Empirical Essays On Volatility Forecasting

XINPENG ZHANG

This thesis is submitted for the degree of PhD
at the

School of Management
University of Stirling
Dec 2023

Declarations

1. Candidate's declarations:

I, XINPENG ZHANG, hereby certify that this thesis, which is approximately 45000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

I was admitted as a research student in January 2019 and as a candidate for the degree of PhD in February 2020; the higher study for which this is a record was carried out in the University of Stirling between 2019 and 2025.

Date: **Signature of candidate: Supervisor's declarations**

I, hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of Stirling and that the candidate is qualified to submit this thesis in application for that degree.

Date: **Signature of supervisor:**

2. Permission for electronic publication

In submitting this thesis to the University of Stirling I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and the abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use unless exempt by award of an embargo as requested below, and that the library has the right to migrate my thesis into new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration or have requested the appropriate embargo below.

The following is an agreed request by candidate and supervisor regarding the electronic publication of this thesis:

Access to printed copy and electronic publication of thesis through the University of Stirling.

Date **Signature of candidate:** **Signature of supervisor**

Empirical Essays On Volatility Forecasting

XINPENG ZHANG

This thesis is submitted for the degree of PhD
at the

School of Management
University of Stirling
Dec 2023

Abstract

This thesis aims to examine and improve forecasting performance for both univariate volatility and multivariate covariance models. This thesis investigated the forecasting ability of volatility models and covariance models including univariate GARCH methods, HAR models, and multivariate DCC process on stock indices in twelve countries. Moreover, several hybrid models combined by the current GARCH genres and neural networks are investigated in three empirical exercises. The accuracy of forecasting by different models is addressed. There are four main contributions of this study. First, the comparison among the univariate normal GARCH genre, HAR model and hybrid models by neural networks reveals that the hybrid models are superior to others which gives an empirical result in a wide comparison. The policymakers can benefit from the results to formulate their policies to avoid risk. Second, with the application of DCC process, the new multivariate model built by neural networks are preferred rather than original DCC GARCH models when forecasting covariance which give some empirical results on multivariate covariance forecasting. The results are able to provide some suggestions for market managers on risk control, especially for the portfolios containing multivariate assets in different countries. Third, the trading volume is found to be useful for improving volatility forecasting in the hybrid process. Finally, the original neural networks are improved by a deep learning model which has more hidden layers than the previous neural networks. The forecasting ability of all the models are investigated and the hybrid model built with deep learning are still superior. This research provides valuable insights and a reliable framework for improving stock volatility predictions.

List of contents

1. Introduction

In the past few decades, with the rapid development of the stock market, lots of events related to risk management in the stock markets happened. Some of them caused huge losses and nearly destroyed the stock market. On 19^{th} October 1987, which is called "Black Monday", the New York Stock Exchange (NYSE) lost more than US$500 billion in market capitalization. Between October 19 and 23, the stock crash has influence on global market include: United Kingdom, Japan, USA, Hong Kong, Frankfurt, Amsterdam, Mexico City and Sydney, etc. (Carlson, 2007). Another one is the global financial crisis happened in United States of America in 2008 and then spread to the whole world, it caused billions of losses, and the market appeared to get down in a long time. Therefore, the stock volatility has attracted lots of attention in the finance literature, lots of papers were proposed to investigate the volatility. Although various models and measurements have been proposed, there is still some space for the improvement of the volatility forecast. With this thesis, we wish to follow the steps of the current literature and conduct research by comparing the existing models and investigate some new hybrid-built models by using machine learning methods. Some key parameters will be explored as well in the improvement of the volatility forecast.

Since the market can be highly volatile, the volatility is directly related to market uncertainty, and it will affect anyone in the markets. By the view of Bhowmik & Wang (2020), in financial markets, volatility is mainly reflected in the deviation of the expected future value of assets, and it represents the uncertainty of the future price of an asset. A good estimate or forecasting of the volatility can present significant investment risk. Investors who deal with derivative securities, accurate and precise volatility forecasting can help them manage their portfolios and get more chance to succeed in trading which will generate more profits or avoid losses.

Moreover, volatility stands an important position in investment, security valuation, risk management and monetary policy making. Poon & Granger (2001) mentioned on page 478 that "*Volatility forecast crucially affects investment choice and is the key input to the valuation of corporate and public liabilities. The forecast of volatility is also the most important parameter affecting prices of market-listed options, of which trading volume has quadrupled in the last decade*."

Since volatility plays an important role in both theoretical and empirical applications in finance, the forecasting of volatility forecasting has obviously become crucial to almost anyone who is involved in the financial markets, it will affect the markets as well as the whole economy in a country even in the world. An accurate volatility forecasting becomes an important mission in the finance area including portfolio pricing, hedging, and option strategy, etc. It will help the market stay stable and avoid a potential crash. In addition, a good understanding and measurement of volatility can be useful across different areas including politics, banking, risk management, individual investors, enterprises and lots of other institutions.

With this thesis, we aimed to add some new knowledge to the literature on volatility forecasting including a hybrid volatility forecasting model with neural networks, and several investigations on different normal models with different parameters to be considered in the modelling process. The reason to use the neural networks is that the method does not need too much formal statistical training. It can not only detect complex nonlinear relationships between dependent and independent variables but detect all possible interactions between predictor variables as well addressed by Tu (1996). Moreover, it can be simple to combine any other different models with their training algorithms.

There are four main contributions of this thesis:

First, the hybrid model built with GARCH genre and simple neural networks were used to make forecasts of univariate volatility. After performing the volatility exercise among the hybrid models, GARCH series and HAR model, it suggests that the hybrid models are superior to the GARCH series or HAR models which gives an empirical result in a wide comparison. The policymakers can benefit from the results to formulate their policies to avoid risk. Second, when forecasting the covariance/correlations, a model which combined the neural networks and the DCC process was considered. After comparing the performance between several DCC GARCH models and hybrid models, the hybrid models appeared to be better when considering the covariance/correlation forecasting which give some empirical results on multivariate covariance forecasting. The results are able to provide some suggestions for market managers on risk control, especially for the portfolios containing multivariate assets in different countries. Third, the hybrid models were improved by a machine leaning method called Deep Learning. More hidden layers were considered in the process and the machine with more hidden layers has a better performance than the original machine. Forth, with the contribution of trading volume on the forecasting ability, both the GARCH model and hybrid models has a better forecasting performance when considering the trading volume and the hybrid models with Deep Learning performs best among all the models. The empirical results of the models with consideration of trading volume provides a view for further research on the volatility forecasting when using deep learning methods. It provides valuable insights and a reliable framework for improving stock volatility predictions. Apart from the main contribution mentioned above, some other issues are considered as well. Since some of the exist literature focus on a single model or very limited data, this thesis provides a more comprehensive comparison among nine models in a wide range of stock index from twelve different countries. Both developed and emerging countries are included. It is better to have a whole view of the trend or dynamics from different angles.

The structure of the thesis contains a literature review chapter, three empirical chapters and a conclusion chapter. The next chapter (**Chapter 2**) is the literature review chapter. The evolution of the volatility as well as a definition of the volatility in stock market are introduced. The importance and implications of forecasting volatility are given by different literatures. Some stylized facts of the data are introduced including non-normality, volatility clustering, long memory, asymmetric volatility phenomenon, co-movements, asymmetric vertical dependence. After that, a list of GARCH models, the HAR model and several machine learning methods are noted. The models introduced in this chapter will be used in the rest of the thesis. After that, a state of the literature is proposed which several of research on volatility in past two decades are reviewed and the research question of this thesis are addressed.

In **Chapter 3,** an empirical comparison exercise among different volatility models is carried out. This is a very popular topic in financial literature. However, there does not exist a general conclusion which model performs the best. Since numerous studies in financial literature have conducted lots of comparisons among the normal volatility models such as simple models like Exponential Smoothing (ES) and the Moving Average model (MA), ARCH/GARCH type models, this chapter uses a hybrid model built with GARCH models and neural networks. A new comparison between three model categories is proposed: the normal ARCH/GARCH models, Heterogenous autoregressive models (HAR), and the hybrid models built with neural networks.

A wide range of stock index from twelve countries is selected in order to make more comprehensive view of the comparison between different models. The daily closing price of the stock index from these countries is obtained. Five measures of comparison techniques are applied in the exercise based on a loss function with the realized volatility. A model confidence set is created at the end to identify a better model at a certain confidence interval. There are two main results: among the comparison of GARCH series model, the long memory models outperform the asymmetric models and HAR model, the other result is among the whole comparison between GARCH series, HAR model and hybrid models, the hybrid-built type models outperform the normal GARCH or HAR models. The asymmetric GARCH models built with neural networks have a better forecasting performance rather than other types of hybrid models.

In **Chapter 4,** another empirical comparison exercise among different covariance models is carried out. There exists lots of papers talking about the volatility behaviors of a univariate asset, while a discussion of the conditional covariance or correlation among multivariate assets is limited. This chapter (**Chapter 4**) makes an empirical comparison exercise among different covariance models. There are different methods to capture the dynamics of the covariance such as the vector error correction model (VEC), the direct extension of univariate GARCH model called BEKK model, the Constant conditional correlation (CCC) model, the standard dynamic conditional correlation (DCC) model, etc. Following the step of previous chapter (**Chapter 3**), the standard dynamic conditional correlation (DCC) GARCH models are selected and a new hybrid model built with neural networks and DCC GARCH models are introduced. Comparisons among the normal multivariate DCC GARCH models and the new built hybrid models are created. The same daily closing price of the stock index is used from **Chapter 3**. Since the covariance matrix of all the countries will be a huge matrix and most of the procedure could be iterated, we take a covariance forecasting between USA and all other countries as an empirical exercise to make the chapter more concise and direct to understand and read. The realized covariance will be used to test the distance between the forecast series and the "actual" with the application of four comparison techniques. A model confidence set will be created as well to test the forecasting ability of different models. There are two main results: The new proposed method built with the DCC model and neural networks outperforms the traditional DCC GARCH models when forecasting covariance, particularly, the EGARCH DCC process built on neural networks has the best forecasting performance within the whole comparison technique. Likewise, the results show that there is no strong evidence that the results from conditional variance forecast in DCC process (Step one) will have greater impact on the forecasting of the conditional covariance.

In **Chapter 5**, the effect of the trading volume on the volatility forecasting is investigated. Moreover, the forecasting ability of a new machine built with more hidden layers will be tested as well. The relationship between the volatility of financial markets and trading volume has attracted a great deal of attention during the past three decades, lots of paper and several theoretical models have been developed to investigate the relationship between the volatility and trading volume. We will explore the impact of trading volume on the volatility forecasting with the application of the GARCH series. Furthermore, an improved machine with more hidden layers which could be called Deep Learning Machine generalized from **Chapter 3** will be introduced and a test of the forecasting ability will be carried out. Several comparisons will be created including standard GARCH with/without considering trading volume, the old hybrid model with/without considering trading volume, the new hybrid model with/without considering trading volume.

The same dataset from **Chapter 3** is used in order to make the comparison with different models in the same out-of-sample forecasting. There are two main objectives of the exercise:

one is to test the impact of the trading volume on volatility forecasting, the other is to test the forecasting ability of new hybrid models with more hidden layers when training the machine. Four same comparison techniques will be used for consistency. The results show that the accuracy of the volatility forecasting will be improved with the consideration of the trading volume not only in traditional GARCH series, but the hybrid models as well. Moreover, with the application of more hidden layers in neural networks, the ability of volatility will be improved.

Finally, **Chapter 6** summarizes the main findings: the univariate volatility forecasting performance among GARCH series, HAR model and hybrid models with neural networks; the forecasting ability conditional covariance of DCC GARCH model and neural networks; the superiority of the GARCH series models, the old machine with less hidden layers, and the new machine with more hidden layers and the effect of the trading volume on volatility forecasting.

2. Literature review

Abstract

The main aim of this chapter is to review literature and related works about the measurement and forecasting of volatility. A list of Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, Heterogeneous Autoregressive (HAR) model and hybrid GARCH model based on Machine Learning methods (ML) are introduced. With the application of the GARCH models which is the one of the most widely used volatility models, the characteristics and dynamics of volatility could be captured. Moreover, with the development of machine learning methods, a hybrid GARCH model based on Machine Learning methods has proved to be effective to enhance the forecasting performance of traditional GARCH models. However, there still existed a debate which model can be considered as the “best” one for volatility forecasting. However, some of the empirical results by hybrid GARCH models were very limited. They are limited to using data from one or two assets, which is not sufficient to give a general result. Likewise, some related work only considered a corresponding hybrid GARCH model which has a good estimation in original GARCH step which has a lack of reporting the full performance of all the hybrid GARCH models. In this chapter, the literature about volatility and covariance is reviewed and the superiority of different models by different empirical exercises are addressed.

2.1 Introduction

In the past twenty years, with the rapid development of the stock market, lots of events related to risk management in the stock markets have happened. There are some events which are very famous. On 19th October 1987, which is called "Black Monday", the New York Stock Exchange (NYSE) lost more than US$500 billion in market capitalization. Between October 19 and 23, the stock crash has influence on global market include: United Kingdom, Japan, USA, Hong Kong, Frankfurt, Amsterdam, Mexico City and Sydney, etc. which is investigated by Goldman Sachs (2019). Another is the 2008 global financial crisis happened in United States of America and then spread to the whole world, it caused lots of billions loss and the market appeared to get down in a long time by Duffie (2019). By the report of Stevenson (2020), the 2015 stock market crisis happened in China which made lots of people lose their house. Therefore, risk measurement appeared to be crucial to avoid loss and market disruption, which leads to a popularity of studying on topic of stock market volatility.

In the early years, funds were judged largely by the performance of fund managers until Markowitz (1952) argued that fund performance should be judged compared to the amount of risk it takes, which can be considered as the early stage for volatility measurement. The "risk" was a vague concept which cannot be described as a number and unobservable so that the volatility is considered to be the "variance" by Markowitz.

Studies used measures of the variance or "volatility" of speculative asset prices to provide evidence against simple models of market efficiency. According to the work of LeRoy & Porter (1981) and Shiller (1979), the measures were interpreted as implying that prices show too much variation to be explained in terms of the random arrival of new information about the fundamental determinants of price in connection with stock price, bonds, dividends, foreign exchange rates, etc.

The entire financial industry started using volatility to measure risk in the early times. After the crash, a financial institution, JPMorgan, began to use a new daily report which would describe the loss of a bank on its trading positions on any given day. The report then comes to be a new tool that is known as "Value-at-Risk" (VaR). The VaR report was first designed by a banker called Till Guldimann (1980), who used the historical volatility of markets to calculate the maximum the bank could lose on any given day, with 95% certainty. Later the VaR estimate is defined as the minimum expected loss with a 1% confidence level (sometimes 5% confidence level) for a given time horizon (usually 1 or 10 days) which is readily available to give volatility estimates and widely used among banks and trading institutions.

The changes in volatility can be influenced by lots of factors. Becketti & Roberts (1990) suggested that stock market volatility can be divided into two types: normal volatility and jump volatility. In a simple word, the normal volatility is the ordinary variability of stock prices, that is, the normal movements in stock prices, while the jump volatility refers to accidental and sudden extreme changes in stock prices such as the price shocks in 1987, the volatility jumped up from roughly 20 percent to over 50 percent although jumps in returns generate more than half of the crash in 1987 while high volatility explains the rest. Eraker et al. (2003) mentioned that "*It is especially important to determine the contribution of jumps to periods of market stress because jump risk, either in returns or in volatility, cannot typically be hedged away, and investors may demand a large premia to carry these risks*". Jumps in stock prices over a relatively short time-period which may lead to a jump of volatility, and it can temporarily disrupt capital markets. Furthermore, with an increased jump volatility, the confidence of investor will be reduced dramatically in the stock market, leading to decreased trading activity, lower liquidity, and higher investment cost, all of which could seriously threaten market stability documented by Darrat & Rahman (1995). For example, stocks are more volatile than bonds, so investors would shift their portfolio to a more stable asset to avoid risk. Similarly, the shares of smaller companies tend to trade less than bigger companies' shares in order to make profits with less risks. In the early time, lots of researcher blame the future trading activities as the source of the increased jump volatility. However, Darrat & Rahman (1995) suggested that futures trading activity (however measured) is not a force behind the recent episodes of jump volatility. Similarly, Becketti & Roberts (1990) argued that recent regulations aimed at reducing the general level of futures activity are unwarranted and would not contribute to a more stable environment in the stock market.

Jones et al. (1994) suggested that public information is the major source of short-term return volatility by investigating the stock price movements with the effect of public information. The main reason could be considered as the action of the investors to the information, which will lead to high fluctuations (volatility) in the market. The crash in 1987 was a good symbol that the investors gave responds to the market investigated by the survey of Campbell & Shiller (1988).

The monetary policy on the stock market is also a factor which cannot be ignored. In general, a loose monetary policy may increase the probability of a rise in the stock market. Relatively, a tight monetary policy may increase the probability of a drop in the stock market. The policymaker can either reduce the market volatility by investing money in the market which will increase the confidence of the market by a loose monetary policy. Alternatively, the policymaker can stay calm with a tight monetary policy, standing as a lender to the market when the probability of a financial crash rises and be ready to save the market addressed by Mishkin (1999).

Volatility plays a really important role almost everywhere in financial market. It is directly linked to market uncertainty or risks and affects the investment behaviors of investors and financial institutions. A high volatility is usually found by a wide range of fluctuations in the prices over short time periods. It means security values are not dependable, and the capital markets are not functioning as well as they should mentioned by Figlewski (1997). In financial markets, the issue of volatility forecasting is indispensable since it is often taken to represent the risk and plan the investment portfolio. In recent years, studies follow the trend to improve the ability of forecasting volatility in financial markets. A reliable forecast of the volatility will direct investor to make a more correct decision, especially for institutions involved in pricing financial derivatives like options trading and portfolio management since almost all modern option-pricing techniques rely on a correct volatility parameter for price evaluation. Mathematically, a model of asset pricing without taking volatility into account will have an unpersuasive result. With the assistance of modelling the volatility, the dependencies between the current values of the financial indicators and their future expected values will be detected.

The main aim of this chapter will be set to explore the issues happened in volatility forecasting, especially, the characteristics of volatility models in stock market which explained by the researcher and investors. For beginning in this section, a list of different models or approaches to estimate volatility are introduced below.

2.2 Modelling Volatility

In the financial market, a clear understanding of both return and risk is the first necessary task for either investing or research. The return is widely considered as a change of one certain asset over a period while the risk is considered to be a potential variability or volatility of the return. The measurement or prediction of the volatility has held the attention of academics and practitioners over the last two decades although several models and methods are introduced. In the next section, an early definition of volatility and models is reviewed.

2.2.1 Early stage of volatility measurement

Some conceptual issues including volatility, standard deviation, and risk are elucidated by the early work of Poon & Granger (2003). In finance, volatility is often used to refer to standard deviation, σ, or variance, σ^2, computed from a set of observations as

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{t=1}^{N}(r_t - \bar{r}_t)^2 \qquad (2.2.1)$$

where $\bar{r}_t$ is the mean return.

The sample standard deviation statistic $\hat{\sigma}$ is a distribution free parameter representing the second moment characteristic of the sample.

Figlewski (1997) noted that since the statistical properties of sample mean make it a very inaccurate estimate of the true mean, especially for small samples, taking deviations around zero instead of the sample mean as in equation (2.2.1) typically increases volatility forecast accuracy. In his work, he constructed a good concept of volatility by using "efficient markets hypothesis" and "random walk" model. In an efficient market, asset price movements can be described by an equation as follows:

$$r_t = \frac{S_t - S_{t-1}}{S_{t-1}} = \mu_t + \varepsilon_t \tag{2.2.2}$$

where $E(\varepsilon_t) = 0;\ var(\varepsilon_t) = \sigma_t^2$

The term, "return" r_t means the rate of change in the asset price *S* that occurs between two time periods from t-1 to *t*. In his work, the return is the sum of a nonrandom mean return μ_t of period *t*, plus a zero mean random disturbance ε_t which is independent of all past and future values. It means that $\{\ldots \varepsilon_{t-1}, \varepsilon_t, \varepsilon_{t+1}, \ldots\}$ are independent. Since $\{\ldots \varepsilon_{t-1}, \varepsilon_t, \varepsilon_{t+1}, \ldots\}$ is independent, he mentioned "*the lack of serial correlation defines characteristic of efficient market pricing: past price movements give no information about the sign of the random component of return in period t*".

A continuous time analogue can be obtained by

$$\frac{dS}{S} = \mu dt + \sigma dz \tag{2.2.3}$$

The formula comes from the logical extension of the random walk model to continuous time which was adopted by Black & Scholes (1973). They aimed to model stock price movements over a very short interval of time by deriving the option price formula so that it is possible for them to consider a trading strategy in a short time interval. The formula seeks to observe the price movements by limiting the time to zero. With this measure, the mean return μ and the standard deviation of return σ which is defined as the volatility will be a constant over a year. *dz* is a time independent random disturbance with mean zero and variance $1 * dt$.

In the long history of research of modelling volatility, there are lots of models generated to measure or capture the volatility. Volatility forecasting becomes a hot topic in recent years. The volatility can be measured as mentioned above by standard deviation or variance of the returns. Alternatively, it can be measured by the Black & Scholes formula. Among these measurements, the most important part is to consider a return process. The model describing the returns of an asset at time *t* can be defined as

$$X_t = \mu + \sigma_t \upsilon_t \qquad (2.2.4)$$

where μ is the mean process which could be an AR process, MA process or ARMA process, ν_t and σ_t are independent.

$$\varepsilon_t = \sigma_t \upsilon_t \qquad (2.2.5)$$

$\{\sigma_t\}$ is non-negative stochastic process *t* for a fixed *t*. $\{\nu_t\}$ is a sequence of independent and identically distributed (i.i.d) random and symmetric random variables. Volatility process is identified by $\{\sigma_t\}$. The time series $\{X_t\}$ and the volatility process $\{\sigma_t\}$ are assumed to be strictly stationary. Stock return is unique from returns of other types of investments because of the implication of continuous compounding concept. Such nature of stock return requires its measurement from a natural.

2.3 Stylized facts of the asset return and volatility

Following the work of Tsay (2005), financial time series data analysis was different from other time series analysis due to their characteristics. The empirical time series contained an element of complex dynamic system with high volatility and a great amount of noise. The uncertainty and noise make the series exhibit some empirical regularity, which are known as stylized facts. At an early stage, research on the empirical distribution of the stock returns found that the distribution often appeared to be different from the normal distribution. The empirical distribution of the stock returns turned out to be leptokurtic or skewed (either right or left). This leads to a non-constant variance over time and the volatility tends to be clustering. Both non-normality and clustering are identified as the stylized fact or empirical regularity of the returns and volatility. There are lots of stylized facts which are believed to be the truth, and the facts are necessary to be identified when making selection of models.

Non-normality

It is a widely accepted stylized fact that most financial asset returns exhibit non-normal distribution characteristics. At the early stage, by the work of Mandelbrot (1963), Fama (1965), and Cootner (1970) when studying the distribution of stock returns, they reported that the empirical distribution of a return series has properties different from the normal distribution. By observing the real asset returns, the extreme event appears to be larger than what is assumed by common data generating process (normal distribution). The returns distribution has fat tails with a more peaked central section. The non-normality of market returns is also confirmed by the work of Sheikh & Qiao (2009). When they try to make an asset allocation decision, the extreme negative events will be observed in a much higher frequency than current risk frameworks allow for caused by the non-normality of asset returns.

Volatility clustering

Mandelbrot (1963) observed a feature of returns that “large changes tend to be followed by large changes of either sign; small changes followed by small changes.” which is represented as the volatility clusters. It can be understood that observing a large (small) return today (whatever its sign) is a good precursor of large (small) returns in the coming days. Since the distribution of the stock returns may not follow the normal distribution in most cases (non-normality mentioned above), the variance of the returns will not keep a constant over time, and this may lead to volatility clustering i.e. volatility is not constant and tends to cluster through time. The research by Niu & Wang (2013) used an autocorrelation analysis to demonstrate the volatility clustering properties for the actual return series.

Long memory

With the work of Ding et al. (1993) by studying the property of the stock market returns, they suggested that not only there is substantially more correlation between absolute returns than returns themselves, but the power transformation of the absolute return also has quite high autocorrelation for long lags. A simple explanation with the work of them is that the volatility exhibits a significant autocorrelation even for very long lags which indicates that the changes of volatility typically have a long-lasting impact on its after performance. The property can be identified as a long memory effect.

Asymmetric volatility phenomenon

The degree of leverage with a firm value will have an impact on its stock volatility which is commonly explained by the asymmetry property or leverage effect. By the early work of Black & Scholes (1973), they discussed the impact of leverage on stock price behavior, and the argument was documented by Merton (1974), Galai & Masulis (1976), Geske (1979). It could be described as the movements of stock price displayed a negative relation with volatility, a decrease of returns will lead to more volatility changes rather than a same amplitude increase of returns. The phenomenon has been confirmed by Christie (1982), Schwert (1989), Glosten et al. (1993), Braun et al. (1995), etc.

Co-movements of volatility

In a simple word, the returns and volatility of different assets and different markets tend to move together. Black & Scholes (1973) documented that when volatility changes, they all tend to change in the same direction. Lots of researches give evidence of this fact. Kim & Rogers (1995) examined the spread of volatility from Japan and USA to Korea. A conclusion made by them is that the spillover effect on the volatility of returns is more than on returns themselves. Subsequently, Liu & Pan (1997) examine stock return and volatility spillover effects from the developed markets to emerging stock market and has discovered that the volatility changes do have a spread over the international stock markets.

Asymmetric vertical dependence

In their work, Gençay et al. (2010) documented a property called asymmetric vertical dependence which indicated that it was asymmetric in the sense that a low volatility state (regime) at a long-time horizon was most likely followed by low volatility states at shorter time horizons. On the other hand, a high volatility stated at long time horizons did not necessarily imply a high volatility stated at shorter time horizons. His work gave evidence that volatility is a mixture of high and low volatility regimes, resulting in a distribution that is non-Gaussian. This result has important implications regarding the scaling behavior of volatility.

Since the volatility and return has some stylized facts or empirical regularities, the simple time series model in the early-stage face difficulties to capture the facts. With the help of the work by Engle (1982), a famous model called the Autoregressive Conditional Heteroscedasticity model (ARCH) was proposed in order to capture the dynamics of the volatility mentioned above.

2.4 ARCH/GARCH modelling

In this section, a series of ARCH/GARCH models are reviewed in finance literature, different GARCH type models are selected in order to capture the stylized facts.

