852	5.775
		(0.708)	(0.746)	(0.755)	(0.762)
S&P/TSX 60	0.939	2.954	3.021	3.304	3.363
		(0.966)	(0.963)	(0.951)	(0.948)
FTSE 100	0.931	9.470	9.964	7.137	7.073
		(0.395)	(0.353)	(0.623)	(0.630)
DAX	0.861	4.456	4.853	3.272	3.046
		(0.879)	(0.847)	(0.953)	(0.963)
SMI	0.941	10.231	10.718	7.200	7.211
		(0.332)	(0.296)	(0.616)	(0.615)
HSI	0.862	6.832	4.520	6.839	4.827
		(0.655)	(0.874)	(0.654)	(0.849)
KOSPI	0.993	1.575	1.012	1.636	0.947
		(0.997)	(0.9994)	(0.996)	(0.999)
NIFTY 50	0.916	14.050	14.354	13.330	14.042
		(0.121)	(0.110)	(0.148)	(0.121)
Nikkei 225	0.860	1.565	1.572	1.696	1.579
		(0.997)	(0.997)	(0.995)	(0.997)

Notes: Column (1) reports the optimal value of the decay factor in the EWMA model, which equals the value of the β coefficient obtained from the estimation of an iGARCH (1,1) model. The results of the one-sample t-test is also displayed, where the significance of the test statistic and its corresponding p-value is represented using stars (*p<0.10, **p<0.05, ***p<0.01). Columns (2), (3), (4), and (5) report the test statistic and the corresponding p-value for the Li-Mak test conducted on the GARCH (1,1) and GJR-GARCH specifications under the null hypothesis that there is no autocorrelation in the squared and standardized residuals of the model. The p-values are reported in parentheses. All statistics are computed using the first 1000 observations of data for each index. All figures are reported to 3 decimal places.

For comparison, the results of the optimization process over the entire data sample for each index is reported in Table A5.1 (see Section A5 in the Appendix).

5.2. Daily VaR modelling using ex-post volatility

5.2.1. Results

Table 5.2.1 reports the expected and realized violations of the VaR for all indices when using different approaches to estimate the conditional volatility. The results of Kupiec's backtest as well as the CCC test are also displayed. The best performing model is determined on the basis of these tests, where the model that results in a significant result for the fewest number of indices is considered to be the optimal approach for modelling the VaR. Considering Kupiec's backtest first, the best performing model is found to be the GJR-GARCH model (assuming that the standardized residuals follow a Gaussian distribution). Of the 13 stock market indices, this model displayed a significantly different failure rate from the expected rate of 5% for three indices i.e., NASDAQ 100, S&P/TSX 60, and the FTSE 100. For all three of these indices, the realized violations substantially exceeded the expected number of violations, resulting in a test statistic that is significant at the 1% level. Interestingly, assuming the standardized residuals to follow a Student's t-distribution worsens the performance of the GJR-GARCH model. In this case, the failure rate is found to significantly differ from 5% for two additional indices, namely the SMI and HSI indices. However, the test statistic is only significant at the 10% level for these two indices, suggesting that the change in performance is marginal.

The model closest in its performance to the GJR-GARCH model is the traditional GARCH model. When assuming the standardized residuals to follow a Gaussian distribution, a significant result for Kupiec's backtest is found for four indices, while this increases to five indices when assuming a Student's t-distribution instead. However, as was seen for the GJR-GARCH model, the additional significant result found when assuming a Student's t-distribution is only significant at the 10% level, which again suggests that the difference in performance when assuming a Gaussian or Student's t-distribution is minimal when modelling the daily VaR. Comparing the performance of the GARCH model to the GJR-GARCH model, it can be seen that the inclusion of the asymmetric term in the GJR-GARCH specification improves its ability to model the daily VaR, thus resulting in significantly different failure rate from 5% for fewer indices.

Table 5.2.1: VaR violations for each statistical model and index

	Expected violations (1)	Historical Mean (2)	AR (3)	EWMA (4)	GARCH (5)	GARCH (t-distribution) (6)	GJR-GARCH (7)	GJR-GARCH (t-distribution) (8)
DAX	736	708###	693###	870***###	752	757#	743	731
DJIA	1131	1019***###	1043***###	1285***###	1122###	1100###	1106##	1096###
FTSE 100	203	211###	211	266***###	232**##	231**##	242***##	243***##
HSI	600	523***###	563###	676***###	567###	559*###	565###	558*###
KOSPI	514	511###	525###	594***###	520###	525###	503#	511
NASDAQ 100	202	236**###	239***##	260***###	243***###	232**#	247***###	235**###
NIFTY 50	335	282***###	297**#	367*##	314	320	314	320
Nikkei 225	593	565###	609###	723***###	623##	619###	625	631
Russell 2000	503	522###	501#	625***###	537	521	528	534
SMI	201	200###	234**###	254***###	235**###	239***###	224	227*
S&P 100	533	520###	521	600***###	535	525	550	537
S&P 500	771	772###	770###	899***###	784###	766###	782	769
S&P/TSX 60	201	217###	218##	281***###	258***###	259***###	254***###	252***###

Notes: this table depicts the performance of each given statistical model in modelling the VaR using the historical data of each specific index. Column (1) displays the expected number of violations based on the number of observations for which the volatility was forecasted multiplied by the significance level at which the VaR will be evaluated, i.e., 5%. Column (2) indicates the number of violations of the VaR computed using the volatility forecasts from the historical mean model while Column (3) and (4) report the same for the autoregressive model and EWMA model respectively. Column (5) and (7) display the performance of the GARCH and GJR-GARCH models when the standardized residuals are assumed to follow a Gaussian distribution, while Columns (6) and (8) display the results when the standardized residuals are assumed to follow a Student's t-distribution for each respective model. The results of Kupiec's two-sided backtest is also displayed in the above table, where the significance of the test statistic and its corresponding p-value is represented using stars (*p<0.10, **p<0.05, ***p<0.01). The results for the CCC test are also shown, where the significance of the test statistic and its corresponding p-value is represented using hashtags (#p<0.10, ##p<0.05, ###p<0.01).

In comparison to the GARCH and GJR-GARCH models, the historical mean and autoregressive models offer a similar level of performance in modelling the daily VaR despite their apparent simplicity. Both models display a significantly different failure rate from 5% for four indices. For three of the four indices in which a significant result is found, the historical mean model overestimates the VaR resulting in the expected number of violations exceeding the realized violations. In fact, this general pattern of the realized violations being less than the expected number of violations can be observed for the historical mean model as well as the autoregressive model albeit to a lesser extent for the latter. This suggests that these models tend to overestimate the conditional volatility in some situations and, hence, also the VaR. This is likely because the volatility decreases over the sample period in these instances and the adjustment to this decreasing level of volatility is delayed for these models. This inference is supported by Figure 5.2.1, which shows the VaR for the DJIA index, which is computed using the conditional volatility from the historical mean model. The moderate upward trend in the VaR indicates that the volatility has generally shown a decline over the sample period. Hence, due to the delayed adjustment to the decreasing volatility over time, the VaR is overestimated in this case. In such instances, a potential consequence for financial institutions of using the such models is that profit-maximizing decisions may be foregone as the risk (as measured by the VaR) is overestimated.

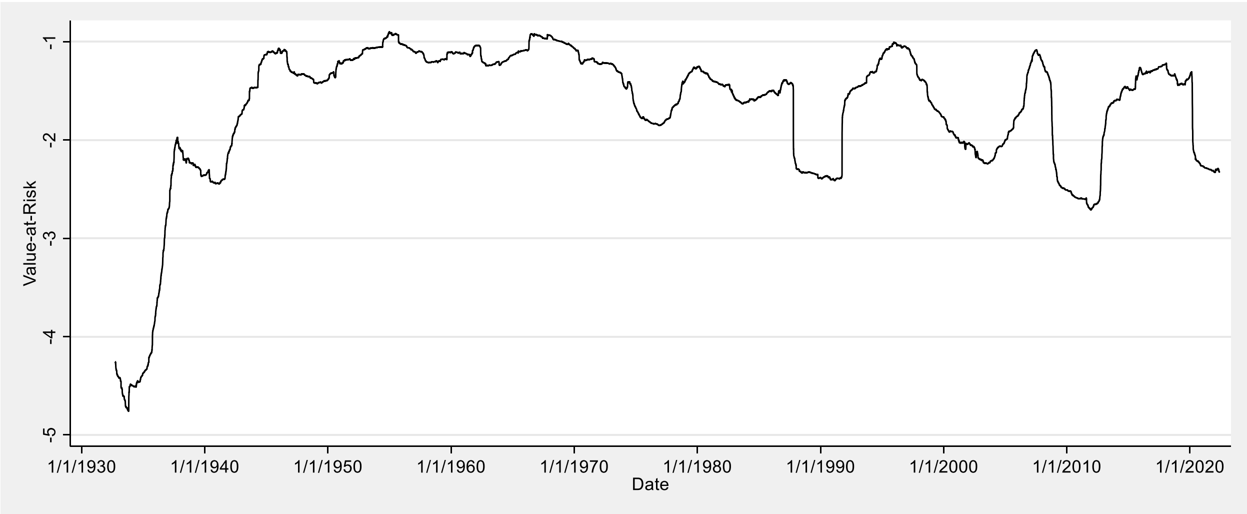

Figure 5.2.1: Evolution of the VaR for the DJIA index computed using the historical mean model

Notes: the above figure displays a time-series line plot that depicts the VaR over the entire sample period for the DJIA index. The VaR is computed using the conditional volatility estimated by the historical mean model. The Y-axis displays the VaR, while the X-axis displays the date. The DJIA is used as an example amongst the indices that displayed a significant result and where the realized violations exceeded the expected violations.

At the opposite end of the spectrum, the EWMA model is seen to display the worst performance amongst the different statistical models in Table 5.2.1. The EWMA model returns a significant result for Kupiec's backtest at the 1% level for all indices (except the NIFTY 50 which shows significance at the 5% level), which indicates that this model fails to appropriately model the daily VaR for all indices as the failure rate is found to significantly differ from the expected rate. A further inspection of the results for the EWMA model shows that the realized number of violations always substantially *exceeds* the expected number of violations. This suggests that the EWMA model consistently underestimates volatility and, hence, the VaR is also consistently underestimated.

The results found for Kupiec's backtest is generally comparable across the different statistical models. However, a greater degree of variation can be found amongst the result for the CCC test. Once again, the GJR-GARCH model is found to be the best performing statistical model in this regard. When it is assumed that the standardized residual follows a Gaussian distribution, a significant result for six indices can be found, which improves to five indices when assuming a Student's t-distribution. However, even in the case of this test, the difference in performance obtained when assuming a Gaussian or Student's t-distribution is marginal, as the additional result is found to only be significant at the 10% level. Overall, these results suggest that the GJR-GARCH is the least prone to excessive consecutive violations of the calculated VaR.

In a similar manner to the results obtained for Kupiec's backtest, the second-best performing model is found to be the GARCH model. In this case, significant evidence is found indicating that the violations of the VaR are correlated for nine indices when a Gaussian distribution is assumed and for ten indices when a Student's t-distribution is assumed instead. A noticeable deterioration in performance can be observed when employing the GARCH model in comparison to a GJR-GARCH specification. This could potentially be due to the ability of the GJR-GARCH model to account for leverage effects, where the additional increase in conditional volatility following a negative innovation leads to the one-day VaR being appropriately adjusted such that excessive consecutive violations are avoided.

Unlike for Kupiec's backtest, the historical mean model is found to perform much worse than the GARCH and GJR-GARCH models on the basis of the CCC test. For the historical mean model, a significant result is obtained for all indices, indicating that the use of this model will likely lead to consecutive violations being seen with excessive frequency. While the autoregressive model does outperform the historical mean model for this test, it still performs rather poorly. A significant result is found for all but two indices for the autoregressive model. While the simplicity of the historical mean and autoregressive models did not appear to hinder their performance with regards to Kupiec's backtest, it is likely that it prevents these models from being able to adequately capture the dynamic structure of the conditional volatility in the data. In turn, this results in the violations of the VaR being correlated over time. Finally, the EWMA model is once again found to be the worst performing model as a significant result for the CCC test is found for every index. Its performance is on-par with the historical mean model and worse than the autoregressive model despite the implementation of an exponential weighting scheme intended to lead to a superior modelling of the conditional volatility (and, hence, the one-day VaR).