Autoregressive conditional heteroskedasticity (ARCH) model

The autoregressive conditional heteroskedasticity (ARCH) model is one of the most well-known models to measure time-varying volatility. The model was proposed by Engle (1982) which is a mean zero serially uncorrelated processes. In his work, this model was used to estimate the means and variances of inflation in the United Kingdom. The process contains two important concepts. One is the conditional variance which is a nonconstant variable conditional on the past information. The other is the unconditional variance which stays constant.

The ARCH effect documented by Engle (1982) which described the conditional variance can be specified as

$$\sigma_t^2 = var(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) \tag{2.4.1}$$

The conditional variance σ_t^2 is time-varying by the past information. In general ARCH (q) can be represented as:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-1}^2 \tag{2.4.2}$$

The financial time series analysis has become more comprehensive by the invention of ARCH. The ARCH models took care of volatility clustered errors, nonlinearity and changes in the econometrician's ability to forecast addressed by Bera & Higgins (1993).

Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models

The GARCH model developed independently by Bollerslev (1986) is a generalization of the Autoregressive Conditional Heteroscedasticity (ARCH) model. It is a combination of a mean model and a variance process in order to model the conditional heteroskedasticity in a parsimonious way. The conditional variance equation in the simplest case, GARCH (1,1) can be written as the form:

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \qquad (2.4.3)$$

The conditional variance is represented as σ_t^2 and it calculates the variance based on the history estimates.

$$\varepsilon_t = \sigma_t \upsilon_t \qquad (2.4.4)$$

The mean error is normally distributed with a zero mean and a conditional variance σ_t^2 which can be changed over a time period.

The GARCH (*p, q)* specification is given by

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon_{t-1}^2 + \beta(L)\sigma_{t-1}^2 \quad (2.4.5)$$

$$\alpha(L) = \alpha_1 L + \alpha_2 L^2 + \cdots + \alpha_q L^q;\ \beta(L) = \beta_1 L + \beta_2 L^2 + \cdots + \beta_q L^p \quad (2.4.6)$$

Since the conditional variance σ_t^2 is clearly positive, the parameter on the right hand of the function should satisfy the "non-negativity constraints", which means $\omega > 0, \alpha \geq 0, \beta \geq 0$. The main idea is that σ_t^2, the conditional variance of ε_t^2 given information available up to time $t-1$, has an autoregressive structure and is positively correlated to its own recent past and to recent values of the squared returns ε_t^2. This captures the idea of volatility (conditional variance) being "persistent" or clustering: large (small) values of ε_t^2 are likely to be followed by large (small) values. The GARCH models permit a wider range of behaviors. It allows the conditional variance to be dependent upon previous own lags, that is the huge values of the lags(q) will not influence the accuracy of the prediction.

Exponential GARCH (EGARCH) models

The GARCH model, mathematically, explains conditional variance by formulating a linear regression between the squared disturbance error term in return process and the past variance. There is a limitation of this method that an equal size of positive and negative news will have the same impact on the conditional variance. In empirical research, an asymmetric effect usually existed which was explained by Black & Scholes (1973). A negative shock will increase more with the conditional variance than a same magnitude of positive shock, which is called a "leverage/asymmetric effect". A widely used asymmetric GARCH model put forward by Nelson (1991) namely the exponential GARCH (EGARCH) model provides a first explanation for the σ_t^2 depends on both the size and the sign of lagged residuals in the return process. In particular,

$$ln(\sigma_t^2) = \omega + \sum_{i=1}^{p} \beta_i ln(\sigma_{t-i}^2) + \sum_{i=1}^{q} \alpha_i \{\phi v_{t-1} + \psi[|v_{t-i}| - E|v_{t-i}|]\} \qquad (2.4.7)$$

Consider p=1 and q=1, $\{v_t\}$ is a sequence of *i.i.d* with a zero mean, the function could be a simple EGARCH (1,1)

$$ln(\sigma_t^2) = \omega + \ ln(\sigma_{t-1}^2) + \alpha\phi v_{t-1} + \alpha\psi|v_{t-i}| \qquad (2.4.8)$$

If define

$$g(v_t) = \phi v_t + \psi|v_t| \qquad (2.4.9)$$

Here $\{g(v_t)\}$ is a zero-mean, *i.i.d* random sequence since the components of $g(v_t)$ are ϕv_t and $\psi|v_t|$ are zero mean. With the assumption in return process that $\{v_t\}$ is normally distributed, the components of $g(v_t)$ are orthogonal, though they are not independent. If $0 \leq v_t < \infty$, with refers to a positive news, $g(v_t)$ is linear in v_t with slope $\phi + \psi$, and if $-\infty \leq v_t < 0$, with refers to a negative news $g(v_t)$ is linear with slope $\phi - \psi$. If $\phi < 0$, $g(v_t)$ allows for the conditional variance process $\{\sigma_t^2\}$ to respond asymmetrically to rises and falls in stock price by $|\phi - \psi| > |\phi + \psi|$ and the clustering is captured by the parameter ψ. In the EGARCH model $\ln(\sigma_t^2)$ is homoscedastic conditional on σ_t^2 , and the partial correlation between v_t and $\ln(\sigma_t^2)$ is constant conditional on σ_t^2.

Threshold-GARCH (TGARCH) models

Another alternative asymmetric model is selected as Threshold-GARCH (TGARCH) model. In GARCH model, by the assumption that the conditional variance is a linear combination of its squared errors and the historical variance, both the negative and positive shocks will cause a same impact on the variance. However, the leverage effect which is mentioned as the stylized fact of the volatility often occurs in empirical studies. It can be represented as a negative relationship between the returns and volatility. The leverage effect describes a negative shock will have a greater impact than the equal positive shocks on the variance. The Threshold-GARCH (TGARCH) model proposed by Zakoian (1994) and GJR GARCH model studied by Glosten et al. (1993) define the conditional variance as a linear piecewise function will capture the effect in most cases. In TGARCH (1,1):

$$\sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \delta D_t \varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \qquad (2.4.10)$$

$$D_t = \begin{cases} 1 & \varepsilon_{t-1} < 0 \\ 0 & \varepsilon_{t-1} > 0 \end{cases} \qquad (2.4.11)$$

It can be clearly seen that the leverage effect is captured by the function D_t with measuring the value of ε_{t-1}. Similar with EGARCH models, ω provides the weighted average of the variance, positive news will reflect by the value of α and negative news will be reflected by $\alpha + \delta$. If $\delta > 0$, the innovation in conditional variance is now positive when returns innovations are negative. The TGARCH model relaxes the linear restriction on the conditional variance dynamics. It could capture the stylized fact that conditional variance tends to be higher after a decrease in return than after an equal increase.

Component GARCH (CGARCH)

Component GARCH models, which were first proposed by Engle & Lee (1999), constitute a convenient method of incorporating long-memory-like features into a short-memory model, at least for the horizons relevant for option valuation. In CGARCH model, the volatility of returns consists of a long-run and a short-run component. With the work of Chan & Maheu (2002), he provides evidence that a component model can capture long-range volatility dynamics by Monte Carlo. The component GARCH (CGARCH) model following is derived by replacing the constant σ_t^2 with a time varying long-run component q_t. The conditional variance changes by a long run component which is calculated by itself, autoregressive of the first order. The CGARCH model specification is:

$$\sigma_t^2 = q_t + \alpha(\varepsilon_{t-1}^2 - q_{t-1}) + \beta(\sigma_{t-1}^2 - q_{t-1}) \qquad (2.4.12)$$

$$q_t = \omega + \rho q_{t-1} + \varphi(\varepsilon_{t-1}^2 - \sigma_{t-1}^2) \qquad (2.4.13)$$

The long-run component q_t will present the effect by computing the parameter ρ larger than $\alpha + \beta$, the value of $(\varepsilon_t^2 - \sigma_t^2)$ will control the time varying movement of the long-run effect. It constitutes a method of making the long-memory-like features into a short-memory model to capture the salient features of speculative returns.

The conditional variance is the main object of interest and there exists varieties of GARCH type models, other parametric specifications for volatility addressed by Hansen and Lunde (2005) are listed below:

IGARCH:

$$\sigma_t^2 = \omega + \varepsilon_{t-1}^2 + \sum_{i=2}^{q} \alpha_i(\varepsilon_{t-i}^2 - \varepsilon_{t-1}^2) + \sum_{j=1}^{p} \beta_j\left(\sigma_{t-j}^2 - \varepsilon_{t-1}^2\right) \quad (2.4.14)$$

TS-GARCH:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i\left|\varepsilon_{t-i}^2\right| + \sum_{j=1}^{p} \beta_j\left|\sigma_{t-j}^2\right| \quad (2.4.15)$$

A-GARCH：

$$\sigma_t^2 = \omega + \sum_{i=1}^{q}(\alpha_i\varepsilon_{t-i}^2 + \gamma_i\varepsilon_{t-i}) + \sum_{j=1}^{p} \beta_j\sigma_{t-j}^2 \quad (2.4.16)$$

NA-GARCH:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i(\varepsilon_{t-i} + \gamma_i\sigma_{t-i})^2 + \sum_{j=1}^{p} \beta_j\sigma_{t-j}^2 \quad (2.4.17)$$

V-GARCH:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i(\varepsilon_{t-i} + \gamma_i)^2 + \sum_{j=1}^{p} \beta_j\sigma_{t-j}^2 \quad (2.4.18)$$

Log-GARCH:

$$ln(\sigma_t) = \omega + \sum_{i=1}^{q} \alpha_i |\varepsilon_{t-i}| + \sum_{j=1}^{p} \beta_j ln(\sigma_{t-j}) \quad (2.4.19)$$

NGARCH:

$$\sigma_t^\delta = \omega + \sum_{i=1}^{q} \alpha_i |\varepsilon_{t-i}|^\delta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta \quad (2.4.20)$$

A-PARCH:

$$\sigma_t^\delta = \omega + \sum_{i=1}^{q} \alpha_i (|\varepsilon_{t-i}| - \gamma_i \varepsilon_{t-i})^\delta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta \quad (2.4.21)$$

GQ-ARCH:

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \varepsilon_{t-i} + \sum_{i=1}^{p} \alpha_{ii} \varepsilon_{t-i}^2 + \sum_{i=1}^{p} \alpha_{ij} \varepsilon_{t-i} \varepsilon_{t-j} + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \quad (2.4.22)$$

H-GARCH:

$$\sigma_t^\delta = \omega + \sum_{i=1}^{q} \alpha_i \delta \sigma_{t-i}^\delta \big(|\varepsilon_t - \kappa| - \tau(\varepsilon_t - \kappa)\big)^\upsilon + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta \quad (2.4.23)$$

2.5 Heterogenous autoregressive models (HAR model)

The standard heterogeneous autoregressive (HAR) model is perhaps the most popular benchmark model for forecasting return volatility mentioned by Clements & Preve (2021) It is estimated using raw realized variance (RV) and ordinary least squares (OLS).

Corsi (2004) first proposed the heterogeneous autoregressive (HAR) model for realized volatility, which recognizes the presence of heterogeneity in trades. The definition of realized volatility involves both the intraday return interval and the aggregation period one day which vary over time. The daily HAR is expressed by

$$\sigma_t = \beta_0 + \beta_D RV_{t-1,t} + \beta_W RV_{t-5,t} + \beta_M RV_{t-22,t} + \varepsilon_{t,t+1} \quad (2.5.1)$$

Where $RV_{t-1,t}$, $RV_{t-5,t}$, $RV_{t-22,t}$ represents to the 1 day, 5 days and 22 days of the realized volatility in a time period which can be viewed as "one trading week" and "one trading month" refer to the average realized volatility of 5 days lagged and 22 days lagged. $\beta_0, \beta_D, \beta_W, \beta_M$ can be estimated with the application of an Ordinary Least Squares (OLS) estimation. The HAR model believes that the latent realized volatility can be observed over time horizons longer than one day. It creates an AR regression of the 1 day, 5 days and 22 days average realized volatility to make forecasting. Xu et al. (2024) and Fan et al. (2023) mentioned that the HAR-RV model offers a more comprehensive and realistic representation of market dynamics by combining realized volatility at different time scales.

2.6 Machine learning approach

The machine learning methods described below refer to the traditional and most widely used algorithm not only in financial literature but other areas as well. For instance, Winter (2019) use the machine learning algorithm to analysis the data of human health. There are some widely used algorithms of machine learning during the past two decades.

K-Nearest Neighbors (KNN) methods

The K-Nearest Neighbors (KNN) is a non-parametric method proposed by Cover & Hart (1967). It is one of the most common and straightforward methods in machine learning methods. The KNN concept aims to make it a good tool for classification in different applications. Particularly, it can be used as a local nonlinear model for regression as well.

In the case of regression, the method allows a simple model to be fitted to the neighborhood of the point to be predicted. The neighborhood of a point in KNN model is defined by taking the k values having the lowest values for a chosen distance notation (usually Euclidean distance) defined on the space of the input vector. Similarly, the nearest neighbors of a test point are selected by looking for the k smallest distances between the test point and the training points. Then the prediction for an unknown input vector x^* is computed as follows:

$$y(x^*) = \frac{1}{k}\sum_{i \in KNN} y(x_i) \qquad (2.6.1)$$

Here, $y(x_i)$ is the output vector based on the ith nearest neighbor of the input vector in the sample. The choice of the optimal number of neighbors *k* will be performed through automatic leave-one-out selection. In a simple word, if the machine receives a sample, the algorithm first searches for its K nearest neighbors in the feature space depending on the feature vectors and defined distance. In this case, every data point is represented in the form (*x, y*) where *x* represents the vector of input values and *y* the corresponding output vector which will usually defined as a forecasting series when doing regression. What is worth to note, the amount of training set data mainly affects the accuracy of the KNN which means that the more historical data fed to the machine, the more accurate result will be get.

Artificial Neural Network (ANN) method

In machine learning, an artificial neural network (ANN) is a network of interconnected elements, which are called neurons. Neural networks are non-linear and non-parametric models that have their roots in biology which is stated by the work of Hornik et al. (1989). The neurons are used to estimate functions based on the inputs. The neurons are connected by joint mechanism which is consisted of a set of assigned weights. The back propagation training (BP) algorithm by Rumelhart et al. (1986) is usually used to minimize the quadratic error by descent maximum gradient. Therefore, the ANN method can be called back propagation neural network (BPNN). The method can be described as follows:

$$\mu_p = \sum_{i=1}^{n} \omega_{p_i} x_i \quad (2.6.2)$$

$$y_p = \varphi\left(u_p + b_p\right) \quad (2.6.3)$$

$\{x_i\}$ is the input data and $\{\omega_{p_i}\}$ describes the connection weights of neurons. u_p is the input combiner, b_p is the bias, $\varphi(*)$ is the activation function and y_p is the output of the neuron. In ANN works, multi-layer feed forward (MLP) is a common approach which has three layers: input layer, output layer, and hidden layer. Neuron takes the values of inputs parameters, sums them up with the assigned weights, and adds a bias. With the application of transfer function, the outputs which are the forecasts of volatility will be displayed.

$$\sigma_{t+h}^2 = \varphi_0^h\left(\underbrace{b_0 + \sum_{i=1}^{m}\omega_{i0}x_{t-i}^2}_{Liner\ AR} + \underbrace{\sum_{j=1}^{H}\omega_{j0}\cdot\varphi_h\left(\sum_{i=1}^{m}\omega_{ij}x_{t-i}^2 + b_j\right)}_{Non-linnear\ Componet}\right) \qquad (2.6.4)$$

It describes the structure of the model for a single forecasting horizon, where the input x_{t-i} can be a matrix of the volatility generated by the GARCH type models and other explanatory variables. The model can be separated into a linear autoregressive component of order and a nonlinear component whose structure depends on the number of hidden nodes which is the hidden layer in ANN.

Support Vector Regression (SVR) methods

Support Vector Regression is a regression methodology, based on the Support Vector Machine (SVM). The key idea behind SVR is that the regression model can be expressed using a subset of the input training samples called the support vectors. In SVM, the methods can be specified as:

$$y_k|\omega^T\phi(x_k) + b| \geq 1 - \xi_k \qquad (2.6.5)$$

$$\min_{a,b} \frac{1}{2} \|\omega\|^2 + C \sum_{k=1}^{N} \xi_k \qquad (2.6.6)$$

where $\{x_k\}$ is the variable that prepared to be predicted called predictor variable, ω is the weight vector and $\{y_k\}$ is the sample classification. The variable $\xi_k \geq 0$ is defined as a tolerance margin in the classification to make the classifier more flexible in an accepting error and the formula (2.6.5) is called a hyperplane condition. Therefore, the SVM methods is to find an optimal solution of formula (2.6.6) under the hyperplane condition where C describes the edge of the hyperplane space. When forecasting volatility, the target is not classification as SVM, but the real value series. Therefore, a regression model needs to be found when SVR is applied not the hyperplane in SVM. The SVR seeks for the linear regression function with an optimal function specified as:

$$f(x, \omega) = \omega^T x + b \qquad (2.6.7)$$

$$|y - f(x, \omega)|_\varepsilon = \begin{cases} 0 & if\ |y - f(x, \omega)| < \varepsilon \\ |y - f(x, \omega)| - \varepsilon & otherwise \end{cases} \qquad (2.6.8)$$

$$R = \frac{1}{2} \|\omega\|^2 + c(\sum_{i=1}^{N} |y_i - f(x_i, \omega)|_\varepsilon) \qquad (2.6.9)$$

During seeking for the linear regression function, a threshold error ε is introduced to be minimized in the expression. The SVR methods like SVM aim to find a minimum value of ε and $\|\omega\|^2$ under the ε -insensitivity loss error function. (formula 2.6.8). A tolerance variable is introduced as well in SVR process, which can be written as

$$R = \frac{1}{2}\|\omega\|^2 + \mathrm{c}\left(\sum_{i=1}^{1} \xi_i + \xi_i^*\right) \tag{2.6.10}$$

$$(\omega^T x_i + b) - y_i \le \varepsilon + \xi_i \tag{2.6.11}$$

$$y_i - (\omega^T x_i + b) \le \varepsilon + \xi_i^* \tag{2.6.12}$$

where ξ is used to describe the excess of error ε and ξ^* comes to be a constraint which will limit the value to the regression target. With the help of tolerance variables, a condition is formulated by (2.6.10) and (2.6.11). Thus, the SVR is seeking for a minimum value of $\|\omega\|^2$ under the conditions with the tolerance variables ξ and ξ^* more than zero.

The above optimization problem can be converted into a more formal term as:

$$y = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) k(x, x_i) \tag{2.6.13}$$

where n is the number of the support vector. In more formal terms, the model is a linear combination over all the n support vector in a kernel function $k(\cdot\ ,\ \cdot)$. It takes the inputs of the data point x which needs to be forecasted and creates the support vector x_i.

$$k(x, x_i) = \phi(x)^T \phi(x) \tag{2.6.14}$$

$$0 \le \alpha_i \le \mathrm{C}; 0 \le {\alpha_i}^* \le \mathrm{C} \tag{2.6.15}$$

where C is the kernel parameters. The coefficients $\alpha_i, \alpha_i{}^*$ are determined through taking the minimum value of an empirical risk function, solved as a continuous optimization problem. The kernel function could be the linear, radial and polynomial functions.

Liner: $k(x, x_i) = x^T x_i$ (2.6.16)

Radial: $ln[k(x, x_i)] = \frac{\|x - x_i\|^2}{2y^2}$ (2.6.17)

Polynomial: $k(x, x_i) = (x^T x_i + 1)^d$ (2.6.18)

Deep learning

Deep learning is recently introduced and applied in some of finance literature. Several literatures mainly focus on three types of approaches: Recursive Neural Networks (RNN), Long Short Term Memory (LSTM) and Convolutional neural networks (CNN). Although they are all the extension of normal neural networks, they have different characters when capture data dynamics.

Recursive Neural Networks (RNN) is a class of neural network but deeper than normal neural networks. RNN can use their internal memory to process arbitrary sequence of inputs. The units which can be calculated as a time varying real valued activation and modifiable weight and will form a circle with connect to the networks. RNNs are created by applying the same set of weights recursively over a graph-like structure. Their hidden units can be expressed as:

$$h^t = f(h^{t-1}, x^t; \theta) \quad (2.6.19)$$

In the case of RNN, the learned model always has the same input size, because it is specified in terms of transition from one state to another by using the same transition function with the same parameters at each step. A special extension of RNN called Long Short Term Memory (LSTM) is proposed by Hochreiter & Schmidhuber (1997) which replaces the hidden layers with LSTM cells. The cells are composed of various gates including input gate, cell state, forget gate, and output gate that can control the input flow. A sigmoid layer is constructed to describe how much of each component should be let through by generating a series of numbers between zero and one. In addition, a tanh layer vector is generated and will be added to the cell state to help the cell state to be updated based on the output gates by point wise multiplication operation σ. Mathematically, it can be specified as:

The input gate which consists of the input vector x_i

$$i_t = \sigma\left(W_{i(h_{t-1}, x_i)} + b_i\right) \quad (2.6.20)$$

The cell gate which constructs the entire network and the information can be added or removed information by the gates vector

$$c_t = tanh\left(W_{c(h_{t-1}, x_i)} + b_c\right) \quad (2.6.21)$$

The forgot gate vector which decides what kind of the information to be allowed

$$f_t = \sigma\left(W_{f(h_{t-1}, x_i)} + b_c\right) \tag{2.6.22}$$

The output gate vector

$$o_t = \sigma\left(W_{o(h_{t-1}, x_i)} + b_o\right) \tag{2.6.23}$$

The output vector:

$$h_t = o_t * tanh(c_t) \tag{2.6.24}$$

Convolutional neural network (CNN) is another kind of neural network for processing data that has a known topological pattern. The network employs a mathematical operation on processing data called convolution which is a special kind of linear operation instead of general matrix multiplication in at least one of their layers. There is a difference between RNN/ LSTM and CNN. The RNN/LSTM consider long term dependencies which is the long memory facts exists in time series data and uses them for future forecasting while CNN focuses on the given input sequence and does not use any previous history or information during the learning process prediction.

2.7 State of literature

Volatility is a hot issue in economic or financial research. It is directly related to market uncertainty or shocks, and it will influence the behavior of investors or companies. The trend of modelling volatility has remained popular over the past years and a huge number of papers and research could be found. Researchers did lots of work on volatility prediction using different models on various assets in different markets. This section will reviewed the literature about the volatility forecasting from early stage to recent popular methods. Different works in last two decades are reviewed and mentioned in this section.

2.7.1 Early stage of volatility forecasting

With application of different methods on the stock returns and volatility, an empirical fact had been found that the correlation among volatility is stronger than that among returns which is confirmed by Diebold (1998). It means that the financial time series has complex structure, especially in volatility. Several related works and research were carried out to forecast or measure the volatility. In the early time, Taylor (1987) studied the use of high, low, and closing prices to forecast one to twenty days DM/US (exchange rate) futures volatility and found a weighted average composite forecast to perform best. In the next ten years, more research appeared to model the volatility. Kroner (1996) gave an explanation on how to create a volatility forecast and how to use volatility to measure or observe the shock of the stock returns. Diebold (1998) found that the volatility had a complex structure when examining the exchange rate. He mentioned that "*Forecast estimates will differ depending on the current level of volatility, volatility structure and the forecast horizon*". In the long history of the volatility measurement, the main topic almost focused on "*If the volatility can be forecasted, which method will give the best performance?*" Since several models were developed to calculate or forecast the volatility, the topic became to find a "best" performance model.

2.7.2 GARCH models vs simple models

There are lots of methods to measure the volatilities, some simple models which is traditional and widely used in the past were proposed to capture the dynamics of the volatility such as Random Walk, Historical Average, Moving average, Exponential Smoothing (ES), Exponential Weighted Moving average or RiskMetrics. When considering about the financial time series, the assumption that the volatility stay constant over the time is not suitable for its structure since many stylized facts mentioned above have indicated that the volatility has some more empirical regularity. Among them, volatility clustering is a typical phenomenon. The clustering can be explained by "large changes tend to be followed by large changes of either sign, small changes followed by small changes", which is caused by the non-normality of the returns. In this case, some simple models like Random Walk which forecasts the volatility by taking the value of the last time period or historical Average models by taking the average value of all the past period volatility are not appropriate enough to identify the clustering.

Therefore, the ARCH models by Engle (1982) or its generalization by Bollerslev (1986) were proposed to capture the dynamics of the volatility. With the application of ARCH/GARCH models, Akgiray (1989) tested the predictive power of GARCH and argued that the forecasts based on the GARCH model are found to be superior rather than the Historical Average and EWMA methods. However, a wider comparison work finished by Brailsford & Faff (1996) among Random Walk model, Historical Mean model, Moving Average model, Exponential Smoothing model, Exponentially Weighted Moving Average (EWMA) model, a simple regression model, two standard GARCH models and two asymmetric GJR-GARCH models suggested that no single model is clearly superior since the rankings of the various model forecasts are sensitive to the choice of error statistic.

More recently, Balaban et al. (2002) evaluated the out-of-sample forecasting accuracy of eleven models for weekly and monthly volatility in fourteen stock markets. A comparison of several models including Random Walk model, Historical Mean model, Moving Average models, Weighted Moving Average models, Exponentially Weighted Moving Average models, Exponential Smoothing model, a regression model, an ARCH model, a GARCH model, a GJR-GARCH model, and an EGARCH model were given out. A result was obtained that the Exponential Smoothing model provides superior forecasts of volatility and ARCH-based models generally prove to be the worst forecasting models.

A later work by examining the performance of GARCH type models, moving average and Exponential Smoothing (ES) in daily exchange rate of 17 countries carried out by (McMillan & Speight (2004) stated that the simple GARCH outperforms the Exponential Smoothing and Moving Average model for the majority of currencies by using cumulative squared returns based on 30-minute intra-day observations as the measure of 'true volatility' .

By the work of Ayele et al. (2017), they used the EWMA method and several GARCH models to forecast the volatility of gold and suggested that the GARCH models with explanatory variables are superior for volatility forecasting after the comparison. However, by the work of Arı (2022), after the comparison among GARCH model, Conditional Autoregressive Range (CARR) and EWMA model, the GARCH model acted the worst model.