Based on the results of Kupiec's backtest and the CCC test, it is clear that the GJR-GARCH model performs best with respect to both criteria. Hence, the first hypothesis which states that the GJR-GARCH model will be most accurate at modelling the daily VaR in comparison to other statistical models is not rejected.

5.2.2. Comparing the models

Despite sharing commonalities in their specifications, a strong divergence in performance can be observed between the autoregressive, EWMA, and GARCH models. When comparing the autoregressive model with the GARCH model, the results for Kupiec's backtest are found to be similar for both models. However, when considering the CCC test, the GARCH model is found to be marginally superior. The primary difference between the autoregressive and GARCH model is the exclusion of the lagged conditional variance term in the former, which suggests that this term is likely required in order to adequately adjust the forecasted volatility in times of extreme returns. Without this term, the adjustment, i.e., the lagged response to an increase in volatility, is likely to take longer, thus leading to excessive consecutive violations and correlation amongst the violations.

However, the poor performance of the EWMA model cannot be attributed to the same reason as its specification contains a lagged squared innovation term as well as the lagged conditional variance (see Equation (8) in Section 2.2.3). Given this, the poor performance of this model is rather surprising given its apparent similarity to the GARCH (1,1) model. However, a few differences between the EWMA and GARCH specifications can be noted. Firstly, the coefficients of the EWMA model, i.e., λ and $(1-\lambda)$, are required to sum to one while this is not the case for the traditional GARCH model. Typically, however, the coefficients in the GARCH model, i.e., α and β, tend to sum to a value that is extremely close to one. This suggests that the restriction imposed on the coefficients of the EWMA model are unlikely to be the source of its relatively poor performance. The other discernible difference between the EWMA and GARCH model is the exclusion of the constant term, i.e., ω, in the former. The constant term in a GARCH model represents a proportion of the long-term unconditional variance and the minimum conditional variance that can result in any period. Hence, omitting the constant term in the specification is likely to lead to an underestimation of volatility as this minimum conditional variance is not accounted for.

In order to verify if the exclusion of the constant term is responsible for the poor performance of the EWMA model, the VaR is computed using the conditional volatility obtained from the estimation of an autoregressive model without a constant. The results of this are presented in Table A6.1 (see Section A6 in the Appendix). In comparison to the results obtained from the standard autoregressive model, the performance can be seen to deteriorate tremendously when the constant term is eliminated from the specification. The realized violations increase considerably and much like the EWMA model, the realized violations are seen to exceed the expected number of violations for all indices. This suggests that the exclusion of the constant term is indeed responsible for the poor performance of the EWMA model as it results in an underestimation of the conditional volatility and, consequently, also the one-day VaR.

5.3. Comparison of daily VaR modelling with ex-ante and ex-post volatility

The results shown in Table 5.3.1 differ from those displayed in Table 5.2.1. In order to facilitate comparison between the *ex-ante* and *ex-post* approaches to modelling the VaR, the results obtained using these approaches are compared over a common time period, i.e., a period over which data for both the stock market index and the corresponding volatility index are available. These results are displayed in Table 5.3.1.

Considering again the results of Kupiec's backtest first, utilizing implied volatility to model the VaR returns a significant result for five indices, which indicates that the failure rate is significantly different from 5%. Amongst the indices for which a significant result was found, the VIX and the VXO display a severe overestimation of the VaR as the expected violations far exceed the realized violations. The same is true for the VXD but the degree of underestimation is not as severe as for the VIX and VXO. On the other hand, for the VDAX and VHSI, the implied volatility estimates tend to underestimate the VaR as the realized number of violations exceed the expected number of violations. In this case, the VDAX in particular fails to accurately model the VaR since it displays a large difference between the expected and realized number of violations. Hence, for those indices that returned a significant result, the volatility indices for the US in particular tend to result in the VaR being overestimated, which in turn implies that the implied volatility estimates derived from these indices tend to be exaggerated. However, this cannot be said for all the US-based volatility indices, as the VaR computed using the implied volatility estimates from the RVX and the VXN do not return a significant result for Kupiec's backtest. For the NASDAQ

100, using the implied volatility from the VXN is the only approach that results in an accurate modelling of the one-day VaR as a significantly different failure rate from 5% is found for all the statistical models.

The behavior of the VIX in modelling the one-day VaR is consistent with the findings of Zakamulin (2016). The VIX displays periodic peaks and troughs in its value that coincides with the respective troughs and peaks in the value of the underlying index, i.e., the S&P 500. However, the average duration of a phase in which the value of the VIX is falling is found to be 1.4 times longer than the duration of a phase in which its value is rising. An implication of this is that when the value of the VIX is at a peak, it tends to stay at this higher level of volatility for longer than would be expected. The slow decline from its peak value to a more typical level of implied volatility is attributed to investor overreaction. As such, it can be posited that the expected number of violations exceeds the realized violations for the VIX due to investor overreaction resulting in an overestimation of the implied volatility. This is likely to also be the case for the VXO and VXD.

Figure 5.3.1 and 5.3.2 compare the calculated VaR when using implied volatility estimates and when using volatility estimates from the GJR-GARCH model (Gaussian distribution). Both figures show extended periods of time during which the VaR computed using the conditional volatility from the GJR-GARCH model lies above that calculated using the implied volatility from the VIX. This supports the prior conclusion that the implied volatility indicated by the VIX tends to be exaggerated. However, when considering all indices overall, utilizing implied volatility to model the daily VaR can be seen to lead to a comparable performance to that obtained when using the historical mean, autoregressive, GARCH, and GJR-GARCH models on the basis of Kupiec's backest. For the GARCH and GJR-GARCH models, a significant result for four indices is obtained when assuming the standardized residuals follow a Gaussian distribution, and for five indices when assuming that they follow a Student's t-distribution. In comparison, the historical mean and autoregressive models display a significantly different failure rate from 5% for four indices.

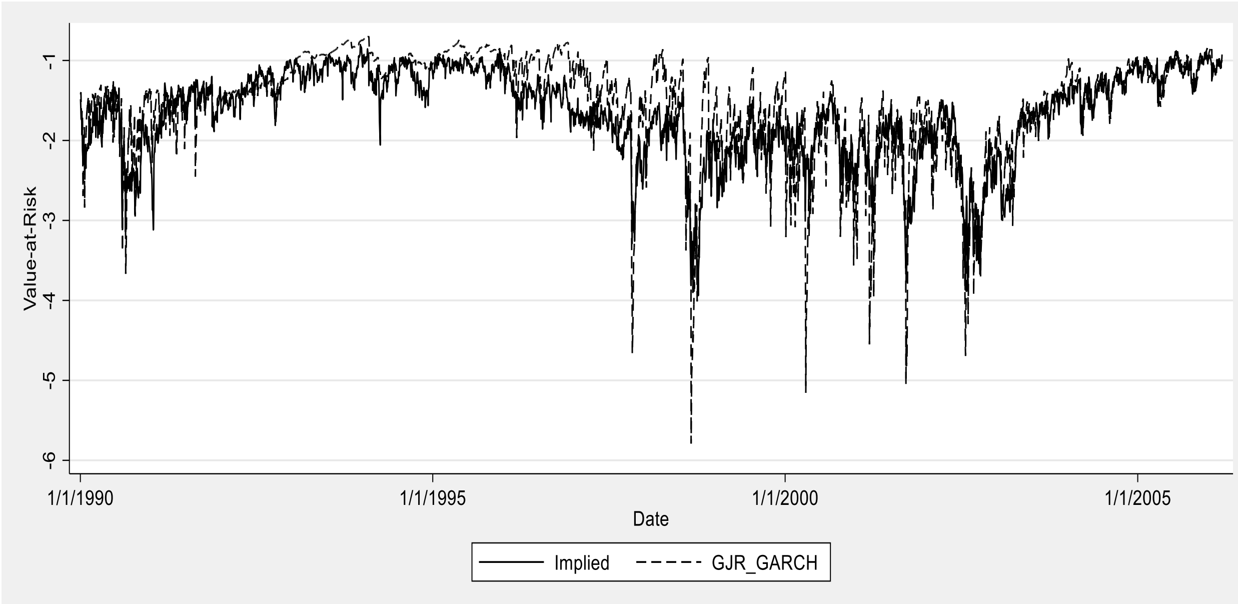

Figure 5.3.1: Comparison of VaR for the S&P 500 modelled with implied volatility and the GJR-GARCH model

Notes: the above figure displays a time-series line plot that depicts a comparison of the VaR computed using the daily implied volatility obtained by standardizing the values of the VIX and the conditional volatility from the GJR-GARCH model. The Y-axis displays the VaR, while the X-axis displays the date. The VaR is computed over the common sample period for which data for the S&P 500 and the VIX are both available. However, Figure 5.3.1 only displays the VaR for the first half of this common sample period, i.e., 02/01/1990 to 10/03/2006, for a clear depiction.

However, once again, the EWMA model is found to perform the worst in its ability to model the one-day VaR. The EWMA model displays a significantly different failure rate for all but one index, namely the S&P/TSX 60. However, little credence can be placed on this given the limited number of overlapping observations for this index and its corresponding volatility index. Hence, using the implied volatility approach offers a superior alternative to using the EWMA model for the computation of the one-day VaR.

Considering the results of the CCC test, the implied volatility approach displays a significant result for five indices. Hence, for these indices, using the implied volatility to model the one-day VaR leads to its violations being correlated over time. Additionally, four of these five indices also returned a significant result for Kupiec's backtest, indicating that these volatility indices in particular are unsuitable for modelling the one-day VaR across both evaluation criteria. In comparison, the GJR-GARCH model can be seen to outperform the implied volatility approach along the basis of the CCC test as it displays a significant result for four and three indices when assuming a Gaussian and Student's t-distribution respectively. This indicates that the GJR-GARCH model offers a superior alternative to using the implied volatility approach in modelling the daily VaR.

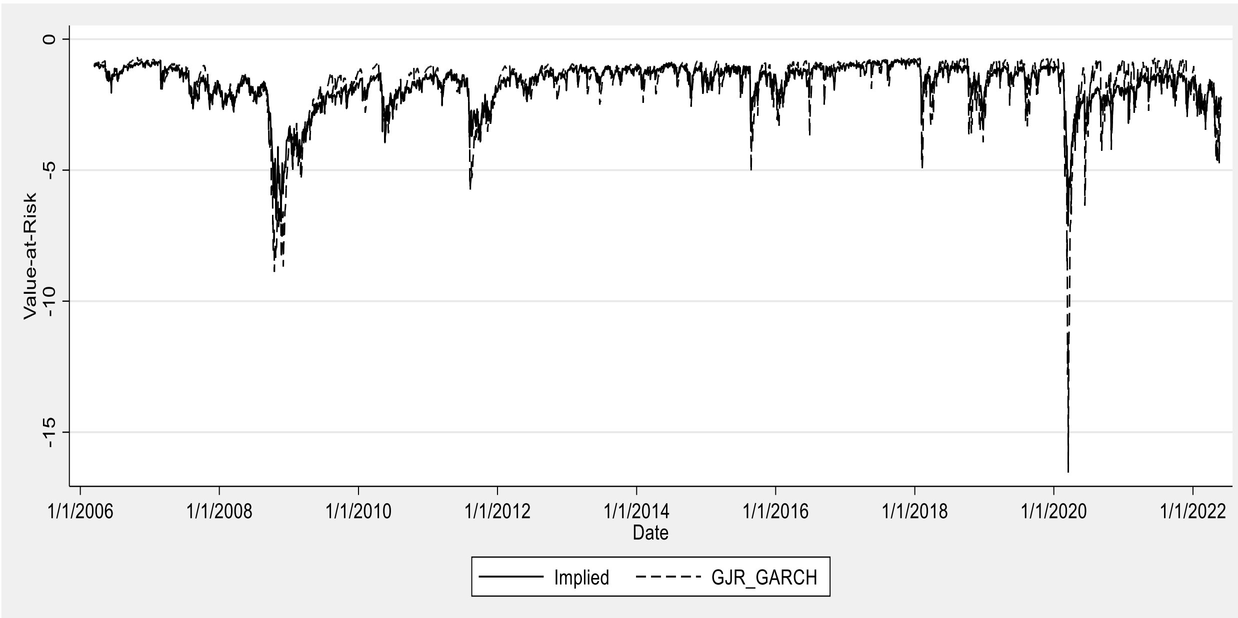

Figure 5.3.2: Comparison of VaR modelled with implied volatility and the GJR-GARCH model

Notes: the above figure displays a time-series line plot that depicts a comparison of the VaR computed using the daily implied volatility obtained by standardizing the values of the VIX and the conditional volatility from the GJR-GARCH model. The Y-axis displays the VaR, while the X-axis displays the date. The VaR is computed over the common sample period for which data for the S&P 500 and the VIX are both available. However, Figure 5.3.2 only displays the VaR for the second half of this common sample period, i.e., 11/03/2006 to 31/05/2022, for a clear depiction.