In recent years, the ARCH/GARCH models were selected as the most popular models, while other simple models or regression models mentioned above proved to be useful approaches to measure the volatility as well. Although the development of the GARCH is more likely to capture more empirical facts of the volatility, the existing literature still contains different evidence which supports different methods regarding to the accuracy of stock market volatility forecasts. Evidence can be found supporting the superiority of GARCH class models, while there is also evidence supporting the superiority of more simple alternatives models and questioned the superiority of the GARCH model. Since 2003, there is a review paper carried out by Poon & Granger (2003) which collected a large variety of volatility models. They reviewed about 93 papers regarding to the volatility forecasting in finance literature and suggested that number of literatures in favor of simple volatility model are roughly equal to the number of GARCH. There is no general conclusion which models were the best.

2.7.3 Standard GARCH vs alternative GARCH

The ARCH/GARCH models were considered to be one of the most widely used approaches to measure volatility and many extensions of them proved to be useful as well. Due to the complex structure of volatility, varieties of models are proposed as extensions of GARCH model in order to capture more empirical facts of volatility like the asymmetric effects of positive and negative shocks or long memory effects. The asymmetric models such as the EGARCH by Nelson (1991) and other asymmetric models like TGARCH or GJRGARCH by (Glosten et al. (1993), Quadratic GARCH (QGARCH) model by Sentana (1995) and the long memory models like Component GARCH (CGARCH) by Engle & Lee, (1999) were proposed. Further, another model called Asymmetric Power (A-PARCH) was proposed by Ding et al. (1993) which allows the power of the heteroskedasticity equation to be estimated from the data.

Since there is no mathematical evidence to show which model has the “best” performance of volatility forecasting, lots of empirical studies are carried out. Franses & Van Dijk (1996) studied the performance of the GARCH model and two of its modification QGARCH and GJRGARCH model, the result gives evidence that the QGARCH is best performance model when the estimation sample did not contain extreme observations such as the 1987 stock market crash. Miron & Tudor (2010) examined daily stock return volatility based on U.S. and Romanian daily stock return data and reported that volatility estimates given by the EGARCH model exhibit generally lower forecast errors and are therefore more accurate than the estimates given by the other asymmetric GARCH models like A-PARCH and TGARCH.

With the work of Nugroho et al. (2019) a variety of GARCH type models including GARCH, GARCH-M, GJR-GARCH, and log-GARCH were investigated. They used two datasets, one is the simulation data by random, the other is the empirical data from stock index. The results showed that the GARCH model is superior to other models when using the simulation data. The GJR-GARCH model was suggested as the best performer in the empirical exercise.

However, a forecasting performance evaluation carried out by Gabriel (2012) states that the TGARCH model is the most successful when forecasting the volatility based on Romania BET index. More recently, Dixit & Agrawal (2019) examined the performance of GARCH, EGARCH, P-GRACH in Bombay Stock Exchange (BSE) and National Stock Exchange (NSE) markets and suggested that the P-GARCH model is most suitable to predict and forecast the stock market volatility. Later by their work in Dixit et al. (2022), after a further analysis with the GARCH family models, they suggested that E-GARCH was the most acceptable model for the purpose of predicting and forecasting the market volatility.

With the work of Sharma et al. (2021), they used the standard GARCH, TGARCH and EGARCH to investigate the forecasting performance of in 5 major emerging countries and the findings indicated that the standard GARCH model are the best performer.

In general, it is difficult to define which type of GARCH is the best performance model. The result was related to the data they were fit since the data has lots of complex structure which might be well captured by different type of GARCH models. There is very little literature discussed if one method is significantly better than another since each research uses different data sets, different time intervals and a variety of evaluation techniques.

2.7.4 GARCH model vs other traditional methods

With the hard identification of a best model on volatility forecasting, the recent literature moved their attention on how to improve the estimation of the parameters in an existing model or how to improve their forecast ability. With the application of GARCH type models, most researchers tended to use GARCH (1,1) model to avoid using the high order one since the high order of GARCH will increase the complexity and difficulty of calculation.

There was some evidence that a small lag such as GARCH (1,1) was sufficient to model the variance changing over long sample periods confirming by the work of Franses & Van Dijk (1996). In financial risk management a method called Value at Risk (VaR) model which is one of the most used models which measure the amount one could lose. Many banks and other financial institutions use the concept to measure the risks faced by their portfolios. The 1% value at risk is defined as the amount of cash that one can be 99% certain exceeds any losses for the next day. Statisticians call this as 1% quantile, because 1% of the outcomes are worse than the rest of 99 %. Oskay et al. (2017) carried out a VaR model based on GARCH family and found the obtained volatility estimates in a VaR-GARCH model should eventually increase the forecasting ability beyond a traditional VaR method.

Echaust & Just (2020) combined the Extreme Value Theory with GARCH model to build a new GARCH-EVT model to estimate out-of-sample Value at Risk (VaR) forecasts. However, the GARCH-EVT model produces similar Value at Risk estimates and there is no improvement of VaR accuracy being observed.

Although methods like GARCH and other simple regression models occupied most of area in measuring the volatility in finance literature, some weaknesses were still discovered. The conditional variance is latent and slowly decreasing, and hence is not directly observable. It can only be estimated among these approaches which will lead to a failure to describe in an adequate manner. Some research by Andersen & Bollerslev (1998); Andersen et al. (2001) suggested that this relative failure of GARCH models arises not from a failure of the model but a failure to specify correctly the 'true volatility' measure against which forecasting performance is measured. They tried to avoid the failure of GARCH by measuring the daily foreign exchange volatility by aggregating 288 squared five-minutes returns. A method of calculating the high frequency data is proposed by Merton (1980): *the variance over a fixed interval can be estimated arbitrarily, although accurately, as the sum of squared realizations, provided the data are available at a sufficiently high sampling frequency.* Moreover, by the work of McMillan & Speight (2004), they suggested that the prediction performance of GARCH model will be improved by using the high frequency intraday data confirmed. The method of calculating the volatility by high frequency data can be introduced as the realized volatility. The realized volatility essentially becomes "observable", it can be modeled directly, rather than being treated as a latent variable. More and more papers used this new volatility measure to evaluate the out-of-sample forecasting performance of GARCH models such as Hansen & Lunde (2005), and Patton (2011).

With the appearance of realized volatility, several comparisons were made by Andersen (2008) including Quadratic Return Variation, Jumps and Bi-power Variation, Conditional Return Variance Efficient Sampling and Microstructure Noise. This comparison gave evidence that the RV measures facilitate direct estimation of parametric models. It gave empirical content to the latent variance variable and was therefore useful for specification testing of the restrictions imposed on volatility by parametric models previously estimated with low frequency data. The HAR-RV model proposed by Corsi (2004) paved a good approach to measure and forecast the realized volatility. It is a simple AR-type model of the realized volatility with the feature of considering different volatility components realized over different time horizons and thus termed as Heterogeneous Autoregressive model of Realized Volatility (HAR-RV). The HAR-RV model successfully captures the main empirical features of financial returns (long memory, fat tails, and self-similarity) in a very tractable and parsimonious way and it has a good forecasting performance. A recent work carried out by Mastro (2014) confirmed that the HAR-RV model outperforms ARCH/GARCH and EGARCH. Huang et al. (2016) combined the HAR model and traditional GARCH model and developed a parsimonious variant of the Realized GARCH model in order to capture more dynamics of volatility. The finding of their work showed that the new model specification better captures the long memory dynamics of volatility.

Bergsli et al. (2022) investigated the forecasting performance of volatility with the application of several GARCH and two heterogeneous autoregressive (HAR) models. By a wide comparison among GARCH series and HAR model, they found that EGARCH and APARCH were superior to other type of GARCH models in GARCH family. HAR models based on realized variance perform better than GARCH models when using daily data, which showed the superiority of HAR models over GARCH models when using short-term/high frequency data.

More recently, by the work of Zahid et al. (2022), they used a model built with HAR model and GARCH model which is introduced as HAR-GARCH model. They further extended the model by incorporating jumps and continuous components. This new extension model provided better forecasting accuracy for Bitcoin volatility as compared to other realized volatility models in their empirical exercise.

Chen et al. (2023) investigated three realized GARCH series models including the realized exponential GARCH (EGARCH), realized heterogeneous autoregressive GARCH (HAR-GARCH), and realized threshold GARCH (TGARCH) models. After considering the out-of-sample result of 5 years by S&P500 stock index, it showed that the realized EGARCH model performs best with regard to volatility forecasts.

2.7.5 GARCH vs Machine Learning

In recent years, some other methods from machine learning concept were being used in forecasting the volatility as well. Although the concept of machine learning methods was proposed in early 60s, their ability to improve the forecast of volatility along with other models were developed in recent years. In most of finance literature regarding with the machine learning methods, three main methods including Artificial Neural Networks (ANN) proposed by Warren McCulloch and Walter Pitts in 1943, Support Vector Machine (SVM) by Vladimir in 1963 and Deep Learning (DL) by Alexey in 1960s were used widely. The main reason of using machine learning methods was their flexible abilities to approximate any nonlinear functions arbitrarily without priori assumptions on data distribution confirmed by S. Haykin (1999). Several literatures about machine learning methods of forecasting volatility tended to combine the machine learning methods with GARCH to make more accurate or reliable forecasts, especially with the application of ANN hybrid GARCH model. In order to enhance the forecasting performance of GARCH family model, Neural Networks (NNs) were brought into this field. The reason is that NNs has the functional flexibility to capture the nonlinear relationship between past return innovations and future volatility. Donaldson & Kamstra (1997) introduced a new nonlinear semi-nonparametric model for conditional stock volatility. In their work, a nonlinear GARCH model based on the Artificial neural network was constructed. The performance of the new model is compared both in-the-sample and out-of-sample with other popular volatility models in four international stock market indices. They state that the new GARCH model based on ANN performs better than the other GARCH approaches. Later by the work of Hu & Tsoukalas (1999), they combined four conditional models including Moving Average, GARCH, EGARCH and IGARCH with a simple averaging model, an ordinary least squares model, and an artificial neural network. The conclusion drawn by them gave evidence that the ANN combined model performed better forecast in the crisis period, and it proved superior to linear combining models like simple averaging and ordinary least squares models.

More recently, Hyup Roh (2007) proposed the hybrid model between the ANN and financial time series models to forecast the volatility of stock price index. He addressed that ANN-time series models can enhance the predictive power for the perspective of deviation and direction accuracy witch a same result is confirmed by Bildirici & Ersin (2009). He discussed the ARCH/GARCH family models and enhanced them with artificial neural networks to evaluate the volatility of daily returns of Istanbul Stock Exchange. The ANN-extended versions of the obtained GARCH models improved forecast results by capturing the strong volatility clustering, asymmetric effect, and nonlinearity facts better. Some more details about the improvement of GARCH by the NNs method have been discussed in the past few years. Kristjanpoller et al. (2014) carried out a study which demonstrated that the ANN-GARCH model improves the forecasts of the GARCH model by 30.6% for the oil spot price volatility and 29.8% for the oil futures price volatility when using 21 days as a horizon. The study found the coefficients of input variables by financial time series process and extracted new variables that greatly influence the results through analyzing stock market.

Liu & So (2020) developed a GARCH model into an artificial neural network (ANN) for financial volatility modeling. The results showed that the new model built with neural networks can outperform the standard GARCH (1,1) model with standardized Student's t distribution.

Mademlis & Dritsakis (2021) combined the neural networks with GARCH models in order to improve the volatility forecasts. The conclusions reveal that the hybrid models based on an EGARCH model provides the best predictive power.

However, there is also some investigation said that the traditional GARCH model is better than machine leaning methods. Ampountolas (2021) made a wide range comparison among Holt–Winters (HW) triple exponential smoothing model, the autoregressive integrated moving average (ARIMA) model, a seasonal autoregressive integrated moving average (SARIMAX) model with exogenous variables, the artificial neural networks model (ANNs), a sGARCH and GJR-GARCH models. He suggested that the GJR-GARCH model shows a superior predictive accuracy at all horizons.

The literature of neural networks confirm their usefulness in modeling the conditional volatility of stock returns due to their data-driven and nonparametric weak properties but one of the important weaknesses of neural networks is that they cannot avoid getting trapped in local minima, which means that the increase of variable freedom will lead the training process to a trap in some local minimum of the high dimensional functional space. More generally, the overfitting problem will exist which can be attributed to the fact that a neural network captures not only useful information contained in the given data, but also unwanted noise. This usually leads to a large generalization error or local minima mentioned above.

Unlike most of the traditional learning machines that adopt the Empirical Risk Minimization Principle, SVMs implement the Structural Risk Minimization Principle, which seeks to minimize an upper bound of the generalization error rather than minimize the training error which will lead to a better generalization than conventional techniques. Cao & Tay (2001) also confirmed the feasibility of SVM in financial time series forecasting by comparing it with a multilayer Back Propagation Neural Network (BPNN) and a result reported that SVM outperforms the BPNN. A SVM-based model proposed by Gavrishchaka & Ganguli (2003) for the volatility forecasting from the multiscale (such as volatility clustering) and high-dimensional market data can efficiently extract information from the high-dimensional inputs of lagged returns and can handle both long memory and multiscale effects of in homogeneous markets without restrictive assumptions and approximations required by other models. Later he applied the SVM-based volatility model to stock market data and concluded "*the SVM-based volatility model can be comparable and often superior to the main-stream models such as generalized ARCH and its generalizations" a*ddressed by Gavrishchaka & Ganguli (2003). Pérez-cruz et al. (2003) used the Support Vector Machines (SVMs) as a new nonparametric tool for regression estimation of a GARCH model to predict the conditional volatility of stock market returns. Theoretically, GARCH models are usually estimated using maximum likelihood (ML) procedures which assumes that the data are normally distributed, while empirical data usually has a different distribution with normal. Pérez-cruz et al. (2003) gave a method to estimate GARCH models by using SVMs and that such estimates have a higher predicting ability than those obtained via common ML methods. In more recent years, a more comprehensive comparison is carried out by Chen et al. (2009), they used four methods of volatility forecasting. First, the simple model, simple moving average. Second, the GARCH type model. Third, traditional ANN-GARCH models and fourth the SVM under the GARCH framework. Empirical results from both simulation and real data reveal that, under a recursive forecasting scheme, SVM-GARCH models significantly outperform the competing models in most situations of one-period-ahead volatility forecasting. In addition, the standard GARCH model also performs well in the case of normality and large sample size, while EGARCH model is good at forecasting volatility under the high skewed distribution.

More recently, Sun & Yu (2019) combined the SVR method and GARCH model together. Unlike SVR-GARCH model, they do not choose the way to replace the maximum likelihood estimation with the SVR estimation method when estimating the GARCH parameters. They propose two SVR-GARCH models on normal distribution and t distribution, respectively when investigating the effect of innovation. The results showed that the GARCH-(t)-SVR models and GJR-(t)-SVR models improve the volatility forecasting ability. (Papadimitriou et al. (2020) used a GARCH-SVM approach to make the out-of-sample volatility forecasting. The model they created by SVM and GARCH reached 79.17% out-of-sample forecasting accuracy. Aras (2021) proposed a new combining method based on support vector machines (SVM) and GARCH type models and suggested that the new method leads to more accurate volatility forecasts than original GARCH type models.

Another approach is the Deep Learning method. It can be seen as an extension method of neural networks. As mentioned above, the neural networks face a problem of overfitting and in fact the increase of freedom will lead the training process to a trap in some local minimum of the high dimensional functional space even when the model is an honest representation of the system by Huang (2009). To solve the problem, some procedures like providing more insight into how they operate on neural networks. A deep learning method will leverage their predictive power and systematically avoiding the problem of overfitting. In particular, there are some new regularization methods and faster training techniques such as using piecewise linear activation functions as opposed to transcendental functions, which allow for neural nets with many hidden layers to be trained easily. In a simple word, the deep learning methods can be seen as a neural network method with enough hidden layers so that the neural networks will reduce the regression error. Some latest works are carried out by using deep learning to study the characteristic of volatility. The study on Chinese Yuan's volatility in the onshore and offshore markets carried out by Lee & Chun (2016) proposed MLP-GARCH (neural networks based) and a DL-GARCH (deep learning based) by employing Artificial Neural Network to the GARCH. In his work, both models show their overall outperformance in forecasting over the GARCH.

Amirshahi & Lahmiri (2023) made a combination of Deep Learning and GARCH-type models and tested the forecasting ability in cryptocurrencies market. Their finding revealed that the deep learning methods will improve the forecasts of GARCH-type models with any distribution assumption. Furthermore, when treating the forecasts of GARCH-type models as informative features, it can also increase the predictive power of the studied deep learning models as well.

Michańków et al. (2023) developed a new approach by combining the common GARCH models with deep learning methods. The GARCH type models are GARCH, EGARCH, GJR-GARCH and APARCH and concluded that the hybrid solutions produce more accurate point volatility forecasts.

Ni & Xu (2021) introduces a new method which is developed on deep learning method and DCC-GARCH models in order to investigate the stock market correlation. The results indicate that the introduction of deep learning can help improve the efficacy of existing correlation forecasting methods.

2.8 Conclusion

In this thesis, we will continue the work of exploring the improvement of existing models for producing a more accurate volatility forecast by an appropriate model. As mentioned in the literature, lots of work are carried out to forecast volatility, a list of modes is introduced for empirical exercise. However, there is no general census on which model has performed the best so far. GARCH models are proposed to capture different dynamics of volatility such as clustering, leptokurtic, leverage effect and long memory. Although the hybrid GARCH model based on machine learning is pointed to be effective in enhancing the forecasting performance of volatility, there is very few literatures concerns about the input variable. Some questions still remains, like how many or what type of explanatory variable should be used to train and test the machine since each empirical result from literature are based on their own dataset and most of them are limited in one or two countries. Which model will capture the long-memory effect in most cases in the stock markets? A basic comparison will be made among the GARCH family models and the hybrid models. With the application of deep learning, theoretically, if enough hidden layers are given, the methods can capture sufficient features of the data. A further comparison will be carried out by different variables as input among these models. Finally, in this thesis, we will explore the forecasting performance of several hybrid models and build some new approaches based on machine learning methods and traditional models.

3.Emprical chapter of volatility forecasting

GARCH type models versus Hybrid GARCH based on Neural Networks & HAR models

Abstract

In this chapter, several comparisons of univariate volatility models are carried out. This chapter investigates the forecasting performance between several models including GARCH type models which have been proved to be able to capture the volatility clustering, leverage effect and long memory effect in the data, their extension based on neural networks and HAR-RV model which measure the volatility in a realized way. Twelve stock indices from different countries and five measures of comparison are applied in the exercise. The results show that the GARCH type models based on neural networks perform better overall than GARCH type models, especially the hybrid asymmetric GARCH models. However, these hybrid models have some weakness to capture the "long memory" effect exists in the data. More specifically, the CGARCH model outperforms than others to identify any "long memory" effect. The results gives an empirical result in a wide comparison. The policymakers can benefit from the results to formulate their policies to avoid risk.

3.1 Introduction

In recent years, stock markets have had a rapid development. With the rapid development, lots of events related to risk management in the stock markets happened. There are some events which are very famous. One is the stock market crash in 1929 caused by excessive leverage, the stock market suffered a huge loss and declined by more than 20%. Another is the 2008 global financial crisis happened in United States of America and then spread to the whole world, it caused lots of billion's loss and the market appeared to get down in a long time. More recent is the 2015 stock market crisis happened in China which made lots of people lose their house. Therefore, measuring the risk is one of the most important tasks to avoid loss and market disruption. It is highly confirmed in financial market that the profit and the risk is related positively with a simple word "high risk comes with a high return". Volatility, as a measurement of risk, reflects a financial situation in a certain period. Moreover, high unexpected volatility usually leads to a loss of the expected returns. Therefore, capturing the movement of volatility in stock market appears to be an important issue. Particularly, in stock market, measuring and forecasting volatility is vital for the investors or the supervisors of the market since investors tend to make profits with less risk and the supervisors like government or institutions are more likely to observe the early appearance of the market disruption through the changes of volatility. Finding a reliable way to measure and forecast volatility is important in stock markets to manage risk and make profits. Lots of studies in the finance literature focus on volatility forecasting models, but different results are generated from varieties of research when identifying the "best performance" models or methods. Followed with the key question of finding a model with a better performance to forecast the volatility, several models are proposed. In general, with application of the work by Poon & Granger (2003), most of the literature focus on the time series models or stochastic volatility models based on option prices. This chapter puts the emphasis on the time series models, especially the GARCH models proposed by

Bollerslev (1986). Although some other regression models are proved to be superior in the literature, the GARCH type models are still the most widely used. In more detail, GARCH type models, as a time series models, measures or forecast the volatility by conditional on its past variance and error term in the return process, which has the ability to capture the volatility clustering recognized by Mandelbrot (1963a). More recently, some literatures report that the volatility forecasting can be improved by a GARCH model based on machine learning methods like neural networks or a direct regression on realized volatility of different time period. Follow the step of the research, a hybrid GARCH model based on the neural networks and a HAR-RV model by Corsi (2004) will be included. This empirical chapter aims to examine the forecasts performance on univariate volatility of different models including GARCH genres, HAR model and hybrid ones and a wide and comprehensive comparison among these models are expected. Likewise, some researchers only concern with the hybrid models from a certain GARCH model with a good estimation in GARCH process without checking the performance of other hybrid models based on different GARCH process. The empirical exercise used four traditional GARCH type models, four hybrid GARCH models, and a HAR-RV model to find a model with better performance. Twelve stock indices from different countries are included in the analysis, a more comprehensive analysis of the hybrid models will be reported which will be contributed to the gaps mentioned above.

The appearance of GARCH models can be seen as the milestone of volatility forecasting history since its better fit for forecasting volatility based on time series when the data exhibits heteroskedasticity and volatility clustering. Comparing with the simple models such as Historical Average (HA), Moving Average (MA), Exponential Weighted Moving Average (EWMA), etc. in the early stage, the GARCH models calculate the conditional variance of the return series with maximum likelihood method instead of using the sample standard deviations. In this chapter, all the analysis will be considered by a GARCH (1,1) type models rather than the more sophisticated model with large lag number of past conditional variance and squared returns for the reason of time consuming. Likewise, there is no evidence that a GARCH (1,1) is outperformed by more sophisticated models which confirmed by (Hansen & Lunde, 2005).

With the development of the first generation of GARCH model, some extensions are proposed to identify more characteristics of the volatility. The "leverage effect", which an unexpected price drop increases volatility more than an analogous unexpected price increase are considered. The exponential GARCH (EGARCH) model generated by Nelson (1991) allows for an asymmetric effect ("leverage effect") between positive and negative asset returns. Another asymmetric model which has been widely used to identify the asymmetric effect is the Threshold GARCH (TGARCH) model, which is also known as the GJR GARCH model by Glosten et al. (1993). With the appearance of the "long memory" effect, which the current information remains important for the forecasts of the conditional variances for all horizon, several "long memory" models are proposed in order to capture the "long memory" effect by their property of "persistent variance". The Fractionally Integrated GARCH (FIGARCH) by Baillie et al. (1996) and Component GARCH (CGARCH) by Engle & Lee (1999) are introduced, which are able to capture the "long memory" in the horizon.

Another type of extension of the GARCH model is based on the machine learning methods, neural networks. The GARCH models mainly focus on the past conditional variance and squared returns using maximum likelihood method, while the neural networks have the ability to concerns more explanatory variables of the volatility by formulating them with an activation function including both linear and nonlinear function like Logistic, Hyperbolic Tangent, SoftMax, etc. Neural networks are considered to be one of most widely used techniques to enhance the forecasting performance of GARCH models. Compared with the traditional GARCH type models, the method does not need to satisfy lots of constraints of GARCH model, it is able to approximate any nonlinear functions arbitrarily without prior assumptions on data distribution. Technically, the method can take more information which may affect the volatility like prices, volumes, returns, etc. by feeding the machine. However, the method will generate overfitting error if too many freedoms of inputs are fed which is also called a local minimum.

This chapter reports the one-day rolling window of forecast ability of a list of volatility models including the GARCH model, the extension of the GARCH model TGARCH and EGARCH which is cataloged as the asymmetric models, the long memory model, CGARCH, some hybrid GARCH models based on neural networks, GARCH-NN, EGARCH-NN, TGARCH-NN, CGARCH-NN and a HAR model as well. Basically, the volatility generated by the GARCH type models will be used as the main input data to train the hybrid models and make a forecasting of the volatility with the application of hybrid models. Four measures of comparison including Mean Absolute Error (MAE), Mean Squared Error (MSE), Quasi-Likelihood (QLIKE) and R^2LOG loss function are used to find the a model with better performance and a wider performance report of different models on twelve stock indices are lay out. The Mincer-Zarnowiz regression test (MZ test) are used to test the regression and the Model Confidence Set (MCS) are used to select a best model.

There are three main contributions of this chapter: First, a series of hybrid-built model which combined the GARCH family and neural networks were introduced. This new approach proved to be superior to the traditional GARCH models on stock index from twelve countries. Second, the forecasting ability in GARCH family was examined in a wide range of stock indices across twelve countries, the long memory model CGARCH performed better than other GARCH series. Third, the empirical results revealed that the neural networks are more likely to capture asymmetric effects. This chapter gave a more comprehensive view of the comparison among nine volatility models including GARCH series, HAR models and hybrid models. For further research, the results can be a reference when choosing volatility models.

The structure of the chapter is organized as follows. Section 2 gives the data series, source, time periods. Section 3 describes the list of volatility forecasting models and a measure of realized volatility. Section 4 shows the method used to compare the performance of the models in the previous section. In section 5, a full report of the performance will be displayed, and section 6 will discuss the conclusions and findings.

3.2 Data

The main data set being used in this paper is the daily stock price index in twelve countries over the world. The countries which are selected are based on the principle of market capitalization in the stock market, particularly, countries with over $1500 billons market capitalization which is presented by the end of year 2017 will be included. The market capitalization represents the total value of a company's stock, the stock with a high market capitalization means the size of the stock market is large. With the application of market capitalization, it allows investors to size up a stock index. The thesis aims to select the stock index with large size to test the forecast ability of different models. Therefore, the twelve daily stock price indexes are selected.

More specifically, the selected countries can be classified in regions: Europe, America, and Asia, including both developed and emerging markets. In more details, the EEA union contains lots of countries while the market capitalization of it is calculated together, therefore, only countries with more comprehensive data to access currently are selected. In order to compare the performance of different methods, a sample which contains the realized volatility of each stock index which act as a “true volatility” in these countries are included in the data set as well. The selected countries in alphabetical orders are Australia, Canada, China, France, German, Hong Kong, India, Japan, Korea, Switzerland, United Kingdom and United States of America.