The results for the GARCH model are similar to those obtained from using the implied volatility approach. The GARCH model (both Gaussian and Student's t-distribution) is found to show significant correlation amongst the VaR violations for five indices. Although the implied volatility approach returns a significant result for the same number of indices, it should be noted that all of the violations are significant at the 1% level for the former, while this is only true of three indices for the latter (the remaining two indices are only significant at the 10% level). This indicates that the GARCH model marginally outperforms the implied volatility approach along this dimension.

Table 5.3.1: VaR violations for implied volatility and ex-post volatility estimates

	Expected Violations (1)	Implied Volatility (2)	Historical Avg. (3)	AR (4)	EWMA (5)	GARCH (6)	GARCH (t-dist.) (7)	GJR-GARCH (8)	GJR-GARCH (t-dist.) (9)
DAX	309	410***###	333###	325	378***###	345**#	346**#	345**#	340*
DJIA	308	237***###	326###	317	379***###	332	315	321	317
FTSE 100	167	160#	165###	166	227***###	188#	190*#	194**#	196**#
HSI	262	292*	255###	272##	323***###	272	266	276	270
KOSPI	162	152	131***###	154###	216***###	187*###	185*##	187*##	181##
NASDAQ 100	202	205	236**###	239***##	260***###	243***###	232**#	247***###	235**###
NIFTY 50	180	167	154**###	157*	204*	172	174	174	175
Nikkei 225	245	243	236###	267	317***###	267	262	270	264
Russell 2000	232	215	245###	233	301***###	255#	246	253#	261*#
SMI	203	183###	200###	235**###	256***###	237**###	241***###	226	229*
S&P 100	447	269***###	439###	430	511***###	458	448	470	461
S&P 500	408	257***###	408###	410##	474***###	410	396	428	419
S&P/TSX 60	14	9	6**##	7**	15	15	14	17	16

Notes: this table displays the performance of implied volatility estimates and backward-looking estimates of volatility in VaR modelling. It should be noted that this analysis is conducted over a time period that is common for both the stock market index and its corresponding volatility index. Column (1) displays the expected number of violations based on the number of observations for which the volatility was forecasted multiplied by the significance level at which the VaR will be evaluated, i.e., 5%. Column (2) indicates the number of violations of the VaR computed using the implied volatility derived from the values of the volatility indices. Columns (3) – (9) report the realized number of violations of the various statistical models shown in Table 5.2.1 over the same time period. The results of Kupiec's two-sided backtest is also displayed in the above table, where the significance of the test statistic and its corresponding p-value is represented using stars (*p<0.10, **p<0.05, ***p<0.01). The results for the CCC test are also shown, where the significance of the test statistic and its corresponding p-value is represented using hashtags (#p<0.10, ##p<0.05, ###p<0.01).

Similar to the GARCH model and the implied volatility approach, the autoregressive model also returns a significant result for five indices and, hence, is seen to offer a comparable level of performance to these approaches to modelling the VaR (on the basis of the CCC test). However, this is not true of the historical mean and EWMA models, both of which perform significantly worse than when using the implied volatility to model the one-day VaR. The historical mean and EWMA model return a significant result for all but two indices, indicating that there is a significant deterioration in performance when using these two models as it results in excessive consecutive VaR violations.

Therefore, the second hypothesis stating that the use of *ex-post* measures of volatility will result in a more accurate modelling of the VaR in comparison to using implied volatility is rejected. Although some measures of *ex-post* volatility do result in an improved modelling of the daily VaR in comparison to using implied volatility, this is not the case for all *ex-post* volatility measures. The EWMA model performs worse than the implied volatility approach on the basis of both Kupiec's backtest and the CCC test, while the historical mean model performs worse on the basis of the CCC test. Hence, a broad generalization that *ex-post* measures of volatility outperform implied volatility in modelling the one-day VaR cannot be made.

6. Conclusion

The primary objective of this paper has been to answer the question: *which approach to modelling volatility results in the most accurate one-day VaR?* To answer this question, two fundamental approaches to estimating the conditional volatility required to model the VaR were considered: using *ex-post* measures of volatility that involve the estimation of a multitude of statistical models on historical data, and an *ex-ante* approach that involves obtaining the implied volatility from volatility indices. The performance of these approaches was evaluated by means of Kupiec's backtest and Christoffersen's conditional coverage test.

Of the *ex-post* measures of volatility, the GJR-GARCH model was shown to outperform all alternate statistical models across both the aforementioned tests. The GARCH model was found to be the next best model as it returned a significant result for the fewest indices (after the GJR-GARCH model) for both Kupiec's backtest and the CCC test. Despite displaying a significantly different failure rate for a similar number of indices as the GARCH and GJR-GARCH models, the autoregressive and historical mean model are found to be suboptimal alternatives as they are more likely to result in correlation amongst the VaR violations over time. This is especially true for the historical mean model, which displays a substantial deterioration in performance along the basis of the CCC test in comparison to the other models. Of all the statistical models considered, the EWMA model is found to offer the least accurate approach to modelling the VaR due to its extremely poor performance for both Kupiec's backtest and the CCC test.

The above findings align with those of Brailsford and Faff's, where they find the GJR-GARCH to be the best performing model followed by the traditional GARCH model when using the MAE and MAPE error statistics. The conclusions of Hansen and Lunde were also found to be in accordance with the results found in this analysis as the GJR-GARCH model did show an improvement in performance over the traditional GARCH model, especially when considering the results of the CCC test. However, the findings of Walsh and Yu-Gen Tsou stand in strong contrast to the results of this paper as the EWMA model is found to be the worst performing model in its ability to model the VaR.

In the analysis of *ex-post* and *ex-ante* measures of volatility, it was found that using implied volatility leads to a more accurate modelling of the daily VaR in comparison to the historical mean and EWMA models. However, the best performing model remained the GJR-GARCH, while the GARCH and autoregressive models can also be seen to marginally outperform the implied volatility approach. Although this study employs a differing methodology, the results show some similarity with those found by Giot as implied volatility on its own does not offer the optimal approach to modelling the daily VaR. Hence, to answer the main research question of this paper, using the GJR-GARCH model provides the optimal approach to modelling the one-day VaR over all other approaches considered.

However, the aforementioned results should be interpreted with some amount of caution due to potential limitations associated with this paper. The proxy for *ex-post* volatility that is utilized in the analysis conducted in this paper represents one such limitation. Following the approach used by Day and Lewis, the square of the daily returns is used to proxy the *ex-post* volatility for each stock market index. However, as noted by Andersen and Bollersev (1998), this measure of volatility is rather noisy as the standardized residual tends to display significant variation between observations relative to the conditional variance. An alternative approach that may be more accurate is to proxy volatility by using data of a higher frequency, i.e., intraday data. Andersen and Bollersev suggest squaring the hourly returns and summing them over the course of the trading day offers a more accurate proxy of the daily volatility. Using such an approach in future research could identify whether the noise involved in the current proxy for daily volatility materially affects the results presented in this paper. If the results are found to align with those presented in this study, then greater confidence can be attached to the insights derived in this paper.

Despite these limitations, the results of this study can be beneficial for financial institutions in their application of the VaR. As shown in this paper, the optimal approach to modelling the one-day VaR is to use a GJR-GARCH model to forecast the conditional volatility. A more accurate computation of the VaR will enable financial institutions to make more prudent financial decisions that reduce the risk of bankruptcy, which in turn promotes greater national and global economic stability. Additionally, financial organizations also stand to gain as a more accurate assessment of their risk exposure could potentially lead to additional profit-maximizing decisions being undertaken. Beyond this, the results of this paper also offer additional insight into the much-deliberated topic of volatility forecasting in the academic literature. The conflicting evidence found in the current stock of papers makes it difficult to obtain a clear consensus on the best approach to modelling volatility within a VaR framework. This paper provides a definitive answer that can resolve this uncertainty, while also introducing a novel approach for treating implied volatility in this context.

7. Bibliography

Akgiray, V. (1989). Conditional Heteroskedasticity in Time Series of Stock Returns: Evidence and Forecasts. *Journal of Business, 62*(1), 55-80.

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International economic review*, 885-905.

Black, F. (1976). Studies of stock market volatility changes. *1976 Proceedings of the American statistical association business and economic statistics section*, 177-181.

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

Bollerslev, T. (2008). Glossary to arch (garch). *CREATES Research paper*, *49*.

Brailsford, T.J., & Faff, R.W. (1996). An evaluation of volatility forecasting techniques. *Journal of Banking and Finance 20*(3), 419-438.

Brooks, C., & Burke, S. P. (2003). Information criteria for GARCH model selection. *The European journal of finance, 9*(6), 557-580.

Brownlees, C. T., Engle, R. F., & Kelly, B. T. (2011). A practical guide to volatility forecasting through calm and storm. *Available at SSRN 1502915*.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International economic review*, 841-862.

Cox, J. C., Ross, S. A., & Rubinstein, M. (1979). Option pricing: A simplified approach. *Journal of financial Economics*, *7*(3), 229-263.

Day, T.E., & Lewis, C.M. (1992). Stock Market Volatility and the Information Content of Stock Index Options. *Journal of Econometrics, 52*(1-2), 267-287.

Dimson, E., & Marsh, P. (1990). Volatility Forecasting Without Data-Snooping. *Journal of Banking and Finance*, *14*(2-3), 399-421.

Duffie, D., & Pan, J. (1997). An overview of value at risk. *Journal of derivatives, 4*(3), 7-49.

Engle, R.F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica, 50*(4), 987-1007.

Engle, R. F. (2001). GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics. *Journal of Economic Perspectives, 15*(4), 157-168.

Engle, R. F., & Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric reviews*, *5*(1), 1-50.

Giot, P. (2005). Implied volatility indexes and daily value at risk models. *The Journal of derivatives*, *12*(4), 54-64.

Glosten, L. R., Jagannathan, R., & Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *The journal of finance*, *48*(5), 1779-1801.

Hansen, P.R., & Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH (1, 1)? *Journal of Applied Econometrics, 20*(7), 873-889.

Hendricks, D. (1996). Evaluation of value-at-risk models using historical data. *Economic policy review*, *2*(1).

Javed, F., & Mantalos, P. (2013). GARCH-type models and performance of information criteria. *Communications in Statistics-Simulation and Computation*, *42*(8), 1917-1933.

J.P. Morgan/Reuters. (1996). *RiskMetrics - Technical Document*. Retrieved from https://www.msci.com/documents/10199/5915b101-4206-4ba0-aee2-3449d5c7e95a

Kupiec, P. H. (1995). *Techniques for verifying the accuracy of risk measurement models* (Vol. 95, No. 24). Division of Research and Statistics, Division of Monetary Affairs, Federal Reserve Board.

Li, W. K., & Mak, T. K. (1994). On the squared residual autocorrelations in non-linear time series with conditional heteroskedasticity. *Journal of Time Series Analysis*, *15*(6), 627-636.

Lundbergh, S., & Teräsvirta, T. (2002). Evaluating GARCH models. *Journal of Econometrics*, *110*(2), 417-435.

Mayhew, S. (1995). Implied volatility. *Financial Analysts Journal*, *51*(4), 8-20.

Merton, R. C. (1973). Theory of rational option pricing. *The Bell Journal of economics and management science*, 141-183.

Pagan, A.R., & Schwert, G.W. (1990). Alternative Models for Conditional Stock Volatility. *Journal of Econometrics, 45*(1), 267-290.

Scholes, M., & Black, F. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, *81*(3), 637-654.

S&P Dow Jones Indices. (2013). *Dow Jones Industrial Average Historical Components.* Retrieved from http://www.djindexes.com/mdsidx/downloads/brochure_info/Dow_Jones_Industrial_Average_Historical_Components.pdf

S&P Dow Jones Indices. (2022). *S&P 500: The Gauge of the Market Economy*. Retrieved from https://www.spglobal.com/spdji/en/documents/additional-material/sp-500-brochure.pdf

Walsh D.M., & Yu-Gen Tsou, G. (1998). Forecasting index volatility: sampling interval and non-trading effects. *Applied Financial Economics*, *8*(5), 477-485.

van den Goorbergh, R.W.J., & Vlaar, P.J.G. (1999). Value-at-Risk Analysis of Stock Returns: Historical Simulation, Variance Techniques or Tail Index Estimation? De Nederlandsche Bank NV.