Table 3.2.1 Sample countries by region				
Region/country	Europe	America	Australia	Asia
1	France	Canada	Australia	China*
2	German	USA		HK
3	Switzerland			Japan
4	UK			Korea*
5				India*
Note: * Emerging/Developing Market				

The markets of selected countries in Europe, Australia and America are all developed market since the market in these countries have already experienced a long history of development while the emerging market are mainly located in the Asian area including 3 markets: China, Korea and India. The data comes from different type markets will give a comprehensive result that whether the performance of the methods is suitable or effective when the market changes. The stock price indices in different countries usually cannot be compared directly since the structure or trading rules are different among countries. The structure of an index is usually sorted by three criterions: market capitalization weighted index, equal weighted index, and price weighted index. In order to replicate of the performance in the market itself, the market capitalization weighted indexes are selected since it is more possible to form a much better representation of the market, which the index will represent the change of a stock by weight and are more likely to avoid bias. The index selected will differ in countries which is the most widely used and representative, Table 3.2.2 will give an overview of the index selected in these countries.

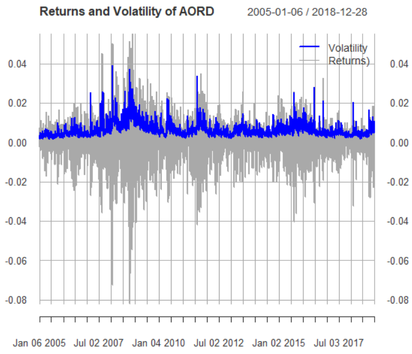
Table 3.2.2 Indices selection in sample countries

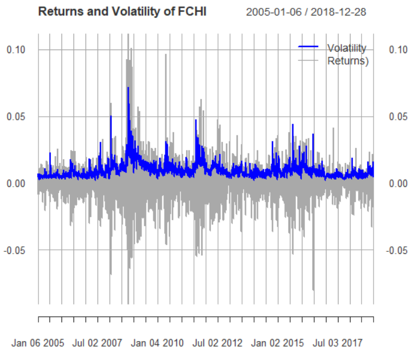
Region	Country	Stock Index
Europe	France	CAC-40
	German	DAX
	Switzerland	SMI
	United Kingdom	FTSE-100
America	USA	S&P-500
	Canada	TSX
Australia	Australia	AORD
Asia	China*	SSEC
	Hong Kong	HSI
	Japan	Nikkei-225
	Korea*	KS-11
	India	Nifti-50

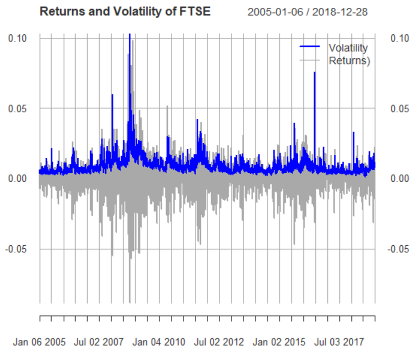
All the price data are obtained by the Application Programming Interface (API) finance data from "Yahoo Finance", while the realized volatility data is obtained by "Oxford Man". In all the countries, the adjusted daily closing price is chosen, the time period is from 5 January 2005 which is the first available trading date in year 2005 to 28 December 2018 which is the final available trading date in year 2018. In order to train the machine (neural networks) with more data, the long time period of data is selected. Since some of the daily stock price data are not comprehensive before 2005, due to the data availability, this time period was selected. The prices of the index are transformed in returns by standard methods which make it more measurable in equation (3.2.1).

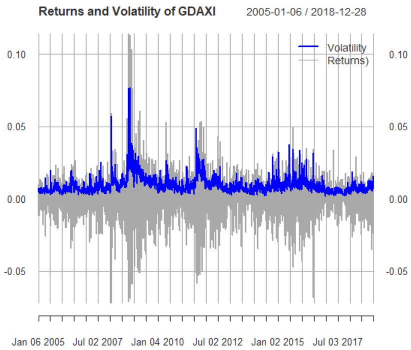
$$r_t = \frac{S_t - S_{t-1}}{S_{t-1}} * 100\% \qquad (3.2.1)$$

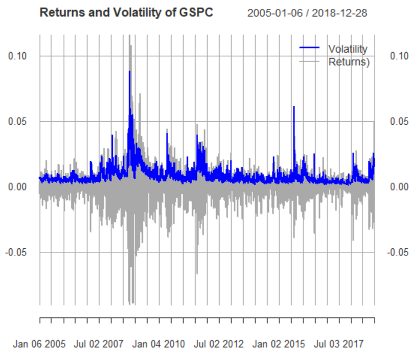
The realized volatility data is calculated by daily high-frequency intraday data (average of 5-minutes returns). The data across all the sample countries are separated into two parts which is the in-sample estimation from 5 January 2005 to 25 November 2014 (2400 observations or approximate 10 years) and out of sample forecasting from 26 November 2014 to 28 December 2018 (1000 observations or approximate 4 years). The returns and volatility can be viewed in Figure 3.2.1. It can be seen that the volatility tends to be the highest during the period of 2007 and 2008, the amplitude of return changes is great as well. This should be referred to the global financial crisis happening in 2008. Moreover, during the period of year 2015, there also a high volatility gathers together, which should be referred to the Chinese stock market crisis in 2015 since the volatility of SSEC in 2015 tends to be strong higher than other countries even higher than the volatility during the 2008 global financial crisis period.

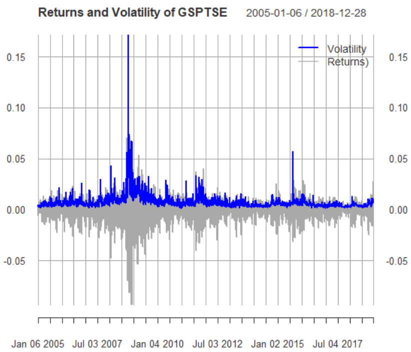

Returns and Volatility of AORD
2005-01-06 / 2018-12-28
Volatility
Returns)
0.04
0.02
0.00
-0.02
-0.04
-0.06
-0.08
Jan 06 2005
Jul 02 2007
Jan 04 2010
Jul 02 2012
Jan 02 2015
Jul 03 2017


Returns and Volatility of FCHI
2005-01-06 / 2018-12-28
Volatility
Returns)
0.10
0.05
0.00
-0.05
Jan 06 2005
Jul 02 2007
Jan 04 2010
Jul 02 2012
Jan 02 2015
Jul 03 2017


Returns and Volatility of FTSE
2005-01-06 / 2018-12-28
Volatility
Returns)
0.10
0.05
0.00
-0.05
Jan 06 2005
Jul 02 2007
Jan 04 2010
Jul 02 2012
Jan 02 2015
Jul 03 2017


Returns and Volatility of GDAXI
2005-01-06 / 2018-12-28
Volatility
Returns)
0.10
0.05
0.00
-0.05
Jan 06 2005
Jul 02 2007
Jan 04 2010
Jul 02 2012
Jan 02 2015
Jul 03 2017


Returns and Volatility of GSPC
2005-01-06 / 2018-12-28
Volatility
Returns)
0.10
0.05
0.00
-0.05
Jan 06 2005
Jul 02 2007
Jan 04 2010
Jul 02 2012
Jan 02 2015
Jul 03 2017


Returns and Volatility of GSPTSE
2005-01-06 / 2018-12-28
Volatility
Returns)
0.15
0.10
0.05
0.00
-0.05
Jan 06 2005
Jul 03 2007
Jan 04 2010
Jul 03 2012
Jan 02 2015
Jul 04 2017

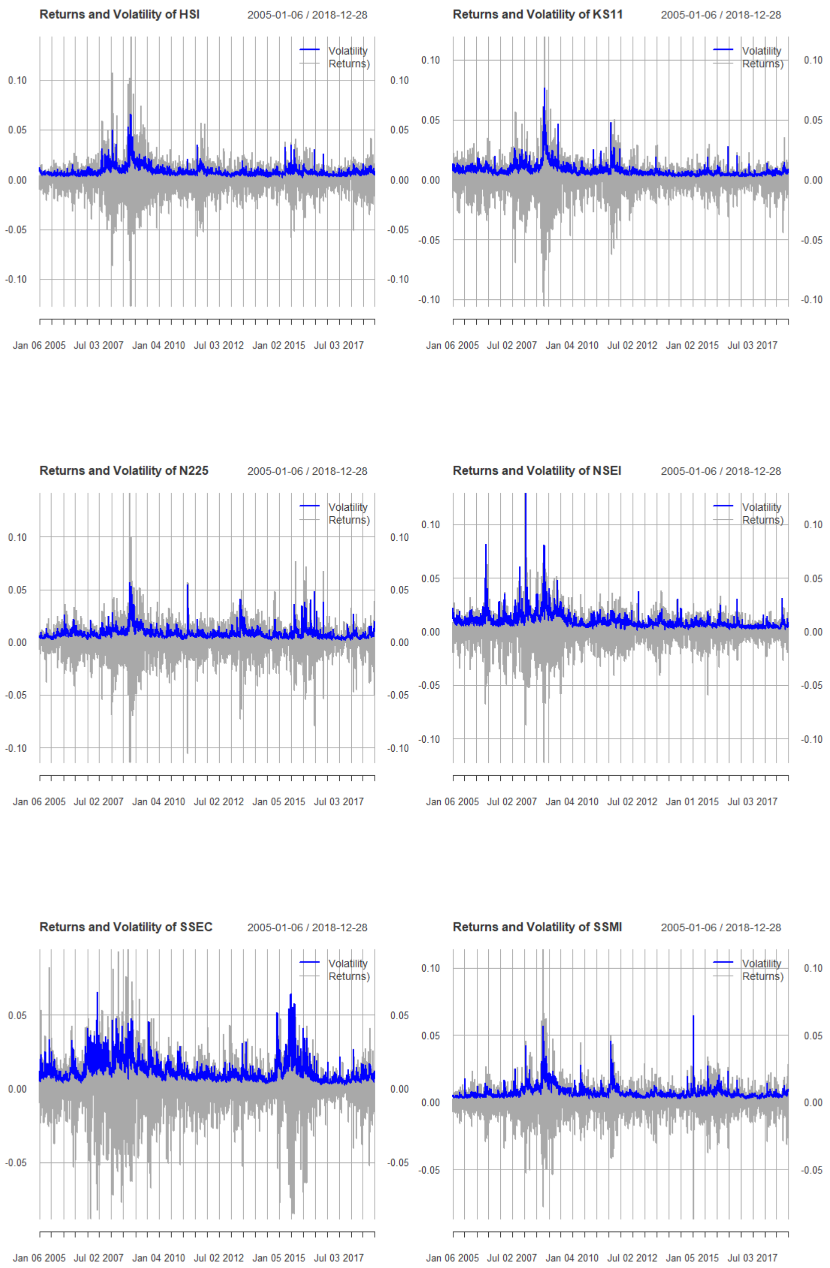


Figure 3.2.1 Returns and realized volatility of the sample index

Some descriptive statistics of the returns are reported in Table 3.2.3. The price of the indices is converted into a return series that jumps around zero which can be observed that the mean of the returns is close to zero. It is not surprising to find that the standard deviation in emerging market is higher rather than in developed countries, particularly, China market appears to be the most unstable while Australia market is the most stable. The skewness and kurtosis of all the return series are calculated and a Jarque-Bera (JB) test is carried out to discover the normality of the return series. The statistics of the JB test give evidence that there exists significant difference in skewness and kurtosis compared with the normal distribution, which means the normality is rejected for all series.

Table 3.2.3 Statistics of returns in all sample countries

	Mean	Maximum	Minimum	Median	St.Dev	Kurtosis	Skewness	J.B test
Australia	0.00015	0.055064	-0.08198	0.000546	0.0102	5.033131	-0.42902	3837.521
Canada	0.000188	0.098233	-0.09324	0.000719	0.010746	11.43234	-0.48617	19258.39
China*	0.000463	0.094551	-0.08841	0.000925	0.016938	4.067061	-0.40052	2197.946
France	0.000148	0.111762	-0.09037	0.000337	0.013602	7.35191	0.147217	8061.94
German	0.000344	0.11402	-0.07164	0.00089	0.0133	6.96382	0.118507	7179.474
HK	0.000289	0.143471	-0.127	0.000679	0.014849	10.15159	0.25123	14807.36
India*	0.00042	0.177441	-0.12203	0.000496	0.01418	13.98701	0.435273	22544.43
Japan	0.000277	0.141503	-0.11406	0.000594	0.015031	8.096362	-0.31853	9420.848
Korea*	0.000318	0.119457	-0.10571	0.000597	0.012291	9.12867	-0.38028	12048.23
Switzerland	0.000169	0.11391	-0.08671	0.000518	0.010913	9.133535	-0.06709	12279.51
UK	0.000161	0.098387	-0.08848	0.000416	0.011374	8.602499	0.029216	10851.19
USA	0.000281	0.1158	-0.09035	0.000634	0.011812	11.93823	-0.12037	20905.68

Note: * Emerging/Developing Market; St.Dev: Standard deviation

3.3 Methodology

This chapter will present a selection of models including GARCH type models, HAR models and hybrid GARCH models based on neural networks. The different GARCH model is built in different structures or features in order to capture the characteristic of volatility empirically like clustering, leverage effect, long memory, while the HAR model is a new approach to measure the realized volatility directly. The hybrid GARCH models based on neural networks are considered to be some extensions of the GARCH genre which aim to improve the forecasting performance. The chapter will focus on the forecasting performance in different methods and the possibility whether the hybrid GARCH will give a better empirical result in selected countries. Four GARCH type models are considered including the symmetric GARCH: standard GARCH model, the asymmetric GARCH: TGARCH by Glosten et al. (1993) and EGARCH by Nelson (1991) models, the long-memory models: CGARCH by Engle & Lee (1999). Four hybrid GARCH based on machine learning are considered, particularly, neural networks in this chapter: the GARCH-NN, TGARCH-NN, EGARCH-NN, CGARCH-NN. The HAR-RV model is used as well to capture the realized volatility more directly. Recent research of the comparison between GARCH and simple models are very comprehensive, however, different views on application of the hybrid ones still exists, an explanation of the hybrid models will be included in this chapter.

In order to document the model more clearly by mathematics, some definition and assumption should be made. The return series X_t is generates by below

where μ is the mean process which could be an AR process, MA process or ARMA process, ε_t is a disturbance which can be written as:

$$X_t = \mu + \varepsilon_t \tag{3.3.1}$$

$$\varepsilon_t = \sigma_t \upsilon_t \tag{3.3.2}$$

$\{\nu_t\}$ is a sequence of independent and identically distributed (*i.i.d*) random variables with a constant mean zero and variance.$\{\sigma_t\}$ is non-negative stochastic process for a fixed *t* which is identified as the volatility process with a forecast value presented by $\hat{\sigma}_t$ in empirical exercise.

The sample data is separated into an in-sample period and out-of-sample period. The "actual volatility", specifically, which the volatility forecasts will be compared to is defined as the realized volatility calculated by 5minutes high frequency intraday data. It can be retrieved from the dataset directly as a time series.

Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models

The GARCH model developed by Bollerslev (1986) is a generalization of the Autoregressive Conditional Heteroscedasticity (ARCH) model. In GARCH model, the disturbance error term ε_t is under an assumption that it is distributed by zero mean and the conditional variance σ_t will change with the time. The conditional variance equation in the simplest case which is the GARCH (1,1) can be written as the form:

$$\hat{\sigma}_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{3.3.3}$$

$$\varepsilon_t = \sigma_t \upsilon_t \tag{3.3.4}$$

The conditional variance is represented as σ_t^2 and it calculated the variance based on the history estimates and the past disturbance error term from the mean model. More general, the GARCH (p, q) can be retrieved by derive the formula (3.3.5)

$$\sigma_t^2 = \omega + \alpha(L)\varepsilon_{t-1}^2 + \beta(L)\sigma_{t-1}^2 \tag{3.3.5}$$

$$\alpha(L) = \alpha_1 L + \alpha_2 L^2 + \cdots + \alpha_q L^q;\ \beta(L) = \beta_1 L + \beta_2 L^2 + \cdots + \beta_q L^p \tag{3.3.6}$$

Since the conditional variance σ_t^2 is clearly positive, the parameter on the right hand of the function should satisfy the "nonnegativity constraints", which means $\omega > 0, \alpha \geq 0, \beta \geq 0$. The main idea is that σ_t^2, the conditional variance of ε_t^2 given information available up to time k–1, has an autoregressive structure and is positively correlated to its own recent past and to recent values of the squared returns ε_t^2 are likely to be followed by large (small) values. The GARCH models permit a wider range of behaviors which allows the conditional variance to be dependent upon previous own lags so that the huge values of the lags (q) will not influence the accuracy of the prediction.

Exponential GARCH (EGARCH) models

The GARCH model, mathematically, explains conditional variance by formulating a linear regression between the squared disturbance error term in return process and the past variance. There is a limitation of this method that an equal size of positive and negative news will have a same impact on the conditional variance. In empirical research, an asymmetric effect usually exists which was explained by Black & Scholes (1973). A negative shock will increase more with the conditional variance than a same magnitude of positive shock, which is called a "leverage/asymmetric effect". A widely used asymmetric GARCH model put forward by Nelson (1991) namely the exponential GARCH (EGARCH) model provides a first explanation for the σ_t^2 depends on both the size and the sign of lagged residuals in the return process. In particular,

$$ln(\hat{\sigma}_t^2) = \omega + \sum_{i=1}^{p} \beta_i ln(\sigma_{t-i}^2) + \sum_{i=1}^{q} \alpha_i \{\phi \upsilon_{t-1} + \psi[|\upsilon_{t-i}| - E|\upsilon_{t-i}|]\} \quad (3.3.7)$$

Consider p=1 and q=1, $\{\upsilon_t\}$ is a sequence of *i.i.d* with a zero mean, the function could be a simple EGARCH (1,1):

$$ln(\hat{\sigma}_t^2) = \omega + \ ln(\sigma_{t-1}^2) + \alpha\phi\upsilon_{t-1} + \alpha\psi|\upsilon_{t-i}| \quad (3.3.8)$$

If define

$$g(\upsilon_t) = \phi\upsilon_t + \psi|\upsilon_t| \quad (3.3.9)$$

Here $\{g(\nu_t)\}$ is a zero-mean, *i.i.d*, random sequence since the components of $g(\nu_t)$ are $\phi\upsilon_t$ and $\psi|\upsilon_t|$ are zero mean. With the assumption in return process that $\{\nu_t\}$ is normally distributed, the components of $g(\upsilon_t)$ are orthogonal, though they are not independent. If 0≤ $\nu_t < \infty$, with refers to a positive news, $g(\nu_t)$is linear in ν_t with slope $\phi + \psi$, and if -$\infty \leq \nu_t < 0$, with refers to a negative news $g(\nu_t)$ is linear with slope $\phi - \psi$. If$\phi < 0$, $g(\nu_t)$

allows for the conditional variance process $\{\sigma_t^2\}$ to respond asymmetrically to rises and falls in stock price by $|\phi - \psi| > |\phi + \psi|$ and the clustering is captured by the parameter ψ. In the EGARCH model $\ln(\sigma_t^2)$ is homoscedastic conditional on σ_t^2 , and the partial correlation between ν_t and $\ln(\sigma_t^2)$ is constant conditional on σ_t^2.

Threshold-GARCH (TGARCH) models

Another alternative asymmetric model is selected as TGARCH model. The Threshold GARCH (TGARCH) model studied by Glosten et al. (1993) define the conditional variance as a linear piecewise function. In TGARCH (1,1):

$$\hat{\sigma}_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \delta D_t \varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \tag{3.3.10}$$

$$D_t = \begin{cases} 1 & \varepsilon_{t-1} < 0 \\ 0 & \varepsilon_{t-1} > 0 \end{cases} \tag{3.3.11}$$

It can be clearly seen that the leverage effect is captured by the function D_t with measuring the value of ε_{t-1}. Similar with EGARCH models, ω provides the weighted average of the variance, positive news will reflect by the value of α and negative news will be reflected by $\alpha + \delta$. If δ >0, the negative news will have a greater impact $\alpha+\delta$ on volatility rather than the same magnitude of positive news by α. The TGARCH model relaxes the linear restriction on the conditional variance dynamics. It could capture the stylized fact that conditional variance tends to be higher after a decrease in return than after an equal increase.

Component GARCH (CGARCH)

The component GARCH (CGARCH) model following by Engle and Lee (1999) is derived by replacing the constant σ_t^2 with a time varying long-run component q_t. The conditional variance changes by a long run component which is calculated by itself, autoregressive of the first order. The CGARCH model specification is:

$$\hat{\sigma}_t^2 = q_t + \alpha(\varepsilon_{t-1}^2 - q_{t-1}) + \beta(\sigma_{t-1}^2 - q_{t-1}) \tag{3.3.12}$$

$$q_t = \omega + \rho q_{t-1} + \varphi(\varepsilon_{t-1}^2 - \sigma_{t-1}^2) \tag{3.3.13}$$

The long-run component q_t will present the effect by computing the parameter ρ larger than $\alpha + \beta$, the value of $(\varepsilon_t^2 - \sigma_t^2)$ will control the time varying movement of the long-run effect. It constitutes a method of making the long-memory-like features into a short-memory model to capture the salient features of speculative returns.

Realized Volatility

With rapid growth in financial markets, the volatility in financial time series plays an important role for both theoretical and empirical need. The volatility of the returns seems to be related to the returns easier to forecast than returns. In general, estimating variances depends on the size of the sample, which is the number of observations, while estimating the means of returns depends on the length of the sample which is the time period observation. Therefore, using realized variance is easier than returns due to the availability of high-frequency intraday data.

However, the conditional variance is latent, and hence is not directly observable. It can be estimated, among other approaches, by the GARCH as mentioned above. Although these GARCH type models have been extended in lots of directions based on the empirical evidence that the volatility process is non-linear, asymmetry, and has a “long memory” effect, researchers have found that those models cannot describe the whole day volatility information well enough because they were developed within low-frequency time series of squared roots will influence the conditional variance. As observed by Bollerslev (1987), the GARCH models fail to describe all the facts which is important in financial time series. McAleer & Medeiros (2008) documented an empirical fact that standard latent volatility models fail to describe in an adequate manner which is low but slowly decreasing, autocorrelations in the squared returns that are associated with high excess kurtosis of returns. Therefore, making the volatility to be “observable” has led to appearance of the realized volatility.

The estimation and forecasting of the conditional variance with stock return becomes a work to measure a high frequency intraday data. With the work of Merton (1980), if the data can be estimated at an enough high sampling frequency, the variance over a fixed interval or time horizon can be estimated arbitrarily, as a sum of squared realizations. The work by Andersen & Bollerslev (1998) showed that the daily foreign exchange volatility can be best measured by aggregating 288 squared five-minutes returns which make the volatility now appear to be “observable”. More recent work by Takahashi et al. (2024) also suggests that realized volatility provides useful information in forecasting volatility. More specifically, the daily realized volatility which will be used as the “actual volatility” in this chapter are calculated as:

$$RV_t = \sum r_{t,i}^2 \qquad (3.3.14)$$

$$r_{t,i} = \frac{s_{t,i} - s_{t,i-\triangle}}{s_{t,i-\triangle}} * 100\% \quad (3.3.15)$$

where the $\triangle$ is a time interval of 5 minutes, the $r_{t,i}$ is calculated every 5 minutes to represent the returns and sum the square returns to calculate the intraday realized variance RV_t.

Since the volatility becomes "observable," it can be modeled directly, rather than being treated as a latent variable. Several recent studies have documented the properties of realized volatilities constructed from high frequency data in the literature. The Heterogeneous Autoregressive Model (HAR) by Corsi (2004) gives a good approach to focus on the realized volatility in different time horizons by formulating an AR regression on different period of the volatility to measure different weight of volatility in different time horizon so that the property that a short-term movement of the markets might act differently to volatility swings compared to a medium or long-term movements will be identified.

Heterogenous autoregressive models (HAR model)

Another model which is able to capture the long memory effect is first proposed by Corsi (2004), particularly, the heterogeneous autoregressive (HAR) model. It proves to successfully achieve the purpose of modeling the long-memory behavior of volatility in a very simple and parsimonious way (although not formally belonging to the class of long-memory models) by taking the realized volatility into account. The daily HAR is expressed by

$$\hat{\sigma}_t = \beta_0 + \beta_D RV_{t-1,t} + \beta_W RV_{t-5,t} + \beta_M RV_{t-22,t} + \varepsilon_{t,t+1} \quad (3.3.16)$$

where $RV_{t-1,t}$, $RV_{t-5,t}$, $RV_{t-22,t}$ represents to the 1 day, 5 days and 22 days of the realized volatility in a time period which can be viewed as "one trading week" and "one trading month" refer to the average realized volatility of 5 days lagged and 22 days lagged. $\beta_0, \beta_D, \beta_W, \beta_M$ can be estimated with the application of an Ordinary Least Squares (OLS) estimation. The HAR model believes that the latent realized volatility can be observed over time horizons longer than one day. It creates an AR regression of the 1 day, 5 days and 22 days average realized volatility to make forecasting.

Hybrid GARCH model based on Artificial Neural Network (GARCH-NN)

Machine learning methods are widely used not only in financial literature but other areas as well. Specifically, in order to improve the forecasting performance of the GARCH model, a hybrid model based on neural networks is introduced. An artificial neural network (ANN) is a network of interconnected elements called neurons. The neurons are used to estimate functions based on the inputs. The neurons are connected with each other by joint mechanism which consists of a set of assigned weights. The method can be described as follows:

$$\hat{\sigma}_t{}^2 = \varphi(\textstyle\sum_{i=1}^{t-1} \omega_t x_i + b_t) \tag{3.3.17}$$

$\{x_i\}$ is the input data and $\{\omega_{p_i}\}$ describes the connection weights of neurons and b_p is the bias, $\varphi(*)$ is the activation function and $\hat{\sigma}_t$ is the output of the neuron. In ANN works, multi-layer feed forward (MLP) is a common approach which has three layers: input layer, output layer, and hidden layer. Neuron takes the values of inputs parameters, sums them up with the assigned weights, and adds a bias. With the application of transfer function, the outputs will be displayed.

A hybrid GARCH model can be built on the conception of neural networks. In neural networks, the input data can be set as an explanatory variable of financial time series, such as returns, squared returns, trading volumes, etc. A hybrid GARCH of volatility forecasting will set the target on the volatility. The input data will be set as some explanatory variables and the volatility generated by the GARCH type models including standard GARCH, EGARCH, TGARCH and CGARCH mentioned above. A simple specification can be interpreted as:

$$\hat{\sigma}_t^2 = \varphi(b_t + \sum_{i=1}^{t-1} \omega_t {x_i}^2) \tag{3.3.18}$$

It describes the structure of the model for a single forecasting horizon, where the input x_{t-1} is a matrix of the volatility generated by the GARCH type. The model can be separated into a linear autoregressive component of order and a nonlinear component whose structure depends on the number of hidden nodes which is the hidden layer in ANN.