Zakamulin, V. (2016). Abnormal Stock Market Returns around Peaks in VIX: The Evidence of Investor Overreaction?. *Available at SSRN 2773134*.

Zhang, Y., & Nadarajah, S. (2018). A review of backtesting for value at risk. *Communications in Statistics-Theory and methods*, *47*(15), 3616-3639.

Appendix

A1. iGARCH

The iGARCH or Integrated GARCH model introduced by Engle and Bollerslev (1986) is an extension of the GARCH model with an additional restriction imposed on it. This restriction requires that the α and β coefficients in a traditional GARCH model must sum to 1, thus allowing Equation (10) (which is redisplayed here from Section 2.2.5 for convenience) to be re-written as shown in Equation (15).

$$\sigma_t^2 = \omega + \Sigma_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \Sigma_{k=1}^{q} \alpha_k \varepsilon_{t-k}^2 \tag{10}$$

$$\sigma_t^2 = \omega + \Sigma_{j=1}^{p} \beta_j \sigma_{t-j}^2 + \Sigma_{j=1}^{q} (1 - \beta_j) \varepsilon_{t-k}^2 \tag{15}$$

A2. Cox-Ross-Rubinstein and Black-Scholes model

The binomial option pricing framework introduced by Cox, Ross, and Rubinstein (1979) offers a simplified approach to option pricing. In this model, the price of the stock (the underlying asset of the option) is assumed to follow a binomial distribution such that there are two possible states of the world that can be considered: one in which the price of the stock is assumed to increase by a fixed factor and the other state of world where its price is assumed to decline by a fixed factor. The price of the option can then be determined by finding the value of a replicating portfolio that consists of a given number of shares of the stock and a specific amount of money that is either borrowed or lent (depending on whether a call or a put option is under consideration). This model can then be extended for multiple periods.

The Black-Scholes model (Merton, 1973; Scholes & Black, 1973) can be seen as a limiting case of the Cox-Ross-Rubinstein binomial framework for option pricing, where an infinite number of periods are assumed to exist between the time at which the option price is obtained and the time at which the option matures. Five inputs are required to price an option under this framework: the option's strike price, the current price of the option's underlying asset, the time to expiration, the risk-free rate, and the volatility.

A3. Formula for Kupiec's backtest

The test statistic for Kupiec's backtest can be computed using the following formula:

$$LR = -2\ln(\propto^{T-N} (1-\propto)^N) + 2\ln([1-(N/T)]^{T-N}(N/T)^N) \tag{16}$$

Where α refers to the significance level, T is the total number of observations, and N is the number of realized violations of the computed VaR.

A4. Formula for Christoffersen's test

For Christoffersen's conditional coverage test, a dummy variable I_t must be defined such that it takes a value of 1 when there is a violation of the VaR at time *t* and 0 otherwise. If $\pi_{i,j} = P(I_t = j|I_{t-1} = i)$, i.e., the transition probability for two consecutive dummy variables, then the likelihood function for the series of I_t can be defined as follows

$$L_1 = \pi_{0,0}^{n_{0,0}}\pi_{0,1}^{n_{0,1}}\pi_{1,0}^{n_{1,0}}\pi_{1,1}^{n_{1,1}} \tag{17}$$

where $n_{i,j}$ is the number of observations for which value *i* is followed by value *j*. The maximized likelihood function can then be seen to equal:

$$\widehat{L_1} = \left[\frac{n_{0,0}}{n_{0,0}+n_{0,1}}\right]^{n_{0,0}} \left[\frac{n_{0,1}}{n_{0,0}+n_{0,1}}\right]^{n_{0,1}} \left[\frac{n_{1,0}}{n_{1,0}+n_{1,1}}\right]^{n_{1,0}} \left[\frac{n_{1,1}}{n_{1,0}+n_{1,1}}\right]^{n_{1,1}} \tag{18}$$

Under the null of first-order independence, the likelihood can then be represented as

$$L_2 = (1-\pi)^{n_{0,0}+n_{1,0}}\pi^{n_{0,1}+n_{1,1}} \tag{19}$$

The maximized likelihood function is then:

$$\widehat{L_2} = \left[1 - \frac{n_{0,1}+n_{1,1}}{n_{0,0}+n_{1,0}+n_{0,1}+n_{1,1}}\right]^{n_{0,0}+n_{1,0}} \left[\frac{n_{0,1}+n_{1,1}}{n_{0,0}+n_{1,0}+n_{0,1}+n_{1,1}}\right]^{n_{0,1}+n_{1,1}} \tag{20}$$

The likelihood-ratio test statistic for first-order independence of the VaR violations is given by

$$LR_{ind} = -2(ln(\widehat{L_2}) - ln(\widehat{L_1})) \tag{21}$$

The likelihood-ratio test statistic for conditional coverage is then given by

$$LR_{cc} = LR_{uc} + LR_{ind} \tag{22}$$

where LR_{uc} represents the test statistic for unconditional coverage (computed as in Section A4).

A5. Full-sample optimization results

The process outlined in Section 4.1 is repeated over the entire data sample in order to facilitate comparison with the results obtained when using only the first 1000 observations. The results are presented in Table A5.1.

For the EWMA model, the optimal lambda varies between 0.915 and 0.952. In comparison to the results displayed in Column (1) of Table 5.1.1, this range of values is smaller. Regardless, the results for the optimal value of lambda over the first 1000 observations and over the entire data sample show substantial similarity. As in Column (1) of Table 5.1.1., none of the optimal lambdas for each index were found to be significantly different from 0.94, which again reinforces the prior conclusion that the *RiskMetrics* methodology for the EWMA model is generally applicable. Additionally, it can also be seen that the optimal value of the smoothing exponents lies close to the maximum possible value of one. Hence, the results found over the entire sample of data can be seen to align with the results found when considering only the first 1000 observations.

Table A5.1: Results of the optimisation process for the EWMA, GARCH, and GJR-GARCH models over the full data sample

	EWMA (1)	**GARCH (Gaussian distribution) (2)**	**GARCH (Student's t-distribution) (3)**	**GJR-GARCH (Gaussian distribution) (4)**	**GJR-GARCH (Student's t-distribution) (5)**
DAX	0.925	13.535	14.667	11.662	12.641
	(0.017)	(0.1399)	(0.1005)	(0.233)	(0.1795)
DJIA	0.941	6.3746	8.0438	5.994	5.9687
	(0.002)	(0.7019)	(0.5297)	(0.7405)	(0.7431)
FTSE 100	0.934	8.8319	7.5989	4.4369	4.946
	(0.005)	(0.4529)	(0.575)	(0.8804)	(0.839)
HSI	0.919	10.885	10.316	9.6924	8.9662
	(0.013)	(0.2837)	(0.3255)	(0.376)	(0.4404)
KOSPI	0.952	3.5244	4.4084	3.2852	4.3095
	(0.002)	(0.9343)	(0.8825)	(0.9519)	(0.8899)
NASDAQ 100	0.932	13.396	9.357	5.2227	4.227
	(0.005)	(0.1455)	(0.405)	(0.8145)	(0.8958)
NIFTY 50	0.923	3.8043	4.0336	4.6134	4.7685
	(0.004)	(0.9238)	(0.9092)	(0.8666)	(0.854)
Nikkei 225	0.915	6.5734	7.0543	7.0543	4.6677
	(0.004)	(0.6814)	(0.6315)	(0.6315)	(0.8623)
Russell 2000	0.926	8.7222	9.7386	12.063	13.592
	(0.004)	(0.4633)	(0.3721)	(0.2098)	(0.1376)
SMI	0.919	6.0078	5.6672	7.5364	7.0192
	(0.005)	(0.7391)	(0.7727)	(0.5815)	(0.6351)
S&P 100	0.942	5.9267	7.5844	2.9909	3.0464
	(0.003)	(0.7472)	(0.5765)	(0.9647)	(0.9624)
S&P 500	0.934	6.7771	7.656	4.1022	4.8326
	(0.003)	(0.6603)	(0.5691)	(0.9046)	(0.8487)
S&P/TSX 60	0.916	14.442	12.941	11.332	10.831
	(0.005)	(0.1074)	(0.1653)	(0.2537)	(0.2875)

Notes: Column (1) reports the optimal value of the decay factor in the EWMA model, which equals the value of the β coefficient obtained from the estimation of an iGARCH (1,1) model. The results of the one-sample t-test is also displayed, where the significance of the test statistic and its corresponding p-value is represented using stars (*p<0.10, **p<0.05, ***p<0.01). Columns (2), (3), (4), and (5) report the test statistic and the corresponding p-value for the Li-Mak test conducted on the GARCH (1,1) and GJR-GARCH specifications under the null hypothesis that there is no autocorrelation in the squared and standardized residuals of the model. The p-values are reported in parentheses. All statistics are computed using the first 1000 observations of data for each index. All figures are reported to 3 decimal places.

However, for the GARCH models, some divergence can be seen between the results over the first 1000 observations and the results for the entire data sample. The GARCH (1,1) specification was found to be optimal for all indices except the DAX and the S&P/TSX 60. In the case of the DAX, the optimal specification was found to be a GARCH (2,1) when assuming the standardized residual to follow a Student's t-distribution, whereas assuming a Gaussian distribution resulted in a GARCH (1,1) specification being optimal. On the other hand, the optimal specifications for the S&P/TSX 60 index displayed considerable divergence from the results for the rest of the indices. When assuming the standardized residuals to follow a Gaussian distribution, the GARCH (3,3) model was the most parsimonious specification that resulted in an insignificant result for the Li-Mak test, while this was the case for the GARCH (6,3) specification when assuming a Student's t-distribution. Overall, the GARCH (1,1) is still found to be the optimal specification for the majority of indices, thus echoing the prior conclusion of this specification being generally suitable when using the GARCH model. However, this slight heterogeneity shows the importance of determining the optimal lag structure on a case-by-case basis.

For the GJR-GARCH model, the results strongly align with those found over the first 1000 observations of data. The optimal specification was found to be the GJR-GARCH (1,1) model for all indices except for the Russell 2000. Specifically, when assuming a Student's t-distribution for the standardized residual, the optimal specification is a GJR-GARCH (2,1) model, but this is the only exception.

A6. Importance of the constant term in the autoregressive model specification

As Table A6.1 shows, excluding the constant term in the autoregressive model results in a severe deterioration in performance. The realized violations increase substantially in comparison to the autoregressive model with a constant, with some indices showing almost five times as many violations. Like the EWMA model, the realized violations are seen to exceed the expected number of violations for every index, suggesting that the volatility and, hence, the VaR have been consistently underestimated. As a result, each index displays a significant result at the 1% level for both Kupiec's backtest and the CCC test. This indicates that the realized failure rate is significantly different from 5% and that the VaR violations are correlated over time for all indices.

The severe decline in performance of the autoregressive model following the exclusion of the constant term suggests that the lack of a constant term in the specification for the EWMA model is indeed responsible for its poor performance.

Table A6.1: VaR violations for the autoregressive model with and without the constant term

	Expected violations	AR model	AR model w/o constant
	(1)	(2)	(3)
DAX	736	693###	4649***###
DJIA	1131	1043***###	6869***###
FTSE 100	203	211	1135***###
HSI	600	563###	3564***###
KOSPI	514	525###	3089***###
NASDAQ 100	202	239***##	1023***###
NIFTY 50	335	297**#	2018***###
Nikkei 225	593	609###	3506***###
Russell 2000	503	501#	2773***###
SMI	201	234**###	1031***###
S&P 100	533	521	3165***###
S&P 500	771	770###	4551***###
S&P/TSX 60	201	218##	1078***###

Notes: this table compares the performance of the autoregressive model with and without a constant in modelling the VaR using the historical data of each specific index. Column (1) displays the expected number of violations based on the number of observations for which the volatility was forecasted multiplied by the significance level at which the VaR will be evaluated, i.e., 5%. Column (2) indicates the number of violations of the VaR computed using the volatility forecasts from the autoregressive model with a constant, while Column (3) reports the same for the autoregressive model without a constant. The results of Kupiec's two-sided backtest is also displayed in the above table, where the significance of the test statistic and its corresponding p-value is represented using stars (*p<0.10, **p<0.05, ***p<0.01). The results for the CCC test are also shown, where the significance of the test statistic and its corresponding p-value is represented using hashtags (#p<0.10, ##p<0.05, ###p<0.01).