The whole dataset is separated into two parts: one is from 5 January 2005 to 25 November 2014 (2400 observations or approximate 10 years), it is the training set for the machine. The other is from 26 November 2014 to 28 December 2018 (1000 observations or approximate 4 years), it is the part to check the 1 day rolling window forecasting performance of the machine.

The reason to use the GARCH genres is that it provides a more real-world context than other models when trying to predict the prices and rates of financial instruments addressed by Kenton (2020). The reason to combine the GARCH with neural networks is that it do not need too much formal statistical training and it can detect not only complex nonlinear relationships between dependent and independent variables but all possible interactions between predictor variables as well which is mentioned by Tu (1996).

3.4 Comparison methods of forecast performance

The accuracy of forecasting is the main concern when using different models. In order to check the performance of several methods, lots of measurements are used to calculate the forecasting error in financial literature. The chapter selected four widely used methods first to make comparison: the Mean Absolute Error (MAE), the Mean squared error (MSE), the Quasi-Likelihood (QLIKE) and R^2LOG loss function. Moreover, a regression test called Mincer-Zarnowiz regression test (MZ test) is applied to make sure that the "actual volatility" is regressed on the forecast series. Finally, a Model Confidence Set (MCS) procedure by Hansen (2011) is considered to test the equal predictive ability (EPA) at certain confidence level α. It tests models on various aspects depending on the chosen loss function mentioned above including MAE, MSE, QLIKE and R^2LOG. To retrieve a "actual volatility", the realized volatility (calculated by 5 mins intraday returns) $\tilde{\sigma}_t$ is considered both the in-sample and out-of-sample period, which are believed to be closer to the unobservable volatility. The reason to use the realized volatility is that the conditional variance is latent, and hence is not directly observable, the realized volatility provides a measure of the historical performance of an asset which implies that one comes to know if the asset's price has been fluctuating a lot or not.

The forecasting values are specified as $\hat{\sigma}_t$.

Mean Absolute Error (MAE) and Mean Squared Error (MSE)

The Mean Absolute Error (MAE) statistics are based on the average absolute forecast error, it measures the average magnitude of the errors in a set of predictions, without considering their direction. It calculates the average difference of the comparison with equal weight of all individual differences. Another method called Mean-Squared Error (MSE) calculates the average squared difference between the estimated values and the actual values. Since the errors are squared before they are averaged, the MSE is sensitive with large errors due to the relatively high weight of them by squared function, which means that the MSE should be more useful when large errors are particularly undesirable. Both comparison measures report the better performance by a lower statistic.

$$MAE = \frac{1}{\delta}\sum_{t=T+1}^{T+\delta}|\tilde{\sigma}_t^2 - \hat{\sigma}_t^2| \quad (3.4.1)$$

$$MSE = \frac{1}{\delta}\sum_{t=T+1}^{T+\delta}(\tilde{\sigma}_t^2 - \hat{\sigma}_t^2)^2 \quad (3.4.2)$$

Quasi-Likelihood (QLIKE) and R^2LOG loss function

The Quasi-Likelihood (QLIKE) loss function is a test of forecast bias implied by a Gaussian likelihood which has a specification:

$$QLIKE_t = ln(\hat{\sigma}_t^2) + \frac{\hat{\sigma}_t^2}{\tilde{\sigma}_t^2} \quad (3.4.3)$$

The QLIKE is found to be robust to noise when used to compare matching volatility prediction models. It means that using a proxy for volatility does not influence the performance ranking as using the true conditional variance is more reliable than MAE. The other loss function is the R^2LOG which measure the goodness of fit of the out-of-sample forecasts.

$$R^2LOG_t = \left[ln\left(\frac{\hat{\sigma}_t^2}{\tilde{\sigma}_t^2}\right)\right]^2 \qquad (3.4.4)$$

Mincer-Zarnowitz Regression test

A useful testing tool called Mincer-Zarnowitz regression test (MZ test) is applied to make sure that the "actual volatility" is regressed on the forecast series. The test procedure was proposed by Mincer & Zarnowitz (1969). Particularly, for volatility forecasting, the test tries to find a linear regression between on the "actual volatility", here, the realized volatility, and the forecast values. The regression specification can be written as:

$$\tilde{\sigma}_t^2 = \alpha + \beta\ \hat{\sigma}_t^2 + \varepsilon_t \qquad (3.4.5)$$

It tests the joint hypothesis that the intercept is 0 and the slope is 1. The R^2 statistics will report how well the regression should be. However, there are still lots of methods to test the regression such as Diebold-Mariano regression, Giacomini-White regression, etc. Since the forecast values by each model are univariate individuals, the Mincer-Zarnowiz regression test is suitable to analysis the regression.

Model Confidence Set (MCS)

Since varieties of models are available in recent years, it is crucial to find a statistical method or procedure that delivers the best performing models with respect to a given criterium. The Model Confidence Set procedure (MCS) as a statistic testing method, permits to construct a set of "superior" models which is called "Superior Set Models" (SSM) by testing whether the null hypothesis of equal predictive ability (EPA) is not rejected at certain confidence level α. The EPA statistic tests are evaluated for an arbitrary loss function which can be customed with individuals. The Model Confidence Set procedure starts from an initial set M_0 including all the models being compared and results in a smaller set $\widehat{M}^*_{1-\alpha}$ (SSM). The procedure will iterate with the EPA hypothesis test until null hypothesis is accepted and create an optimal SSM set. Otherwise, the EPA should be tested again after eliminating the worst model. The procedure specification of volatility can be viewed by some steps:

First the loss function can be defined as equation (3.4.5), where the σ_t is the actual volatility and $\hat{\sigma}_{i,t}$ is the output of model *i* at time *t*. The loss differential between models *i* and *j* is specified as equation (3.4.6)

$$l_{i,t} = l(\tilde{\sigma}_t, \hat{\sigma}_{i,t}) \qquad (3.4.6)$$

$$d_{ij,t} = l_{i,t} - l_{j,t} \qquad (3.4.7)$$

The loss function for volatility in this chapter are specified as the MAE, MSE, QLIKE, R^2LOG mentioned above and there are more measures proposed by Hansen & Lunde (2005). The average loss of model *i* relative to any other model *j* at time *t* can be defined as equation (3.4.7), where the M is the initial set of m competing models at the first step mentioned above

$$d_{i\cdot,t} = (m-1)^{-1} \sum_{j \in M} d_{ij,t} \qquad (3.4.8)$$

The EPA hypothesis mentioned above for the given set M can be formulated in two alternative ways:

Null hypothesis: $H_{0,m}: E\left(d_{ij}\right) = 0$ for all *i, j* = 1, 2, …, m (3.4.9)

Alternative hypothesis: $H_{A,m}: E\left(d_{ij}\right) \neq 0$ for some *i, j* = 1, 2, …, m (3.4.10)

Or

Null hypothesis: $H_{0,m}: E(d_{i\cdot}) = 0$ for all *i, j* = 1, 2, …, m (3.4.11)

Alternative hypothesis: $H_{A,m}: E(d_{i\cdot}) \neq 0$ for some *i, j* = 1, 2, …, m (3.4.12)

Two statistics corresponding with the test procedure are created by Hansen (2011):

$$t_{ij} = \frac{\bar{d}_{ij}}{\sqrt{\widehat{var}(\bar{d}_{ij})}} \text{ and } t_{i\cdot} = \frac{\bar{d}_{i\cdot}}{\sqrt{\widehat{var}(\bar{d}_{i\cdot})}} \quad (3.4.13)$$

$$T_{R,M} = \max_{i,j\in M}\left|t_{ij}\right| \text{ and } T_{max,M} = \max_{i,\in M}\left|t_{i\cdot}\right| \quad (3.4.14)$$

where $\bar{d}_{ij} = \frac{\sum_{t=1}^{T} dij,t}{T}$ measures the relative sample average loss between the model *i* and model *j* during the whole time period T. Likewise, $\bar{d}_{i\cdot} = \frac{\sum_{j\in M}\bar{d}_{ij}}{m-1}$ measure the sample loss of the model i relative to the averages losses across models in the set M. The $\widehat{var}(\bar{d}_{ij})$ and $\widehat{var}(\bar{d}_{i\cdot})$ are the bootstrapped variance estimates. The test statistics defined in formula (3.4.13) can be used to test the two hypotheses mentioned above, respectively.

The MCS procedure will eliminate the worst model at each step if the null hypothesis of equal predictive ability (EPA) is not accepted and the procedure will iterate until all models with EPA hypothesis entered in the "Superior Set Models" (SSM). The choice of the model is made by an eliminated rule with equation (3.4.14) corresponding to the statistics defined above. Ideally, the best scenario is when the final set consists of a single model.

$$e_{R,M} = \arg\max_{i} \left\{ \underbrace{sup}_{j \in M} \frac{\bar{d}_{\mathrm{ij}}}{\widehat{var}(\bar{d}_{\mathrm{ij}})} \right\} \text{ and } e_{max,M} = \arg\max_{i \in M} \frac{\bar{d}_{i\cdot}}{\widehat{var}(\bar{d}_{i\cdot})} \tag{3.4.15}$$

Hansen et al. (2010) pointed out the advantage of MCS procedure is that it is able to make statements about significance that are valid, in the traditional sense.

In the comparison of predictive ability with different volatility models, common comparing measures usually does not report a unique result since different models are usually built to identify different stylized facts such as clustering, asymmetry and "long memory". It is hard to point out that a single model outperforms others by the reason of their statistically equivalent property or not enough information from the data. The MCS procedure provides a new approach to identify the best fitting model.

The reason to choose MAE and MSE is that MAE is useful for consistent error measurement across all data points, while MSE is preferred when penalizing significant errors is crucial.

However, the MSE method penalizes symmetrically. Therefore, QLIKE function which Sturesson & Wennström, (2023) states that is an asymmetrical loss function that penalizes under-prediction heavier than over-prediction is selected. The R^2LOG is selected as well since it is often easier to interpret since it doesn't depend on the scale of the data.

3.5 Results and analysis of empirical exercise

All the data and models and comparisons are applied with R studio in an R language environment.

Table 3.5.1 shows the MAE statistics of the volatility forecasting models including GARCH type models, their extension based on neural networks and a HAR-RV model as well. The realized volatility which is calculated based on the five-minutes intraday returns each day acts as the “actual” volatility. The value of the MAE statistics in the table is multiplied by 10^4 to make it simpler to read and compare in an appropriate term. The hybrid GARCH models based on neural networks are written as GARCH-NN, EGARCH-NN, TGARCH-NN and CGARCH-NN which correspond to the different GARCH type models in the table.

From the table, it can be seen that the CGARCH model has a better performance than others in 4 out of 12 sample countries which is same with result of the work by Kambouroudis (2012). The second is the hybrid model GARCH-NN which is a standard symmetric GARCH model based on neural networks in 3 countries followed by the TGARCH-NN and CGARCH-NN for in 2 countries, respectively. Finally, the EGARCH-NN reports a minimum MAE value in 1 country. In addition, there is no minimum MAE value reported for GARCH, EGARCH, TGARCH, and HAR models which means that these models do not appear to be the best model when using MAE statistics in the12 sample countries.

The second-best performing models can be found by the second minimum MAE values reported in each row. Both HAR model and CGARCH-NN appear to be the first with a second minimum MAE value in 4 countries, respectively. TGARCH-NN comes to be the second of the 2 countries. The rest are the GARCH-NN and EGARCH-NN in 1 country each. Again, there is no second minimum MAE values reported for GARCH, EGARCH, TGARCH and CGARCH, which means the whole GARCH type models are excluded from the second-best performer.

The third-best performing models can be observed by the third minimum MAE values in each row. The results are in a wide dispersion. The GARCH, GARCH-NN and TGARCH-NN takes the first in 3 countries, second is TGARCH, CGARCH, and EGARCH-NN in 1 country each.

Finally, Table 3.5.1 also reported the worst performing models of volatility forecasting by considering the maximum MAE statistics in every country. Here, the EGARCH appears to be the worst performer with a maximum MAE value in 5 countries.

With looking the whole performance of the selected models, although the best models are considered to be a "long memory" model, the CGARCH - in 4 countries out of 12, the hybrid GARCH type models based on neural networks also perform well overall, particularly, they act as the best model in the rest of 8 countries. An overall conclusion can be drawn that the hybrid GARCH models outperform the GARCH type models and HAR model. Moreover, the second-best models are still taken by these hybrid GARCH models, especially the extension of a long memory GARCH models, CGARCH-NN, in 4 out of 12 countries. These hybrid GARCH models come to be superior when considering the third best model as well, which performs better in 7 countries out of twelve rather than the GARCH type models in 5 countries. Although the HAR model is marked as the second-best performer as well as CGARCH-NN, it has not been reported as the best or third best models in the rest of the countries.

Since the MAE methods are the simplest way like an "naïve method" to compare the error which takes the difference with an equal weight, the MSE measure is carried out to identify if there exist some outlier predictions with huge errors.

Similarly, Table 3.5.2 reports the MSE statistics for the volatility forecasting models in 12 sample countries. The MSE will puts larger weight on the errors due to the squaring part of the function which will enlarge the outlier prediction error. Unlike when using MAE statistics, the TGARCH-NN model comes to be the first with a minimum MSE value in 5 out of 12 countries. The second is the GARCH-NN in 3 countries and the third is the CGARCH-NN in 2 countries. There are only 1 minimum MSE values reported for each country of CGARCH and HAR models although the CGARCH are the best performer when using MAE. Again, there is no minimum MSE value reported for GARCH, EGARCH, TGARCH and EGARCH-NN.

The second-best model is the CGARCH-NN in 4 countries, TGARCH-NN and EGARCH-NN in 3 countries, the CGARCH and HAR model in 1 country each. The GARCH, EGARCH, TGARCH are found to have no countries where they perform as best or second best. Some similar results are also obtained by the research of Chen (2023).

The third-best model is recorded as EGARCH-NN in 4 countries. The second is the GARCH, CGARCH and GARCH-NN in 2 countries, respectively. The TGARCH-NN and CGARCH-NN come to the last with a minimum MSE value in 1 country each.

The EGARCH once again comes to be the worst performing model with the highest MSE statistics in 6 out of 12 sample countries.

By taking a whole view of Table 3.5.2, the conclusion is different of the measure MSE. The hybrid model TGARCH-NN has come to be the best performing model. Moreover, these hybrid models appear to be superior to others in the best, second, and third best performer category, which outperforms in 10 out of 12 countries when considering the best and second, outperforms 8 out of 12 countries when considering the third best models.

With the application of MZ test, there is a disadvantage: a large value will has a larger impact on the regression and it may generates bias on the test results. Therefore, with the work of Pagan & Schwert (1990) and Kambouroudis (2012), two regressions are introduced in the general form of MZ test but has the ability to rescale the parameters:

$$\tilde{\sigma}_t^2 = \alpha + \beta\ \hat{\sigma}_t^2 + \varepsilon_t \qquad (3.5.1)$$

$$\log\left(\tilde{\sigma}_t^2\right)\tilde{\sigma}_t^2 = \alpha + \beta \log\left(\ \hat{\sigma}_t^2\right) + \varepsilon_t \qquad (3.5.2)$$

In Table 3.5.3, the R^2 statistics for the coefficient of the regression of all the models are reported. The EGARCH-NN performs better in 6 out of 12 sample countries followed by TGARCH-NN in 3 countries. The EGARCH, TGARCH and HAR models has a higher R^2 in

1 country, respectively. In addition, there is no maximum R^2 value reported for GARCH, CGARCH, GARCH-NN and CGARCH-NN.

The second-best performer can be found by estimating the second highest R^2 value. The TGARCH-NN outperforms the rest of the models in 6 out of 12 countries. The second place is the TGARCH and CGARCH-NN with a second maximum R^2 value in 2 countries, respectively. The third is EGARCH and EGARCH-NN in 1 country each. Again, the GARCH, CGARCH and GARCH-NN models are found to have no countries where they perform as best or second best.

The third-best performing model is the GARCH-NN and CGARCH-NN model in 4 out of 12 countries, respectively. The rest are the EGARCH, CGARCH, EGARCH-NN and HAR comes with a third maximum R^2 value in 1 country each.

The worst performing model can be identified by observing the lowest R^2 value in Table 3.5.3. The GARCH model comes to be the worst performer in 8 out of 12 sample countries. The HAR model are reported to be the worst in 3 countries and EGARCH, GARCH-NN, CGARCH-NN in 1 country, respectively.

The overall performance of the volatility models by regression test suggests that the hybrid GARCH type models, especially the asymmetric hybrid models (TGARCH, EGARCH) based on neural networks has a better regression of forecast values than the rest of the models.

There are some weaknesses or limitations of the MAE, MSE or MZ regression test. The MAE only compares the error which takes the difference with an equal weight, which will average some outlier predictions with huge errors. Although the MSE measure can enlarge the error with its squaring part of function, the squaring part will magnify the error if the model only makes a single bad prediction, and this will affect the overall judgement of a model. Likewise, the MZ regression test only concerns the linear regression of the coefficient. In order to observe the forecasting performance of each volatility model more detailly, the MCS procedure is introduced.

Loss of the volatility

Model Confidence Set procedure (MCS) is an iteration of a sequence of statistic tests which permits to construct a set of “superior” models, the “Superior Set Models” (SSM), under the null hypothesis of equal predictive ability (EPA) is not rejected at certain confidence level α based on a loss function. There are six loss function of volatility evaluating based on the (MCS) addressed by Hansen and Lunde (2005) which is mentioned above in section 3.5. Before applying this statistic tests procedure, a direct comparison based on the loss function are taken out. The details of the “Superior Set Models” (SSM) will be displayed as a conclusion in next section.

The QLIKE and R^2LOG are selected to check if there is any significant difference. The loss function specification is:

$$QLIKE_t = ln(\hat{\sigma}_t^2) + \frac{\hat{\sigma}_t^2}{\tilde{\sigma}_t^2} \qquad (3.5.3)$$

$$R^2LOG_t = \left[ln\left(\frac{\hat{\sigma}_t^2}{\tilde{\sigma}_t^2}\right)\right]^2 \quad (3.5.4)$$

where $\tilde{\sigma}_t$ refers to the "actual volatility" and evaluated volatility is specified as $\hat{\sigma}_t$.

The R^2LOG statistics prefers a value closer to zero which indicates that the distance between the "actual" volatility and forecasting volatility is small enough. The value is no doubt nonnegative due to the squaring part of the function so that a lower R^2LOG statistics is preferred when comparing different models.

With the application of the QLIKE loss function in equation (3.5.1), it is hard to say that a lower statistic indicates a better performance. In order to make it more appropriate to compare directly, a specification is defined as:

$$QLIKE_{t+1} = \underbrace{\left(ln(\tilde{\sigma}_{t+1}^2) - ln(\hat{\sigma}_{t+1}^2)\right)}_{1} + \underbrace{\left(\frac{\hat{\sigma}_t^2}{\tilde{\sigma}_t^2} - 1\right)}_{2} \quad (3.5.5)$$

The two parts of the QLIKE statistic measures the log difference between the "actual" volatility and forecasting volatility and their rates. Both two parts tends to be zero if the forecasting volatility tends to be same as the "actual" volatility so that a QLIKE statistics closer to zero will be preferred when comparing different models.

In Table 3.5.4 the QLIKE statistics based on equation (3.5.3) are reported. Both GARCH-NN and TGARCH-NN model comes to be the first with a minimum QLIKE value in 3 out of 12 countries, respectively. The CGARCH and CGARCH-NN appear to be the second both with a minimum QLIKE value in 2 countries. The third is the TGARCH and HAR in 1 country each. Similarly, there is no minimum QLIKE value reported for GARCH, EGARCH and EGARCH-NN.

The second-best performing models can be found by the second minimum QLIKE statistics reported in each row. CGARCH-NN appears to be the first in 4 countries followed by CGARCH in 3 countries. The third is the HAR model in 2 countries and the rest are GARCH-NN, EGARCH-NN, TGARCH-NN in 1 country each. Again, there is no second minimum QLIKE value reported for GARCH, EGARCH and TGARCH.

The third-best performing models can be observed by the third minimum QLIKE values in each row. The results are similar as the MAE and MSE results in a wide dispersion with no doubt. The EGACH-NN is the first in 3 counties followed by GARCH, TGARCH-NN and HAR in 2 countries, respectively. Other models including TGARCH, CGARCH and GARCH-NN in 1 country each.

Finally, Table 3.5.4 also reported the worst performing models of volatility forecasting by considering the maximum QLIKE statistics in each country. The EGARCH appears to be the worst performer with a maximum QLIKE value in 4 countries.

The overall conclusion of Table 3.5.4 generated by the whole performance of the selected models are similar with the measures applied above. The hybrid GARCH models perform better than others, but the best models are considered to be in a difference when using different measures: CGARCH when using MAE, TGARCH-NN when using MSE, EGARCH-NN when using MZ test and here GARCH-NN and TGARCH-NN when using QLIKE. Likewise, with the application of QLIKE measure these hybrid models appear to be superior when considering the second and third best models as well. These hybrid models come to be a second-best performer in 7 countries while GARCH type in 3 and HAR in 2. When considering the third-best performer, these hybrid models outperforms in 6 countries while GARCH type in 4 and HAR in 2.

Another measure of the loss function based on the MCS procedure mentioned above is the R^2LOG. Unlike the QLIKE measure it only concerns the LOG difference between the "actual" volatility and the forecasting values.

Similar as the QLIKE measure, Table 3.5.5 reports the R^2LOG statistics of the loss function for the volatility forecasting models in 12 sample countries. A lower value is preferred when selecting a better performance model. Unlike when using QLIKE, the R^2LOG statistics indicated that CGARCH, GARCH-NN and TGARCH-NN performs better in 3 out of 12 countries, respectively, which both include the "long memory" GARCH type models and hybrid GARCH models. The second is CGARCH-NN in 2 countries followed by the third of TGARCH in 1 country. Again, there is no minimum R^2LOG value reported for GARCH, EGARCH and HAR model

The second-best model is the CGARCH-NN in 4 countries, HAR model in 3 countries, the CGARCH in 2 countries and the GARCH-NN, EGARCH-NN, TGARCH-NN in 1 country, respectively. GARCH and EGARCH models are found to have no countries where they perform as best or second best.

The third-best model is recorded as EGARCH-NN in 4 countries. The second is the GARCH, EGARCH and TGARCH-NN in 2 countries, respectively. The TGARCH and CGARCH come to the last in only 1 country each.

The EGARCH not surprisingly comes to be the worst performing model with the highest R^2LOG values in 4 out of 12 sample countries.

Again, in attempt to look at the overall performance of the models using R^2LOG loss function in Table 3.5.5, the conclusion is similar to the results from QLIKE loss function. The hybrid model GARCH-NN, TGARCH-NN and CGARCH-NN takes most of the position when identifying a best model. Moreover, the "long memory" GARCH type models, CGARCH also has a good performance in R^2LOG statistic report. Likewise, the hybrid models are still superior when considering the second-best models by outperforming in 7 countries out of 12, while the GARCH type models comes to more when considering the third-best models. Both GARCH type models and hybrid ones appear to be the third-best models in 6 countries.

Table 3.5.1 MAE statistics of volatility forecasting models in all samples*10^-4

	GARCH	EGARC	TGARC	CGARC	GARCH-	EGARCH-	TGARCH-	CGARCH-	HAR
USA	0.48441	0.51465	0.49894	0.44456	0.62636	0.61987	0.61254	0.6277 *	0.46697
UK	0.53067	0.52384	0.51752	0.48433	0.59988	0.61635 *	0.61389	0.60099	0.48942
Switzerl	0.61921	0.67593	0.62578	0.55404	0.37285 c	0.37394	0.35887 b	0.35448 a	0.58010
Korea	0.43238	0.47351	0.49277	0.39475	0.23831 a	0.39495	0.39979	0.37565 b	0.45634
Japan	1.27871	1.22068	1.20895	1.19354	0.64895	0.63071 c	0.46068 a	0.47930 b	1.25448
India	0.52288	0.61883	0.62880	0.48621	0.79524	0.81621	0.82035 *	0.56671	0.52174
HK	0.88333	0.95334	0.92955	0.81122	0.45170 b	0.47399	0.46842 c	0.34402 a	0.89501
German	0.77622	0.8964 *	0.86955	0.69921	0.61901 c	0.61372 b	0.56369 a	0.62163	0.74994
France	0.72936	0.81053	0.75651	0.65132	0.42408 a	0.55692	0.53319 c	0.50092 b	0.70659
China	1.69475	1.66042	1.68216	1.63767	1.56216 c	1.26546 a	1.26919 b	1.56578	1.65563
Canada	0.32501	0.36212	0.36086	0.29920	0.55342	0.54569	0.55450 *	0.55305	0.32267
Australia	0.39618	0.41877	0.41015	0.35537	0.22846 a	0.24558	0.24451 c	0.23784 b	0.37577

Note: a: Best performer, b: Second best performer, c: Third Best performer, *: Worst performer

Table 3.5.2 MSE statistics of volatility forecasting models in all samples*10^-8

	GARCH	EGARC	TGARC	CGARC	GARCH-	EGARCH-	TGARCH-	CGARCH-	HAR
USA	1.73871	1.74640	1.73445	1.66423	1.51509 c	1.49050 b	1.47796 a	1.52203	1.58341
UK	3.92500	3.89797	3.8998	3.83667	3.35137	3.33733 c	3.33479 b	3.32971 a	3.66762
Switzerl	2.76779	3.17322	2.58144	2.523	0.43491 c	0.45052	0.41406 a	0.41595 b	2.04765
Korea	0.32819	0.40103	0.41606	0.29151	0.16698 a	0.29294	0.29674	0.28510 b	0.35112
Japan	4.08200	3.54615	3.52607	3.68236	1.65084	1.50502 c	1.22938 a	1.29482 b	3.7658
India	0.56489	0.78016	0.75729	0.51944	1.15968	1.20112	1.20607 *	0.81878	0.56014
HK	1.47589	1.6714	1.59425	1.26938	0.79192	0.77488 c	0.77098 b	0.54767 a	1.48275
German	1.53319	1.90869	1.79847	1.32796	0.82028	0.74986 b	0.67570 a	0.79255 c	1.33551
France	1.60642	1.90909	1.66952	1.34897	0.53877 a	0.78453	0.75180 c	0.65052 b	1.37202
China	11.05978	10.6890	10.9525	10.4777	12.05234	10.24947 b	10.24578 a	12.14279 *	10.76963
Canada	1.13559	1.14506	1.13827	1.09928	1.32261	1.32722	1.3252	1.33363 *	1.08457
Australia	0.34851	0.36462	0.35119	0.30285	0.15324 a	0.15439	0.15407 b	0.15574	0.29552

Note: a: Best performer, b: Second best performer, c: Third Best performer, *: Worst performer

Table 3.5.3 R^2 statistics for MZ test of volatility forecasting models in all samples

	GARCH	EGARC	TGARC	CGARC	GARCH-	EGARCH-	TGARCH-	CGARCH-	HAR
USA	0.1931 *	0.234	0.2355	0.2081	0.3703	0.39560 a	0.3954 b	0.3711 c	0.2641
UK	0.06256	0.0677	0.06642	0.06805	0.2418	0.25070 a	0.2503 b	0.2485 c	0.00027
Switzerl	0.05692	0.05612	0.06155	0.0588	0.82880 c	0.8253	0.8410 a	0.8335 b	0.2236
Korea	0.1129 *	0.1772	0.1854	0.1209	0.19840 c	0.21790 a	0.2169 b	0.1748	0.1729
Japan	0.1523	0.2069	0.2094	0.1555	0.4012	0.42930 c	0.4688 a	0.4355 b	0.09955
India	0.1297	0.17210	0.17490	0.1333	0.11520 *	0.1398	0.1363	0.1172	0.1588 c
HK	0.1347 *	0.1542	0.1524	0.1392	0.2935	0.31110 a	0.3087 b	0.2985 c	0.1903
German	0.2161 *	0.2595	0.2689	0.2208	0.494	0.53580 b	0.5410 a	0.4955 c	0.2209
France	0.1966 *	0.2513	0.2664	0.2069	0.5510 c	0.57320 a	0.5729 b	0.5497	0.3028
China	0.3432	0.36740	0.36250	0.35840	0.2653	0.3488	0.349	0.2592	0.1257 *
Canada	0.07989	0.09704	0.09840	0.08681	0.07474	0.07067	0.07389	0.06886 *	0.1124 a
Australia	0.2388 *	0.2844	0.2856	0.2568	0.52070 c	0.53730 a	0.5359	0.5193	0.3763

Note: a: Best performer, b: Second best performer, c: Third Best performer, *: Worst performer

Table 3.5.4 QLIKE statistics of loss function for volatility forecasting models in all samples

	GARCH	EGARC	TGARC	CGARC	GARCH-	EGARCH	TGARCH	CGARCH	HAR
USA	1.52444	1.5019	1.46946	1.46959	1.96871	1.95389	1.94353	1.97670 *	1.48402
UK	1.20461	1.19183	1.18545	1.16115	1.40993	1.43061 *	1.42672	1.41474	1.09933
Switzerl	1.31612	1.35015	1.32494	1.25112	1.09763	1.09051 c	1.07623 b	1.07081 a	1.24739
Korea	1.55011	1.55738	1.6125 *	1.47048	1.07508 a	1.4839	1.49608	1.44918 b	1.57521
Japan	1.99759	1.9529	1.95022	1.93956	1.46920 c	1.47088	1.22710 a	1.25361 b	1.95871
India	1.48918	1.53855	1.58129	1.42755	1.80398	1.81484	1.82244 *	1.50198	1.47878
HK	1.60804	1.64965	1.64188	1.5561	1.09433 b	1.1263	1.11923 c	0.97700 a	1.60591
German	1.31212	1.37694	1.36067	1.24178	1.26127	1.25491 c	1.17834 a	1.27638	1.28717
France	1.27441	1.29394	1.25498	1.20756	1.02615 a	1.21185	1.18269 c	1.14861 b	1.25396
China	1.51456	1.49163	1.50179	1.51512	1.65396	1.41565 b	1.41514 a	1.66194 *	1.46190
Canada	1.28333	1.3252	1.33973	1.23525	1.86666 *	1.83435	1.86157	1.86313	1.27530
Australia	1.36356	1.39022	1.39255	1.28202	1.02758 a	1.08064 c	1.08211	1.06211 b	1.30063

Note: a: Best performer, b: Second best performer, c: Third Best performer, *: Worst performer

Table 3.5.5 R^2LOG statistics of loss function for volatility forecasting models in all samples

	GARCH	EGARC	TGARC	CGARC	GARCH-	EGARCH-	TGARCH-	CGARCH-	HAR
USA	1.21002	1.11153	1.07756	1.10200	2.22134	2.18026	2.16455	2.24107 *	1.18445
UK	0.65939	0.62713	0.61213	0.57872	1.03004	1.06117 *	1.05683	1.04023	0.59089
Switzerl	0.80376	0.8248 *	0.79353	0.71449	0.57566	0.56453 c	0.55512 b	0.55016 a	0.73997
Korea	1.11389	1.13316	1.20694	1.00384	0.54897 a	1.02045	1.04311	0.98795 b	1.16001
Japan	2.04339	1.90321	1.89302	1.91886	1.10977	1.09838 c	0.73133 a	0.76111 b	1.97342
India	1.05888	1.1451	1.20065	0.97043	1.63095	1.64896	1.66746 *	1.09698	1.03996
HK	1.23678	1.30964	1.28775	1.14389	0.56651 b	0.61308	0.60575 c	0.4032 a	1.24516
German	0.90173	0.93108	0.91095	0.80207	0.85422	0.82134 c	0.70896 a	0.87833	0.87545
France	0.85304	0.84412	0.79041	0.75017	0.46766 a	0.76045	0.72457 c	0.66612 b	0.83026
China	1.01867	0.95169	0.97069	1.01399	1.23875	0.80039 b	0.7998 a	1.24500 *	0.99893
Canada	0.83209	0.87692	0.89581	0.76359	1.87878	1.83242	1.88031 *	1.87814	0.82804
Australia	0.89984	0.93265	0.92505	0.77931	0.5003 a	0.57666 c	0.57741	0.54616 b	0.84522

Note: a: Best performer, b: Second best performer, c: Third Best performer, *: Worst performer

3.6 Discussion and Findings

In order to determine the best and worst performance models, Table 3.6.1 which summary the empirical results are created. It recorded the best, and worst performer based on five measures of evaluating the forecast among GARCH type models, their extension based on neural networks and HAR model in the 12 sample countries.

It cannot make a clear suggestion that a certain model has the best performance. By ranking all the individual models in the table above, the results showed that the TGARCH-NN ranked the first, followed is GARCH-NN and CGARCH. The CGARCH-NN ranked fourth and EGARCH-NN fifth. The TGARCH and HAR model are the sixth, the EGARCH eighth and finally GARCH ninth. If ranking by the category, the hybrid GARCH type models is the first, the traditional GARCH models is the second, the HAR-RV model is the third. Based on the classification, the asymmetric hybrid model first, the standard hybrid GARCH second, the traditional "long memory" model third, the "long memory" hybrid model fourth followed by HAR and traditional asymmetric model.

3.6.1 GARCH estimation for Neural Networks

The ranking above raises a question that whether a good estimating performance by GARCH will results in a good forecasting performance in the hybrid GARCH based on neural networks. For this reason, the table is separated into two parts which recorded the estimation of GARCH type models and the forecasting performance of the hybrid models based on the corresponding GARCH models.

Table 3.6.2 summarizes a result of the best and worst performers in the estimation of GARCH type models, similarly, using the five comparing techniques mentioned in previous chapter. The asymmetric GARCH type models, especially EGARCH in the exercise, come to be the best performance model when estimation followed by the TGARCH which indicates the asymmetric models as well. The worst performer is identified as the standard GARCH.

Table 3.6.3 summarizes the results of the best and worst performers in hybrid GARCH type models based on neural networks. The TGARCH-NN comes to be the best performer followed by GARCH-NN. However, the standard GARCH model

is identified as the worst of estimation in Table 3.6.2. Likewise, although the EGARCH is record as the best model of estimation in the Table 3.6.2 above, EGARCH-NN is not identified as the best forecasting hybrid models but the second-worst performer for 15 times in Table 3.6.3.

3.6.2 Model Confidence Set

As mentioned in the comparing technique section (section 3.5), a "Superior Set Models" (SSM) will be generated after the Model Confidence Set procedure (MCS) to test the equal predictive ability (EPA) hypothesis. The models which enters the SSM set will be assumed to have an equal predictive ability under a confidence level. With the direct comparison of the measure of MAE, MSE, QLIKE and R^2LOG statistics, although some basic results are reported in section 3.5, it is still necessary to consider the forecasting ability of each model.

Table 3.6.4 reports the compositions of the Superior Set of Models discriminating by model based on the volatility loss function of at a 95% confidence level, particularly, the MAE, MSE, QLIKE and R^2LOG measures. The different values in each column represent the number of models that enters the Superior Set Model at the end of the MCS procedure, when the null hypothesis of equal predictive ability (EPA) is not rejected at the 95% confidence level.

It can be seen that the CGARCH and GARCH-NN rank the first by "surviving" from the procedure of 24 times followed by TGARCH-NN with the number of 23. The fourth place is the HAR model with number of 19. The EGARCH and EGARCH-NN ranked the fifth with the number of 18, respectively. The seventh and eighth is the GARCH and TGARCH models, while the CGARCH-NN appears to be the last. This result gives evidence that the "long memory", CGARCH as well as the hybrid GARCH model are more likely to capture the dynamics existed in the data. However, the hybrid GARCH based on a "long memory" model perform the worst.

3.6.3 Conclusion

Taking an overview of all these tables both in section 3.5 and section 3.6, some conclusions can be generated. The first thing is that the hybrid GARCH model based on neural networks will improve the forecasting performance of a traditional GARCH model which is contributed to the literature such as Hyup Roh, (2007), Bildirici & Ersin (2009) and Werner & Hanka (2016). Second, by comparing the performance in GARCH class models, the long memory model comes to be the first followed by asymmetric ones and the symmetric GARCH models comes to the last which can be found in lots of literatures. Third, there is no strong evidence to support that the success in GARCH type estimation will yield in a success of the application of a corresponding hybrid GARCH models based on neural networks which can be generated from the comparison in section 3.6.2. It means that although the input which generated from a "winner" among GARCH type models, there do not exist empirical evidence that the input will give a more accurate forecasts than others. Since some of the literature concerned with the hybrid models only selected one or two tradition GARCH type models to build hybrid ones, or just fit the model in a certain asset, which makes the results very limited or not comprehensive. The conclusion of the exercise makes a new approach to compare four GARCH type models and four hybrid GARCH in twelve countries with five comparison measures, which fill the gap of limited models or samples mentioned above. For further conclusion, the results from the Superior Set of Models (SSM) of the Model Confidence Set (MCS) procedure gives empirical evidence that that "The hybrid GARCH models will not have equal predictive ability as their corresponding GARCH models.

Moreover, the hybrid GARCH model based on neural networks are more likely to capture asymmetric effects. The sum number of TGARCH-NN and EGARCH-NN in MCS procedure are far more than the sum of traditional EGARCH and TGARCH. The results in section 3.5 also indicate that either the TGARCH-NN or EGARCH-NN models have a better performance overall. However, by looking with the surviving number of "CGARCH-NN" by MCS procedure, which is the lowest number in the table, it gives some hints that the hybrid long memory GARCH model, CGARCH-NN has some weakness to identify the "long memory" effect. In more detail, although the CGARCH model performs better than others, it does not mean that the CGARCH-NN will outperform. The CGARCH model are more suitable to identify the "long memory". The results is useful for future research, the hybrid models can be applied directly when forecasting univariate volatility. For economics, the policymakers can benefit from the results to formulate their policies to avoid risk. The investors can use to appropriate models in this empirical chapter to forecast more recent volatility to avoid risk and loss or to revise their portfolio to make more profits.

Table 3.6.1 Summary for all models										
Measure	Performance	GARCH	EGARCH	TGARCH	CGARCH	GARCH-NN	EGARCH-NN	TGARCH-NN	CGARCH-NN	HAR
MAE	Best	0	0	0	4	3	1	2	2	0
	Worst	2	5	1	0	0	1	2	1	0
MSE	Best	0	0	0	1	3	0	5	2	1
	Worst	2	6	1	0	0	0	1	2	0
R^2	Best	0	1	1	0	0	6	3	0	1
	Worst	6	1	0	0	1	0	0	1	3
QLIKE	Best	0	0	1	2	3	0	3	2	1
	Worst	1	4	2	0	1	1	1	2	0
R^2LO	Best	0	0	1	3	3	0	3	2	0
	Worst	2	4	1	0	0	1	2	2	0
Total	Best	0	1	3	10	12	7	16	8	3
	Worst	12	20	5	0	2	3	6	9	3

Table 3.6.2 Summary performance of GARCH type estimation					
Measure	Performance	GARCH	EGARCH	TGARCH	CGARCH
MAE	Best	0	9	1	2
	Worst	10	1	1	0
MSE	Best	0	10	1	1
	Worst	11	0	0	1
R^2	Best	0	10	2	0
	Worst	8	0	0	4
QLIKE	Best	0	0	6	6
	Worst	9	2	1	0
R^2LOG	Best	0	0	7	5
	Worst	9	1	2	0
Total	Best	0	29	17	14
	Worst	47	4	4	5

Table 3.6.3 Summary performance of hybrid GARCH forecasting					
Measure	Performance	GARCH-NN	EGARCH-NN	TGARCH-NN	CGARCH-NN
MAE	Best	4	2	3	3
	Worst	1	5	3	3
MSE	Best	4	0	5	3
	Worst	4	2	2	4
R^2	Best	1	7	4	0
	Worst	6	1	0	5
QLIKE	Best	4	1	4	3
	Worst	2	4	3	3
R^2LOG	Best	4	1	4	3
	Worst	2	3	4	3
Total	Best	17	11	20	12
	Worst	15	15	12	18

Table 3.6.4 Number of models that belong to SSM					
Models/Loss function	MAE	MSE	R2LOG	QLIKE	Total
GARCH	4	5	3	5	17
EGARCH	2	4	6	6	18
TGARCH	3	4	3	5	15
CGARCH	6	6	5	7	24
GARCH-NN	5	9	5	5	24
EGARCH-NN	3	8	3	4	18
TGARCH-NN	4	10	4	5	23
CGARCH-NN	2	2	5	5	14
HAR	4	4	4	7	19

4.Emprical chapter of covariance Forecasting

Covariance forecasting using DCC-GARCH and a hybrid-built Artificial Neural Networks

Abstract

This chapter introduced a hybrid model combining DCC process and Artificial Neural Networks in order to promote the forecasting ability of the traditional approaches when forecasting covariance or correlation. The covariance between the S&P 500 stock index and other eleven stock indices from different countries will be investigated by using four standard DCC GARCH models and four hybrid models built with neural networks. A wide comparison by four measures among these models is created to check the forecasting performance of these models. Three findings are addressed by this chapter. First, the proposed model could fit the covariance matrix well and give a simple way to deal with the high-dimension problem that may occur in forecasting covariance by traditional GARCH model. Second, the results revealed that the new hybrid method outperforms the traditional DCC GARCH models. Finally, the EGARCH DCC process built on neural networks has the best forecasting performance within the whole comparison technique. The results are able to provide some suggestions for market managers on risk control, especially for the portfolios containing multivariate assets in different countries.

4.1 Introduction

Covariance forecasting has been a hot topic in risk management. Modeling and forecasting the covariance matrix of financial assets are essential for portfolio allocation and risk management. Lots of papers and research focused on the traditional models like VEC, BEKK, CCC/DCC process in the past. With the development of technology these days, many studies appear to concentrate on machine learning approaches since it will not be restricted to the heavy parameter or high dimension estimation. Most of them used these approaches individually. Moreover, it is hard to tell which model or approaches are the better one. In this chapter, the study aims to explore the covariance forecasting by using traditional DCC-GARCH models and a hybrid-built model based on neural networks.

In this chapter, a new approach which combines neural networks and the DCC GARCH model will be introduced. A wide comparison among the different approaches will be created. The forecasting power will be explored by using both DCC GARCH methods and several hybrid-built models based on neural networks. Four measures of comparison including Mean Absolute Error (MAE), Mean Squared Error (MSE), Quasi-Likelihood (QLIKE) and R^2LOG loss function will be used to discover the forecast ability. The Model Confidence Set (MCS) will be created at the end as well to find a better performing model.

There are three main contributions in this chapter: First, a approach built with DCC GARCH genres and neural networks is introduced in this chapter. The forecast ability of the new method has been investigated. Second, since there does not exist a general accepted conclusion to state the best models or best methods in the covariance forecasting, this chapter will go into the discussion of the covariance forecasting with the application of several models including both the traditional DCC GARCH model and the new hybrid approach methods in order to identify a better model for covariance forecasting. Most of today's study are talking about the traditional multivariate GARCH or machine learning methods individually when forecasting covariance. This chapter aimed to provide a more comprehensive view of the covariance forecasting performance of traditional DCC GARCH, and hybrid approaches built with neural networks. Furthermore, this exercise aims to introduce a new approach which is more efficient to forecast the covariance. Four different GARCH models including GARCH, EGARCH, TGARCH and CGARCH are used in the first step of the DCC process in order to identify whether different models used in first step will affect the forecast accuracy of the covariance in the results. Likewise, since some of the study only considered about the forecasting performance among very limited samples by only one or two models, this chapter aims to make a wider comparison among 12 stock indices from different countries with the application of 8 models.

The structure of the chapter is organized as follows. Section 2 summarizes the background of the covariance forecasting and some main questions which have been investigated in this area. Section 3 gives the data series, source, time periods. Section 4 describes the list of covariance forecasting models, and a hybrid-built model is proposed as well. Section 5 shows the method used to compare the performance of the models. In section 6, a full report of the performance will be displayed, and section 7 will discuss the conclusions and findings.

4.2 Background

In financial literature, it is important to consider the volatility of an asset since it is linked with the "risk" directly. However, it is not enough to only observe a univariate asset since the behaviors among different assets still play an important role when estimating the conditional distribution of returns which is confirmed by the work of Bollerslev, Engle and Wooldridge (1988). There are several methods to estimate or forecast the covariance including historical methods, MA, EWMA and a series of multivariate GARCH (MGARCH) models. Since a large number of univariate GARCH models are proposed, it is not hard to make some extension to them to multivariate ones for the measurement of the covariance. The multivariate GARCH are divided into four categories by Silvennoinen & Terasvirta (2008) including the models calculated the conditional covariance directly like the VEC and BEKK models; the factor models which assume the return process r_t is generated by a (small) number of unobserved heteroskedastic factors; the semiparametric and nonparametric approaches that can offset the loss of efficiency of the parametric estimators which do not impose a particular structure (that can be mis-specified) on the conditional covariance matrices and finally the models of conditional variances and correlations which built on the idea of modelling the conditional variances and correlations instead of straightforward modelling of the conditional covariance matrix like the Constant Conditional Correlation (CCC) model, Dynamic Conditional Correlation (DCC) model which will be introduced in following part. Bollerslev et al. (1988) first introduced a vector error correction model (VEC) which is a multivariate generalized autoregressive conditional heteroscedastic process. They found that the conditional covariances are quite variable over time and are a significant determinant of time-varying risk premia. With the later work by Engle & Kroner (1995), another direct extension of univariate GARCH model called BEKK which named after Baba, Engle, Kraft and Kroner were proposed. They built a new parameterization of the multivariate ARCH process which gives an effective way for the estimation of covariance between different assets. However, these models are usually heavily parameterized and may only be suitable for a small number of series. Some other alternative models are proposed such as the orthogonal GARCH model. It is a n-dimension GARCH model based upon the principal components of a constant unconditional covariance matrix which is able to reduce the dimension of the covariance matrix by (Ding & Engle (2008) and Klaassen (2002). Van der Weide (2002) also proposed a new type of multivariate GARCH model in which potentially large covariance matrices can be parameterized with a fairly large degree of freedom while estimation of the parameters remains feasible.

Furthermore, study took more attention to the conditional correlation models since it can be used in a more widely environment by its muti-step procedure. The constant conditional correlation model (CCC) was first introduced by Bollerslev (1990) which parameterizes each of the conditional variances as a univariate GARCH process. Another model called the standard dynamic conditional correlation (DCC) was proposed by Engle (2002) which allows the dynamic conditional correlation. Engle states that *"The DCC process have the flexibility of univariate GARCH coupled with parsimonious parametric models for the correlations."* Since the process can be estimated with univariate or two-step methods by likelihood function, it has been proved to perform well in lots of empirical situations by the work of Engle. Cappiello et al. (2003) introduced a new generalization of the standard DCC model with Engle called asymmetric generalized dynamic conditional correlation (AG-DCC) model which allows for series-specific news impact and smoothing parameters. Conditional asymmetries in correlation dynamics are also permitted as well. It gives a suitable method to examine the dynamics of correlation between different type of assets.

However, by the work of Aielli (2011), they proved that DCC large system estimator can be inconsistent which may lead to misleading conclusions by the traditional interpretation of the DCC correlation parameters. Therefore, they suggested a more tractable dynamic conditional correlation model called cDCC-GARCH models which the large system estimator are proven to be consistent. Both DCC and cDCC were tested in two different datasets, one is a small dataset with 10 equity indices and the other is a larger dataset with 100 equity indices. The results show that the cDCC correlation forecasts perform as well as or significantly better than the DCC correlation forecasts.

Since varieties of models are introduced to measure or forecast the covariance or correlation, increasing attention has been devoted to the comparison of different models. There still existed a debate on which model performs better. Caporin & McAleer (2008) examined the forecasting performance using a set of multivariate GARCH models including BEKK, DCC, cDCC, CCC of Bollerslev (1990), Exponentially Weighted Moving Average, and covariance shrinking of Ledoit & Wolf (2004). The results show that many models have similar forecasting performance including both direct and indirect evaluation methods. They cannot give a best performer since the preferred models is not stable over alternative approaches and sample periods. A later empirical study by Huang et al. (2010) made a comparison among BEKK GARCH and DCC GARCH Models using the dataset of Euro zero-coupon bonds with different maturities. They found that the DCC model performs better by its simple two-step estimation method and give a conclusion that "*significant fitting and forecasting performances originate from the trade-off between parsimony and flexibility of the MGARCH models.*" Basher & Sadorsky (2016) use 3 methods including DCC, ADCC and GO-GARCH to model volatilities and conditional correlations between emerging market stock prices, oil prices, VIX, gold prices and bond prices. The results are different between different assets. The ADCC models are preferred for emerging market stock prices with oil, VIX, or bonds, while the GO-GARCH are more effective for emerging market stock prices with gold in some instances.

In the past twenty years, the study of the correlation or covariance using different models among financial assets has never been stopped. Several new models and new approaches are proposed. Lots of comparisons among models are created. Fiszeder et al. (2019) suggest an extension of DCC model that incorporates high and low prices into the DCC framework. They made an empirical study on currencies, stocks, and commodity exchange traded funds with the application of their new models. The new models have proved to be superior to the standard DCC model and alternative range-based DCC model. Besides the traditional parametric models, the machine learning methods also appeared to be an effective way. Since the traditional GARCH has limitations to deal with the high-dimensional cases, more and more study are focusing on machine learning methods. The neural networks, which were introduced in the previous chapter, are the one of the most widely used techniques when making forecasts among time series. Cai et al., (2012) forecasted a large-scale conditional volatility and covariance using neural network. Their experiment results indicate that the neural networks are better in modeling for large-scale cases rather than traditional GARCH models. Fang et a. (2021) proposed a neural networks methodology to forecast the realized covariance matrix which are able to handle high-dimensional realized covariance matrices consistently. The machine was trained with the historical realized covariance matrices with the application of a nonlinear mapping. Their results proved that the method they proposed has more excellent forecasting ability than other several traditional advanced volatility models. Liu et al., (2021) made a comparison between the multivariate dynamic covariance GARCH models and the artificial neural networks. They measure the portfolio optimization and the distances between the actual covariance matrices and predicted covariance matrices by using two different models and found that the neural networks outperforms the DCC model in terms of mean portfolio returns. Support vector regression, as another effective way for time-series forecasting, is able to capture the dynamics of the financial processes and has an overall stable performance. Fałdziński et al. (2020) introduced a methodology using support vector regression (SVR) to forecast the covariance matrices. The results from the new method were proved to be more accurate than the forecasts from the benchmark dynamic conditional correlation model.

Although there exist lots of methods to investigate the dynamics of the covariance, it is hard to tell which method has a better performance. Most of the research talks about a single method like a traditional DCC process, a Neural Network or a Support vector machine and make comparisons among them between very limited assets. This study proposed a new approach which combined the traditional DCC process and neural networks. A wider comparison will be created as well to test the forecasting ability of the hybrid machine and traditional DCC process.

4.3 Data

The target of this exercise is to check the forecasting performance of several models which the main work is to calculate the conditional covariance. The main data set being used in this paper is the daily stock price index in twelve countries over the world. The countries which are selected are based on the principle of market capitalization in the stock market, particularly, countries with over $1500 billons market capitalization which is presented by the end of year 2017 will be included. The time period was selected as same as the previous chapter due to the data access availability and the convenience for comparison across different chapter in the future. The market capitalization represents the total value of a company's stock, the stock with a high market capitalization means the size of the stock market is large. With the application of market capitalization, it allows investors to size up a stock index. The thesis aims to select the stock index with large size to test the forecast ability of different models. Therefore, the twelve daily stock price indexes are selected. The markets of selected countries in Europe, Australia and America are all developed market since the market in these countries have already experienced a long history of development while the emerging market are mainly located in the Asian area including 3 markets: China, Korea and India. The data comes from different type markets will give a comprehensive result that whether the performance of the methods is suitable or effective when the market changes.

The daily adjusted closing price is selected, the time period is from 5 January 2005 which is the first available trading date in year 2005 to 28 December 2018 which is the final available trading date in year 2018. In order to train the machine (neural networks) with more data, the long time period of data is selected. Since some of the daily stock price data are not comprehensive before 2005, due to the data availability, this time period was selected. The prices of the index are transformed in returns by standard methods which make it more measurable in equation. All the price data are obtained by the Application Programming Interface (API) finance data from "Yahoo Finance"

The returns are calculated by the standard method (4.3.1). In order to compare the performance of different methods, the realized covariance of these returns which act as the "true covariance/ correlation" is calculated by the method (4.3.2)/ (4.3.3). The corresponding high frequency returns are obtained from the source of "Oxford Man", the source provides the 5-mins intraday returns of different assets and the covariance were calculated with the method of formula (4.3.2) and (4.3.3).

The index prices are transformed in returns by standard methods which make it more measurable in equation:

$$R_t = \frac{S_t - S_{t-1}}{S_{t-1}} * 100\% \tag{4.3.1}$$

The daily realized covariance between two assets computed from intraday returns recorded at m subintervals each day and equals to the sum of cross products of the intraday returns by equation:

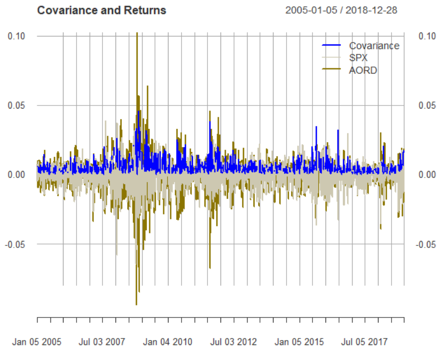
$$RCov_{12,t}^{m} = \sum_{j=1}^{m} R_{1,t-1+j/m} R_{2,t-1+j/m} \qquad (4.3.2)$$

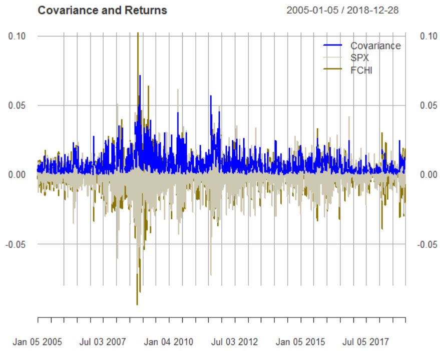
Further, the realized correlation can be obtained by:

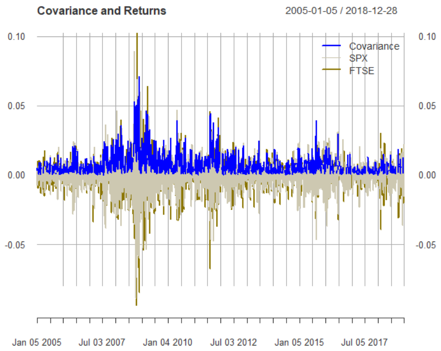
$$RCorr_{12,t}^{m} = \frac{RCOV_{12,t}^{m}}{\sqrt{RV_{1,t}^{m} RV_{2,t}^{m}}} \qquad (4.3.3)$$

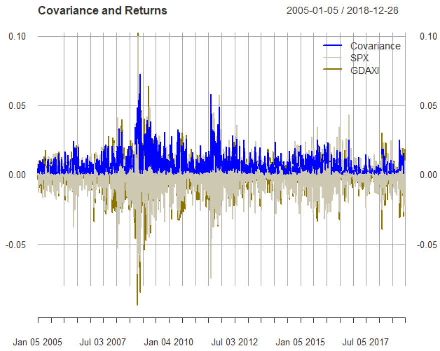
where $RV_{1,t}^{m}$ and $RV_{2,t}^{m}$ are the daily realized variance of two assets which has been introduced in **Chapter 3**.

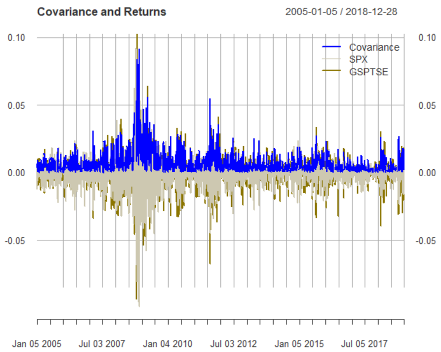
The returns and realized covariance between two indices were plotted for a basic view of the data. Since the space is limited, it is impossible to plot all the covariance and returns in pairs. Here only the covariance between USA and other countries were presented. It can be seen that the covariance or movements between USA and other countries tends to be similar, most of high covariance tends to appear during the period of year 2007 and 2008 in all the samples as well as the amplitude of return changes. This should be referred to the global financial crisis happening in 2008.

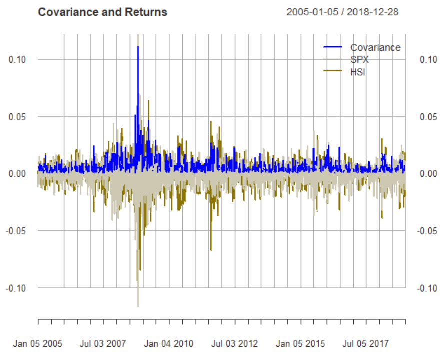

Covariance and Returns
2005-01-05 / 2018-12-28
Covariance
SPX
AORD
0.10
0.05
0.00
-0.05
Jan 05 2005
Jul 03 2007
Jan 04 2010
Jul 03 2012
Jan 05 2015
Jul 05 2017

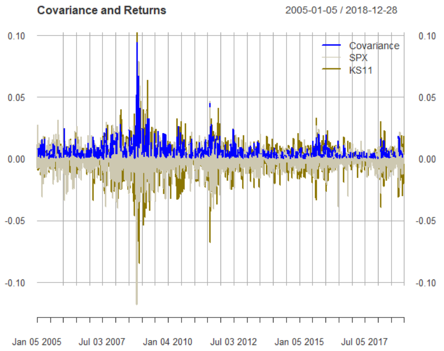

Covariance and Returns
2005-01-05 / 2018-12-28
Covariance
SPX
FCHI
0.10
0.05
0.00
-0.05
Jan 05 2005
Jul 03 2007
Jan 04 2010
Jul 03 2012
Jan 05 2015
Jul 05 2017

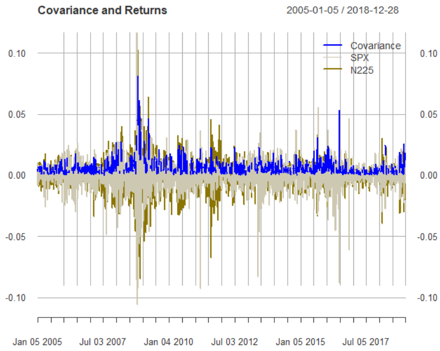

Covariance and Returns
2005-01-05 / 2018-12-28
Covariance
SPX
FTSE
0.10
0.05
0.00
-0.05
Jan 05 2005
Jul 03 2007
Jan 04 2010
Jul 03 2012
Jan 05 2015
Jul 05 2017

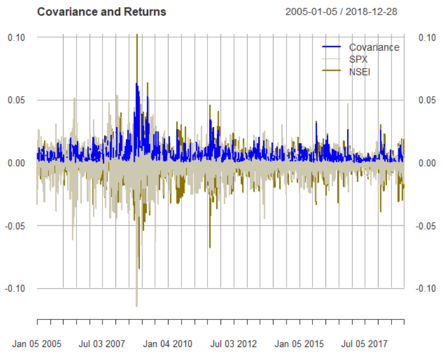

Covariance and Returns
2005-01-05 / 2018-12-28
Covariance
SPX
GDAXI
0.10
0.05
0.00
-0.05
Jan 05 2005
Jul 03 2007
Jan 04 2010
Jul 03 2012
Jan 05 2015
Jul 05 2017

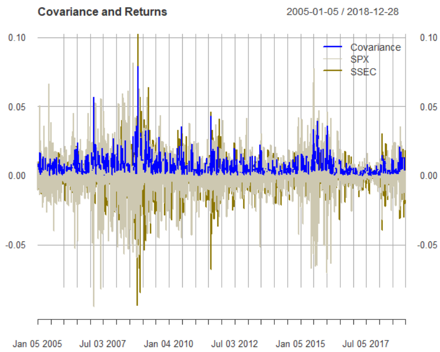

Covariance and Returns
2005-01-05 / 2018-12-28
Covariance
SPX
GSPTSE
0.10
0.05
0.00
-0.05
Jan 05 2005
Jul 03 2007
Jan 04 2010
Jul 03 2012
Jan 05 2015
Jul 05 2017


Covariance and Returns
2005-01-05 / 2018-12-28
Covariance
SPX
HSI
0.10
0.05
0.00
-0.05
-0.10
Jan 05 2005
Jul 03 2007
Jan 04 2010
Jul 03 2012
Jan 05 2015
Jul 05 2017


Covariance and Returns
2005-01-05 / 2018-12-28
Covariance
SPX
KS11
Jan 05 2005
Jul 03 2007
Jan 04 2010
Jul 03 2012
Jan 05 2015
Jul 05 2017


Covariance and Returns
2005-01-05 / 2018-12-28
Covariance
SPX
N225
Jan 05 2005
Jul 03 2007
Jan 04 2010
Jul 03 2012
Jan 05 2015
Jul 05 2017


Covariance and Returns
2005-01-05 / 2018-12-28
Covariance
SPX
NSEI
Jan 05 2005
Jul 03 2007
Jan 04 2010
Jul 03 2012
Jan 05 2015
Jul 05 2017


Covariance and Returns
2005-01-05 / 2018-12-28
Covariance
SPX
SSEC
Jan 05 2005
Jul 03 2007
Jan 04 2010
Jul 03 2012
Jan 05 2015
Jul 05 2017

4.4 Methodology

Modelling

In order to make a clear explanation about the methods to be used, a multivariate returns process $\boldsymbol{r_t}$ is given by:

$$\boldsymbol{r_t} = \boldsymbol{\mu_t} + \boldsymbol{\varepsilon_t} \qquad (4.4.1)$$

where $\boldsymbol{\mu_t}$ is a *n* × *1* vector of the mean process of different assets. Similarly, the *n* × *1* vector $\boldsymbol{\varepsilon_t}$ which is the error term can be written as:

$$\boldsymbol{\varepsilon_t} = \boldsymbol{H_t^{1/2} z_t} \qquad (4.4.2)$$

where $\boldsymbol{z_t}$ is *n* × *1* vector of independent and identically distributed (i.i.d) random variables with a constant mean zero and variance. H_t is a *n* × *n* matrix of conditional variances to be estimated and forecast.

The DCC model

The DCC model was introduced by Engle & Sheppard (2001). The idea of the models in this class is that the covariance matrix $\boldsymbol{H_t}$, can be decomposed into conditional standard deviations $\boldsymbol{D_t}$ and a correlation matrix R_t.In the DCC-GARCH model both $\boldsymbol{D_t}$ and R_t are designed to be time-varying.

$$\boldsymbol{r_t} = \boldsymbol{\mu_t} + \boldsymbol{\varepsilon_t} \tag{4.4.3}$$

$$\boldsymbol{\varepsilon_t} = \boldsymbol{H_t^{1/2} z_t} \tag{4.4.4}$$

$$\boldsymbol{H_t}=\boldsymbol{D_t R_t D_t} \tag{4.4.5}$$

where $\boldsymbol{D_t} = diag(\sqrt{h_{i,t}})$ is a $n \times n$ diagonal matrix of conditional volatility of $\boldsymbol{\varepsilon_t}$ which can be obtained from a univariate GARCH model $\boldsymbol{R_t}$ is the $n \times n$ conditional correlation matrix which can be time varying.

The conditional correlation $\boldsymbol{R_t}$ is a matrix of the standardized disturbance ϵ_t which can be specified as:

$$\epsilon_t = D_t^{-1}\boldsymbol{r_t} \sim \boldsymbol{N(0, R_t)} \tag{4.4.6}$$

Obviously, $\boldsymbol{R_t}$ is a symmetric matrix which means that $\rho_{ij,t} = \rho_{ji,t}$ and $\rho_{ii,t} = 1$.

A simple $\boldsymbol{R_t}$ with two assets can be specified as:

$$\begin{bmatrix} 1 & \rho_{12,t} \\ \rho_{21,t} & 1 \end{bmatrix} \tag{4.4.7}$$

The covariance matrix $\boldsymbol{H_t}$ should be positive definite which means that $\boldsymbol{R_t}$ should be positive definite as well since $\boldsymbol{D_t} = diag(\sqrt{h_{i,t}})$ is positive definite and all the elements in the correlation matrix $\boldsymbol{R_t}$ need to be equal to or less than one by correlation definition. Therefore $\boldsymbol{R_t}$ is decomposed as:

$$R_t = Q_t^{*-1} Q_t Q_t^{*-1} \tag{4.4.8}$$

$$Q_t = (1 - a - b)\bar{Q} + a\epsilon_{t-1}\epsilon_{t-1}^T + bQ_{t-1} \tag{4.4.9}$$

where $\bar{Q} = cov(\epsilon_t \epsilon_t^T)$ is the unconditional covariance matrix of ϵ_t and can be estimated by:

$$\bar{Q} = cov(\epsilon_t \epsilon_t^T) = \frac{1}{K}\sum_{t=1}^{K} \epsilon_{t-1}\epsilon_{t-1}^T \tag{4.4.10}$$

Q_t^* is a diagonal matrix with the square root of the diagonal elements of Q_t

$$Q_t^* = diag(\sqrt{q_{ii,t}}) \tag{4.4.11}$$

Therefore, $|\rho_{ij}| = \left|\frac{q_{ij,t}}{\sqrt{q_{ii,t} q_{jj,t}}}\right| \leq 1$ are satisfied by the rescaling of Q_t^*. Further, the scalars a and b should satisfy: $a \geq 0, b \geq 0$ and $a + b < 1$ to make sure the requirement of positive definite.

The one-step ahead forecasting of the conditional covariance $\boldsymbol{H_{t+1}}$=$\boldsymbol{D_{t+1} R_{t+1} D_{t+1}}$ when the history up to time t is known can be done by two steps, particularly, the forecasts of $\boldsymbol{D_{t+1}}$ and $\boldsymbol{R_{t+1}}$ can be done separately.

Step 1: Forecasting the conditional variances in $\boldsymbol{D_{t+1}}$

As mentioned above the $\boldsymbol{D_{t+1}} = diag(\sqrt{h_{i,t+1}})$ is a $n \times n$ diagonal matrix of conditional volatility of $\boldsymbol{\varepsilon_{t+1}}$. The forecast of $h_{i,t+1}$ which means the conditional variance of asset i can be done by the univariate GARCH models. Moreover, the specification of the univariate GARCH models is not limited to the standard GARCH. Therefore, the varieties of GARCH specification which were introduced in **Chapter 3** could be used for the forecasting of $h_{i,t+1}$

Standard *Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models*

$$h_{i,t+1} = \omega + \alpha\varepsilon_{i,t}^2 + \beta h_{i,t} \tag{4.4.12}$$

Exponential GARCH (EGARCH) models

$$ln(h_{i,t+1}) = \omega + \beta ln(h_{i,t}) + \alpha\varphi\left(\frac{\varepsilon_{i,t}}{\sqrt{h_{i,t}}}\right) + \alpha\psi\left|\left(\frac{\varepsilon_{i,t}}{\sqrt{h_{i,t}}}\right)\right| \tag{4.4.13}$$

Threshold-GARCH (TGARCH) models

$$h_{i,t+1} = \omega + \alpha\varepsilon_{i,t}^2 + \delta I_t\varepsilon_{i,t}^2 + +\beta h_{i,t} \tag{4.4.14}$$

$$I_t = \begin{cases} 1 & \varepsilon_{i,t} < 0 \\ 0 & \varepsilon_{i,t} > 0 \end{cases} \tag{4.4.15}$$

Component GARCH (CGARCH)

$$h_{i,t+1} = g_{i,t+1} + \alpha\left(\varepsilon_{i,t}^2 - g_{i,t}\right) + \beta\left(h_{i,t} - g_{i,t}\right) \tag{4.4.16}$$

$$g_{i,t+1} = \omega + \gamma g_{i,t} + \varphi\left(\varepsilon_{i,t}^2 - h_{i,t}\right) \tag{4.4.17}$$

Step two: forecasting the conditional correlation matrix $\boldsymbol{R_{t+1}}$:

The forecasting of $\boldsymbol{R_{t+1}}$ can be transferred to the calculation of Q_t and $\bar{Q}$ which is mentioned above. Therefore, the forecasts of $\boldsymbol{R_{t+1}}$ can be specified as:

$$R_{t+1} = Q_{t+1}^{*-1} Q_{t+1} Q_{t+1}^{*-1} \tag{4.4.18}$$

$$Q_{t+1} = (1 - a - b)\bar{Q} + a\epsilon_t\epsilon_t^T + bQ_t \tag{4.4.19}$$

The neural networks built on DCC GARCH model

With the hope to improve the forecasting performance of the multivariate GARCH model, a hybrid model based on neural networks are introduced. An artificial neural network (ANN) is a network of interconnected elements called neurons. The neurons are used to estimate functions based on the inputs. The neurons are connected with each other by joint mechanism which consists of a set of assigned weights. The method can be described as follows:

$$Cov_t = \varphi\left(\sum_1^{t-1} \omega_p x_i + b_t\right) \tag{4.4.20}$$

$\{x_i\}$ is the input data which is the variance and covariance, and $\{\omega_p\}$ describes the connection weights of neurons., b_t is the bias, $\varphi(\cdot)$ is the activation function and Cov_t is the output of the neuron which is the forecasting covariance. In ANN works, multi-layer feed forward (MLP) is a common approach which has three layers: input layer, output layer, and hidden layer. Neuron takes the values of inputs parameters, sums them up with the assigned weights, and adds a bias. With the application of transfer function, the outputs will be displayed. The algorithm can be viewed directly in Figure 4.4.1.

With the conception of neural networks, a hybrid multivariate GARCH model can be built, particularly here a hybrid DCC GARCH model. In neural networks, the input data can be set as an explanatory variable of financial time series, such as returns, squared returns, trading volumes, etc. Since the target of the DCC GARCH model is to forecast the covariance, the input data will be set as the variance generated by univariate GARCH models including the standard GARCH, EGARCH, TGARCH and CGARCH mentioned above and the correlation as well. It will take two steps when forecasting the covariance.

Step 1: Train the machine:

A training set which includes the train variances of two assets and the correlation will be entered into the neural networks and to calculate the train covariance. An error will be calculated between the train covariance and the "true" covariance and run back to the system until the error becomes convergence. The whole dataset is separated into two parts: one is from 5 January 2005 to 25 November 2014 (2400 observations or approximate 10 years), it is the training set for the machine. The other is from 26 November 2014 to 28 December 2018 (1000 observations or approximate 4 years), it is the part to check the 1 day rolling window forecasting performance of the machine.

Step 2: Forecast the covariance:

A testing set which includes the test variances of two assets and the correlation will be entered into the neural networks and to calculate the test covariance which here is the forecasting values. Since the machine is trained in step one, it has formulated a function to deal with the testing set and been able to output the forecasting values.

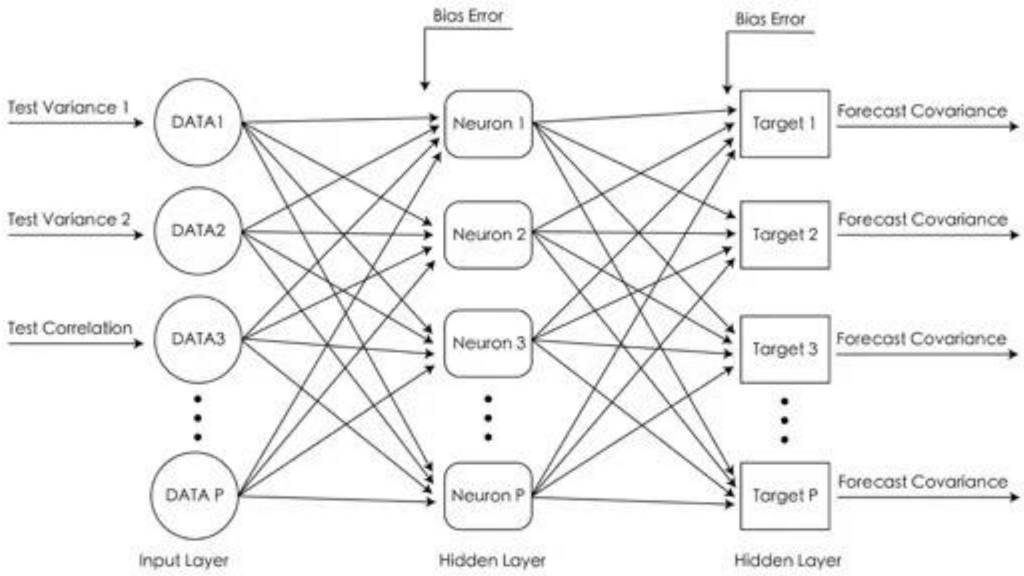


Figure 4.4.1 Hybrid model demonstration

4.5 Comparison techniques for forecast performance

The exercise will mainly focus on the covariance forecasting between USA and other countries since the market value of USA are the highest in the dataset. More analysis could be done by repeating the same procedure in the software and will be added into the appendix. As introduced in chapter 3, the Mean Absolute Error (MAE) and Mean squared error (MSE) are selected. The Model Confidence Set (MCS) procedure by Hansen (2011) are selected as well to test the equal predictive ability (EPA) at certain confidence level α depending on the loss functions including MAE, MSE, QLIKE and R^2LOG. To retrieve a "actual" covariance, the realized covariance $\widetilde{Cov}_{12,t+1}$ (calculated by the product of 5 mins intraday returns with two selected countries) is considered both the in-sample and out-of-sample period, which are believed to be closer to the unobservable covariance. The reason to use the realized covariance is that the conditional covariance is latent and unobservable directly, the realized covariance provides a measure of the historical performance among two different assets, here the stock indices, which implies that the fluctuations of one asset will have effect on the other.

The forecasting values are specified as $\widehat{Cov}_{12,t+1}$. Since the covariance of the two selected assets are a symmetric matrix as: $\begin{bmatrix} Var_1 & Cov_{12} \\ Cov_{12} & Var_2 \end{bmatrix}$, the Cov_{12} will be retrieved into a series of time series to make the comparison with the realized covariance.

Mean Absolute Error (MAE) and Mean Squared Error (MSE)

The MAE calculates the average difference of the comparison with equal weight of all individual differences. The MSE calculates the average squared difference between the estimated values and the actual values. The comparison with covariance is similar with volatility. Both comparison measures report the better performance by a lower statistic.

$$MAE = \frac{1}{\delta}\sum_{t=T+1}^{T+\delta}\left|\widetilde{Cov}_{12,t} - \widehat{Cov}_{12,t+1}\right| \quad (4.5.1)$$

$$MSE = \frac{1}{\delta}\sum_{t=T+1}^{T+\delta}\left(\widetilde{Cov}_{12,t} - \widehat{Cov}_{12,t+1}\right)^2 \quad (4.5.2)$$

Quasi-Likelihood (QLIKE) and R^2LOG loss function

The Quasi-Likelihood (QLIKE) loss function is a test of forecast bias implied by a Gaussian likelihood which has a specification:

$$QLIKE_{t+1} = log\left(\widehat{Cov}_{12,t+1}\right) + \frac{\widehat{Cov}_{12,t+1}}{\widetilde{Cov}_{12,t+1}} \quad (4.5.3)$$

The other loss function is the R^2LOG which measure the goodness of fit of the out-of-sample forecasts.

$$R^2LOG_{t+1} = \left[log\left(\frac{\widehat{Cov}_{12,t+1}}{\widetilde{Cov}_{12,t+1}}\right)\right]^2 \quad (4.5.4)$$

Model Confidence Set (MCS)

The statistical method called Model Confidence Set (MCS) are selected to deliver the best performing models with respect to a given criterium since it is hard to point a single model outperforms others by the reason of their statistically equivalent property or not enough information from the data. The MCS procedure will eliminate the worst model at each step if the null hypothesis of equal predictive ability (EPA) is not accepted and the procedure will iterate until all models with EPA hypothesis entered in the “Superior Set Models” (SSM). An optimal SSM set will be created to make a clear view of the selected models.

The reason to choose MAE and MSE is that MAE is useful for consistent error measurement across all data points, while MSE is preferred when penalizing significant errors is crucial. However, the MSE method penalizes symmetrically. Therefore, QLIKE function which Sturesson & Wennström, (2023) states that is an asymmetrical loss function that penalizes under-prediction heavier than over-prediction is selected. The R^2LOG is selected as well since it is often easier to interpret since it doesn't depend on the scale of the data.

4.6 Results and analysis of empirical exercise

All the data and models and comparisons are applied with R studio in an R language environment.

Since the covariance matrix among 12 countries will be a huge matrix (12*12), and most of the procedure is iteration process, which is easy to repeat, the exercise takes the results of forecasting covariance between the stock index of USA and other 11 countries in order to make a tidy and direct report. With the application of 8 different models, the forecasting covariance were compared with the realized covariance which is calculated by the product of intra-day returns of two countries mentioned above. The hybrid models based on neural networks are addressed with "NN" to make it clear in the table. The different GARCH model recorded in the results correspond to the different GARCH type models in the "step one" of DCC estimation.

Table 4.5.1 gives the MAE statistics of the traditional DCC GARCH process and the hybrid models. From the table, the neural networks built on TGARCH DCC model has a better performance than others in 5 out of 11 sample countries. The second is also the hybrid model GARCH DCC-NN in 3 countries followed by the EGARCH DCC in 2 countries and CGARCH DCCNN in 1 country.

The second-best performing models can be found by the second minimum MAE values reported in each row. The neural networks built on EGARCH DCC-NN model appears to be the first with a second minimum MAE value in 5 countries. The rest are TGARCHDCC-NN in 3 countries, TGARCHDCC in 2 countries, and CGARCHDCC-NN in 1 country.

The third-best performing models can be observed by the third minimum MAE values in each row. The results are in a wide dispersion. The GARCH DCC-NN, takes the first in 7 countries followed by CGARCH DCC-NN in 2 countries. The rest are EGARCH DCC-NN, TGARCH DCC-NN in 1 country, respectively.

Finally, Table 4.5.1 also reported the worst performing models of covariance forecasting by considering the maximum MAE statistics in each row. Here, the CGARCH DCC appears to be the worst performer with a maximum MAE value in 8 countries.

Looking at the whole performance of the selected models, the best models can be identified as the neural networks built on asymmetric models since the best and second-best performer are recorded as TGARCH DCC-NN and EGARCH DCC-NN, respectively. (Both of them gives 5 minimum and second-minimum MAE values). Moreover, the third-best performer are also recorded as the neural networks built on GARCH DCC models. The EGARCH DCC also have a good performance with 2 best cases, but the CGARCH DCC appears to be the worst performer for 8 times. Therefore, An overall conclusion can be drawn that the hybrid model based on neural networks outperforms the original models by using MAE comparison techniques.

Since the MAE methods are the simplest way like an “naïve method” to compare the error which takes the difference with an equal weight, the MSE measure is carried out to identify if there exist some outlier predictions with huge errors.

Similarly, Table 4.5.2 reports the MSE statistics for the covariance forecasting results between USA and other 11 countries by using different models. The MSE will puts larger weight on the errors due to the squaring part of the function which will enlarge the outlier prediction error.

Like using MAE statistics, the TGARCH DCC-NN model comes to be the first with a minimum MSE value in 7 out of 11 countries. The rest are EGARCH DCC and CGARCH DCC-NN which both report 2 minimum MSE value, respectively.

The second-best model is the neural networks built on GARCH DCC and EGARCH DCC models. They both reported a second minimum MSE value in 5 countries each, while the TGARCH DCC-NN appeared to be second performer in 1 country.

The third-best model is recorded as GARCH DCC-NN and EGARCH DCC-NN in 4 out of 11 countries, respectively. The rest are CGARCH DCC-NN in 2 countries and TGARCH DCC in 1 country.

Unlike the results with using MAE statistics, the standard GARCH DCC comes to be the worst performing model with the highest MSE statistics in 5 out of 11 sample countries. Moreover, the CGARCH DCC was identified as the worst performer 3 times as well. The hybrid model TGARCH DCC-NN also performed the worst in 2 countries.

By taking a whole view of Table 4.5.2, the conclusion is similar with the measure MAE. The hybrid model TGARCH DCC-NN again comes to be the best performing model although it has become the worst performer 2 times. The CGARCH DCC-NN and EGARCH DCC has a good performance as well.

Loss function of the covariance

Model Confidence Set procedure (MCS) is an iteration of a sequence of statistic tests which permits to construct a set of "superior" models, the "Superior Set Models" (SSM), under the null hypothesis of equal predictive ability (EPA) is not rejected at certain confidence level α based on a loss function. It is introduced in chapter 3 and the MCS are applied as well to check the predictive ability. A direct comparison based on the loss function QLIKE and R^2LOG are taken out. The details of the "Superior Set Models" (SSM) will be displayed as a conclusion in next section.

The QLIKE and R^2LOG are selected to check if there is any significant difference. The loss function specification for covariance is specified as:

$$QLIKE_{t+1} = log\left(\widehat{Cov}_{12,t+1}\right) + \frac{\widehat{Cov}_{12,t+1}}{\widetilde{Cov}_{12,t+1}} \tag{4.6.1}$$

$$R^2LOG_{t+1} = \left[log\left(\frac{\widehat{Cov}_{12,t+1}}{\widetilde{Cov}_{12,t+1}}\right)\right]^2 \tag{4.6.2}$$

where $\widetilde{Cov}_{12,t+1}$ refers to the "actual" covariance and evaluated covariance is specified as $\widehat{Cov}_{12,t+1}$.

The R^2LOG statistics prefers a value closer to zero which indicates that the distance between the "actual" covariance and forecasting covariance is small enough. The value is no doubt nonnegative due to the squaring part of the function so that a lower R^2LOG statistics is preferred when comparing different models.

With the application of the QLIKE loss function in equation (4.6.1), it is hard to say that a lower statistic indicates a better performance. In order to make it more appropriate to compare directly, a specification is defined as:

$$QLIKE_{t+1} = \left(log\left(\widetilde{Cov}_{12,t+1}\right) - log\left(\widehat{Cov}_{12,t+1}\right)\right) + \left(\frac{\widehat{Cov}_{12,t+1}}{\widetilde{Cov}_{12,t+1}} - 1\right) \tag{4.6.3}$$

QLIKE statistic measures the log difference between the “actual” covariance and forecasting covariance and their rates. The two parts tend to be zero if the forecasting values tend to be same as the “actual” value so that a QLIKE statistic closer to zero will be preferred when comparing different models.

In Table 4.5.3 the QLIKE statistics based on equation (4.6.3) are reported.

The TGARCH DCC-NN comes to be the first with a minimum QLIKE value reported in 6 out of 11 countries followed by ERCH DCC-NN in 3countries. There is 1 minimum QLIKE value reported by EGARCH DCC and GARCH DCC-NN, respectively.

The second-best performing models can be found by the second minimum QLIKE statistics reported in each row. The EGARCH DCC-NN appears to be the first in 5 country samples. The rest are GARCH DCCNN in 4 countries and TGARCH DCC and TGARCH DCC-NN with a minimum QLIKE values in 1 country each.

The third-best performing models can be observed by the third minimum QLIKE values in each row. The GARCH DCC-NN are reported to be the first with a third minimum QLIKE in

6 countries followed by EGARCH DCC-NN in 3 countries. There are 2 third minimum QLIKE value reported by TGARCH DCC-NN as well.

Finally, Table 4.5.3 also reported the worst performing models of the covariance forecasting by considering the maximum QLIKE statistics in each row. The CGARCH appears to be the worst performer with a maximum QLIKE value in 7 countries.

The overall conclusion of Table 4.5.4 generated by the whole performance of the selected models are similar with the measures applied above. The hybrid GARCH models perform better than others. In more detail, the TGARCH DCC-NN appears to be the first, the EGARCH DCC-NN second and the GARCH DCC-NN third. Likewise, with the application of QLIKE measure, these hybrid models appear to be superior since all the best performers are located in the neural networks built on a GARCH DCC-NN models.

Another measure of the loss function based on the MCS procedure mentioned above is the R^2LOG. Unlike the QLIKE measure it only concerns the LOG difference between the "actual" volatility and the forecasting values.

Similar as the QLIKE measure, Table 4.5.4 reports the R^2LOG statistics of the loss function for the covariance forecasting models in 11 country samples. A lower value is preferred when selecting a better performance model.

The TGARCH DCC-NN appears to be the best performer like using other methods above. A minimum R^2LOG value is reported in 5 countries. The second is the EGARCH DCC and GARCH DCC-NN with a minimum R^2LOG value recorded in 2 countries, respectively. The

CGARCH DCC and EGARCH DCC-NN are identified as the best performers for once each in 1 country.

The second-best model is identified as the EGARCH DCC-NN in 4 countries followed by TGARCH DCC-NN with a minimum R^2LOG values in 3 countries. The TGARCH DCC and GARCH DCC-NN comes to be the third with the second minimum R^2LOG in 2 countries, respectively.

The third-best model is recorded as GARCH DCC-NN in 6 countries. The second is the EGARCH DCC-NN in 3 countries. The GARCH DCC and the TGARCH DCC-NN are counted as the third best performer in 1 country each.

The CGARCH DCC not surprisingly comes to be the worst performing model with the highest R^2LOG values in 5 countries and GARCH DCC in 6 countries as well.

Again, in attempt to look at the overall performance of the models using R^2LOG loss function in Table 4.5.4, the conclusion is similar to the results from other comparison techniques. The hybrid model TGARCH DCC-NN and GARCH DCC-NN takes most of the position when identifying the best model. Likewise, the hybrid models are still superior when considering the second-best models and third-best performer while the worst performer are identified as the CGARCH DCC within all the comparison techniques.

Table 4.5.1 MAE statistics of covariance forecasting between USA and others*10^4

	GARCH DCC	EGARCH DCC	TGARCH DCC	CGARCH DCC	GARCH DCC NN	EGARCH DCC NN	TGARCH DCC NN	CGARCH DCC NN
India	3.690572	2.874285 a	3.393725 b	3.99503	3.394603	3.927742	4.155585 *	3.672583
Japan	3.369914	2.814441	2.971616	3.419902 *	2.142856	2.456087	2.246456 b	2.423253 c
HK	1.517489	1.445276	1.408752	1.663692 *	1.078639	0.980402 b	0.944003 a	1.121172
Shanghai	1.66805	1.539765	1.558486	1.699778 *	1.295733	1.291157 b	1.283065 a	1.351175
Canada	3.663669	3.172149	3.325087	3.969414 *	2.010629	1.913547 b	1.874288 a	2.04797
Korea	3.664317	3.167314	3.533336	3.84337 *	2.1149 c	2.055067 b	2.028915 a	2.344618
UK	1.344896	1.214718	1.246533	1.598052 *	0.960976	0.869733 b	0.862461 a	1.001043
Switzerland	2.473614	2.220362	2.300264	2.602498 *	1.954078	2.170092	2.123891 c	2.013752 b
Germany	1.177534	1.135293	1.209485 *	1.136931	0.782673	0.771796 c	0.769442 b	0.758306 a
France	2.004241	1.775475 a	1.9243 b	2.082953	1.924542	2.065823	2.123252 *	1.999491
Australia	1.934025	1.774887	1.821399	1.947678 *	1.559681	1.663716	1.594886 b	1.653135 c

Note: a: Best performer, b: Second best performer, c: Third Best performer, *: Worst performer

Table 4.5.2 MSE statistics of covariance forecasting between USA and others*10^8

	GARCH DCC	EGARCH DCC	TGARCH DCC	CGARCH DCC	GARCH DCC NN	EGARCH DCC NN	TGARCH DCC NN	CGARCH DCC NN
India	35.51075	23.59438 a	32.6401	35.92501	28.4739	34.98252	38.4666 *	30.16806 c
Japan	27.85094	21.14378	23.10047	26.54387	18.7973	19.37703 c	18.07444 a	19.84317
HK	8.814232	8.447946	8.065478 c	9.217663 *	8.404563	7.942574 b	7.420931 a	8.941905
Shanghai	11.13296	10.92372	10.76245	10.61298	9.605392	9.85914 c	9.921447	9.153391 a
Canada	35.58475	28.96521	31.7025	37.24929 *	25.51451	24.74233 b	22.5915 a	25.71767
Korea	29.3516	22.44452	27.8065	29.19685	17.30415	15.37459 b	15.02408 a	18.37346
UK	8.052703	7.262643	7.481294	9.127645 *	6.538181	6.292464 b	5.881663 a	6.654394
Switzerla	15.32804	13.84765	15.16279	14.94123	12.55509	13.47103 c	13.12081 b	12.54327 a
Germany	4.721058	4.401143	5.049929 *	4.021955	3.48104	3.344706 b	3.224585 a	3.661296
France	6.09458	4.978764 a	5.84549	6.129682	5.463598	6.049523	6.340748 *	5.622056 c
Australia	5.397065	4.713749	4.923081	5.270975	4.449487	4.516243 c	4.36479 a	4.569199

Note: a: Best performer, b: Second best performer, c: Third Best performer, *: Worst performer

Table 4.5.3 QLIKE statistics of covariance forecasting between USA and others

	GARCH DCC	EGARCH DCC	TGARCH DCC	CGARCH DCC	GARCH DCC NN	EGARCH DCC NN	TGARCH DCC NN	CGARCH DCC NN
India	1.620268	1.556203 b	1.62603	1.762823 *	1.522785	1.6179 c	1.611965	1.6639
Japan	1.650193	1.544704	1.541044	1.760817 *	1.441178	1.427016 b	1.360048 a	1.519536
HK	1.275135	1.233984	1.180817	1.408715 *	1.18084 c	1.005245 b	0.974893 a	1.265415
Shanghai	1.761977	1.776606 *	1.731536	1.734815	1.612759	1.602763 a	1.607642 b	1.624737
Canada	1.726873	1.636807	1.611832	1.852188 *	1.55291 c	1.472905 b	1.404275 a	1.581144
Korea	1.735927	1.626694	1.634683	1.882638 *	1.459084	1.347305 a	1.530857 c	1.562336
UK	0.962459	0.923167	0.901463	0.975963 *	0.897764	0.834696 b	0.808894 a	0.926055
Switzerland	1.624317	1.503962	1.467238	1.721115 *	1.506685	1.474402 b	1.409626 a	1.549114
Germany	1.65205	2.217297	2.218893 *	1.631933	1.365962	1.376466 c	1.361602 a	1.380238
France	1.347123	1.321339 a	1.349417	1.402719	1.327679	1.346179 c	1.44381 *	1.364391
Australia	1.379201 *	1.337171	1.335688	1.321854	1.294502	1.288547 a	1.299977 c	1.326933

Note: a: Best performer, b: Second best performer, c: Third Best performer, *: Worst performer

Table 4.5.4 R^2LOG statistics of covariance forecasting between USA and others

	GARCH DCC	EGARCH DCC	TGARCH DCC	CGARCH DCC	GARCH DCC NN	EGARCH DCC NN	TGARCH DCC NN	CGARCH DCC NN
India	1.398383	1.253566 a	1.320811 b	1.743776 *	1.360137	1.639644	1.627587	1.701939
Japan	1.506424	1.302246	1.302846	1.765792 *	0.962097	1.12885 c	0.993733 b	1.144148
HK	0.968431 *	0.867435	0.814948	0.964326	0.612913	0.471855 b	0.452127 a	0.687054
Shanghai	1.966066 *	1.967048	1.898308	1.712144 a	1.735476 b	1.835352	1.76833 c	1.801481
Canada	1.765644	1.574308	1.547924	1.988337 *	1.009451	0.937465 b	0.91215 a	1.037906
Korea	1.837591 *	1.571613	1.593554	1.322847	1.090272	1.063539 b	1.026273 a	1.346871
UK	0.786791 *	0.711378	0.69167	0.746231	0.593027	0.521451 b	0.508202 a	0.643341
Switzerland	1.506479 *	1.288881	1.239184	1.493321	1.168657 b	1.215305 c	1.11448 a	1.285848
Germany	1.680947	1.743541	1.723968	1.826098 *	1.189523	1.170072 a	1.147035 b	1.207451
France	1.255454	1.191627 a	1.221701 b	1.372442 *	1.238933	1.35482	1.350036	1.379261
Australia	1.321566 *	1.234746	1.235011	1.23256	1.073264	1.155485 c	1.089382 b	1.162713

Note: a: Best performer, b: Second best performer, c: Third Best performer, *: Worst performer

4.7 Discussion and Findings

In order to determine the best and worst performance models, Table 4.7.1 which summary the empirical results are created. It sums up the times of each model to be the best performer or worst performer.

A strong suggestion can be made that the neural networks built on a DCC model will improve the forecast performance which contributed to the results of Cai et al., (2012). The neural networks build on a TGARCH DCC model acted as the best model for 23 times. Others like the neural networks build on standard GARCH DCC and EGARCH DCC also outperforms the rest of the specifications in some of the cases. The original EGARCH DCC model showed to be the best for 7 times and other original GARCH DCC models haven't been counted for the best model in the whole comparison except for the CGARCH DCC for 1 time. It provides the evidence that the neural networks will not only improve the forecasting ability of the univariate GARCH model but will improve the DCC GARCH models which is a multivariate GARCH specification as well.

Model Confidence Set

As mentioned in the comparing technique section (section 4.5), a "Superior Set Models" (SSM) will be generated after the Model Confidence Set procedure (MCS) to test the equal predictive ability (EPA) hypothesis. The models which enter the SSM set will be assumed to have an equal predictive ability under a confidence level. With the direct comparison of the measure of MAE, MSE, QLIKE and R^2LOG statistics, although some basic results are reported in section 4.5, it is still necessary to consider the forecasting ability of each model.

Table 4.7.3 count the times of the entrance of different models reports to the Superior Set of Models based on loss function of at a 95% confidence level, particularly, the MAE, MSE, QLIKE and R^2LOG techniques. The different values in each column represent the number of models that enter the Superior Set Model at the end of the MCS procedure, when the null hypothesis of equal predictive ability (EPA) is not rejected at the 95% confidence level.

The neural networks build on TGARCH DCC process rank the first by “surviving” from the procedure of 40 times. Following is the traditional EGARCH DCC process with 34 times. The GARCH DCC combined with neural networks comes to be the third model.

Take an overview of the Superior Set of Models, it can be found that the covariance generated by a symmetric model will perform better than others, both the original EGARCH DCC and neural networks build on a symmetric GARCH(TGARCH/EGARCH) outperforms other models. However, the long memory GARCH (CGARCH) acted as the worst performer again that the accuracy of forecasting covariance takes the short run rather than a long memory.

Table 4.7 Summary of best/worst performer										
Measure	**MAE**		**MSE**		**QLIKE**		**R^2LOG**		**Total**	
Performance	**Best**	**Worst**	**Best**	**Worst**	**Best**	**Worst**	**Best**	**Worst**	**Best**	**Worst**
GARCH DCC	0	0	0	5	0	1	0	6	0	**12**
EGARCH DCC	2	0	2	0	1	1	2	0	**7**	1
TGARCH DCC	0	2	0	1	0	1	0	0	0	4
CGARCH DCC	0	8	0	3	0	7	1	5	1	**23**
GARCH DCC-NN	3	0	0	0	1	0	2	0	**6**	0
EGARCH DCC-NN	0	0	0	0	3	0	1	0	4	0
TGARCH DCC-NN	5	1	7	2	6	1	5	0	**23**	4
CGARCH DCC-NN	1	0	2	0	0	0	0	0	3	0

Table 4.7.1 Number of models that belong to SSM					
Models/Loss function	MAE	MSE	QLIKE	$\boldsymbol{R^2 LOG}$	Total
GARCH *DCC*	7	2	2	5	16
EGARCH *DCC*	9	9	8	8	**34**
TGARCH *DCC*	4	4	5	5	18
CGARCH *DCC*	1	3	1	2	7
GARCH *DCC*-NN	10	5	9	6	30
EGARCH *DCC*-NN	6	6	10	6	28
TGARCH *DCC*-NN	10	11	10	9	**40**
CGARCH *DCC*-NN	7	8	6	6	27

4.8 Conclusion and Implications

In this chapter, a methodology for covariance forecasting based on neural networks was introduced, and it is able to capture the dynamics of the financial assets, which is our main contribution. There exists co-movements and correlations among stock indices in different countries. The proposed model could fit the covariance matrix well and give a simple way to deal with the high-dimension problem that may occur in forecasting covariance by traditional GARCH model which linked with the work of Fang et a. (2021). The empirical results show that the proposed model has an outstanding performance on covariance forecasting rather than traditional DCC process. Although to our knowledge, this study is not the first attempt to apply neural networks to model the conditional variance, it is a good method to build the DCC GARCH model with the neural networks and it gives a wide comparison among twelve different assets in a large data scale, which includes more than 3600 intraday stock prices. The procedure ensures the positive value of the forecasting results, which is the covariance matrix. With the application of the method, several traditional models in finance literature could be combined with the concept of neural networks and path a new way to make forecasting for time series. The related factors including returns, volatility of single asset and realized covariance are "fed" to the machine. Afterwards, the machine was applied to forecast the covariance matrix as a result after training by these factors.

The proposed procedure was applied to analyze the stock index pairs from different countries including Australia/USA, France/USA, Germany/USA, Switzerland/USA, Korea/USA, Canada/USA, China/USA, HK/USA, Japan/USA, India/USA. The second primary contribution of this chapter was to give empirical results to show that there do exist movements and correlations between different stock indices and the correlation or covariance can be forecasted. The covariance forecasts of these 11 pairs obtained by the hybrid model based on the neural networks were more accurate than the forecasts obtained by traditional DCC process. Moreover, the empirical results showed that there is no direct evidence that a good fitness in the first step of DCC process will lead to a more accurate forecasting covariance. The third primary contribution of this chapter was to investigate the forecasting ability of covariance models in a wide range of data set across 12 countries with the application of 8 different models. Several different GARCH type models were selected in the first step of DCC process including: GARCH, EGARCH, TGARCH and CGARCH, 4 corresponding hybrid models built with neural networks were proposed and their ability of covariance forecasting was tested. This will fill the gap in the related literature which only considered very few assets and make comparisons with limited models

The main conclusion of the study is strong with the application of 4 comparison techniques including: MSE, MAE, QLIKE and R^2LOG. In this chapter, we used the DCC GARCH model as a competitor since it is one of the most widely used models in covariance forecasting. The procedure of DCC is simple to perform and related factors and parameters are easy to estimate when the dataset is in a large scale. Furthermore, since it is convenient to obtain the related factors and parameters in the DCC process, it is possible to use these factors to train the machine (neural networks), which makes the proposed method to be a robust model. Although other multivariate GARCH models were proved to be efficient in covariance forecasting, the DCC process was the most appropriate for rebuilding with the neural networks. The proposed procedure in this chapter was an effective approach to forecast the covariance in pairs, however, it can still be improved by some techniques. For example, we can add more hidden layers in the neural networks which contribute to a deep learning networks and control the activate function is also an efficient way for further study in the future. These issues were not the primary objective of this work but can be investigated in future studies. The results is useful for future research, the hybrid models can be applied directly when forecasting multivariate covariance or correlations among different assets. The results are able to provide some suggestions for market managers on risk control, especially for the portfolios containing multivariate assets in different countries. For economics, the investors can use to appropriate models in this empirical chapter to forecast the covariance and investigate the co-movements of different assets across the world which is useful for them to observe risk and revise their portfolio.

5.Emprical chapter of volatility forecasting

Volatility forecasting using Hybrid GARCH based on deep learning with trading volume

Abstract

In this chapter forecasting performance of several volatility models are investigated with the consideration of trading volume. Since the trading volume proved to be useful information when forecasting volatility, this empirical chapter will consider trading volume in the univariate volatility models. A wide volatility forecasting comparison with trading volume using four traditional GARCH models and four hybrid-built models with deep learning models will be addressed in this empirical exercise. Different stock indices from twelve countries will be investigated and the forecasting performance will be tested. Four measures of comparison are applied in the exercise based on a loss function with the realized volatility. The results show that the GARCH and TGARCH models based on deep learning method perform better than original GARCH type models and HAR models. Moreover, the results showed a positive effect of trading volume, which gives evidence that the trading volume provide additional forecasting power to the volatility forecasting when using both traditional GARCH models and hybrid GARCH models. It provides a view about the effect of trading volume on univariate volatility forecasting by both traditional GARCH genres and the hybrid models built by neural networks and deep learnings. Our findings highlight the importance of the trading volume in forecasting the volatility. It also gives a valuable insights for improving stock volatility predictions.

5.1 Introduction

Volatility forecasting has been a main topic in recent years. The trading volume as a key factor in financial assets plays an important role in volatility forecasting. The trading Volume is a variable which can be observed directly on the market. It is a variable changed daily like the returns which is relayed to the market movements. There exists a positive relationship between volatility and traded volume by the work of Álvarez et al. (2025) after examining a large number of crypto-assets. In this chapter, the exercise aimed to explore the effect of trading volume on the volatility forecasting. Up to today's work and research, the trading volume has been considered in both traditional model and machine learning approaches individually when forecasting volatility in lots of paper. Most of them are talking about volatility forecasting by using GARCH model or machine learning models alone. Some of them only focus on the price changes and omit the possible effect caused by trading volume. With the development of machine learning, it is easy to consider the effect of the trading volume in volatility forecasting. In this chapter, the empirical exercise aims to examine the forecasting performance of different models with the consideration of trading volume. A hybrid model which combines the traditional GARCH method and machine learning method (Deep Learning) together are introduced to investigate the effects of trading volume in volatility forecasting since there only exist a small number of studies to investigate a hybrid model.

In this chapter, a new approach which a deep learning (machine learning method) built with different GARCH approaches will be introduced. Unlike the hybrid model in **Chapter 3,** more hidden layers were added into the neural networks in order to make the machine work more efficiently. The new neural networks could be called a deep learning method. A deep learning method is made of multiple layers of interconnected nodes, each building upon the previous layer to refine and optimize the prediction or categorization. The trading volume will be considered as an input variable when using deep learning method. The forecasting power will be explored by using both GARCH methods and several hybrid models with deep learning methods. Basically, the volatility series generated by the GARCH type models will come into the neural networks as well as the trading volume to train the machines and a series of volatility forecasts will be generated by this deep learning method.

This chapter reports the forecast ability of a list of traditional GARCH models including the GARCH model, TGARCH, EGARCH and CGARCH. The HAR model is considered as well. A comparison between the new machine learning approach and traditional GARCH models are listed as well.

Four measures of comparison including Mean Absolute Error (MAE), Mean Squared Error (MSE), Quasi-Likelihood (QLIKE) and R^2LOG loss function will be used to discover the forecasting ability and report a wider performance in 12 countries of the hybrid models. The Model Confidence Set (MCS) will be created at the end to find a better performing model.

The structure of the chapter is organized as follows. Section 2 shows the background of the relationship between volatility and trading volume and some studies on it. Section 3 gives the data series, source, time periods. Section 4 describes the list of volatility forecasting models and a measure of realized volatility. Section 5 shows the method used to compare the performance of the models in the previous sections. In section 6, a full report of the performance will be displayed. In section 7, some further and deeper analysis will be carried out and section 8 will discuss the findings and conclusions.

This chapter described a machine learning approach built by traditional GARCH models when forecasting univariate volatility. A more efficient machine will be figured out when training the machine by adding the trading volume variable. A list of both traditional GARCH models and machine learning approaches also make some positive contributions to the lack of attention of the effect of trading volume when forecasting volatility in most of the research.

5.2 Background

The relationship between the volatility of financial markets and trading volume has attracted a great deal of attention during the past three decades. At the early stage, Karpoff (1987) reviewed previous research on the relation between price changes and trading volume in financial markets and made a conclusion that volume is positively related to the magnitude of the price change. After that, lots of papers and several theoretical models has been developed in order to investigate the relationship between the volatility and trading volume. The first one of them is the Mixture of Distribution Hypothesis (MDH) by Clark (1973). According to his work, he described a class of finite-variance distributions for price changes and suggested that "f*inite-variance distributions subordinate to the normal fit cotton futures price data better than members of the stable family*." The MDH model indicates that the volatility is positively related with the volume at a same time interval since the relation is dependent on the rate of information flow into the market. Tauchen & Pitts (1983) got some similar results with MDH. They made research on the relationship between the variability of the daily price change and the daily volume using the data of 90-day T-bills futures market which implies that the past volume will not make any additional positive contributions on the future volatility movement. In addition, many papers proposed by others like Epps & Epps (1976); Harris & Gurel (1986) and Andersen (1996) also have some findings to support the MDH model.

However, there are also findings which are inconsistent with MDH. Bessembinder & Seguin (1992) partitioned each trading activity series into expected and unexpected components, and document that while equity volatility covaries positively with unexpected futures-trading volume, it is negatively related to forecastable futures-trading activity. Later work carried out by Aggarwal & Mougoue (2010) also implies a lack of support for the mixture of distributions hypothesis (MDH). They tested competing hypotheses on the possible relationship between volatility and trading volume using data for three major currency futures contracts in foreign exchange markets (British pound, Canadian dollar and Japanese yen calculated by American dollars) and find that trading volumes and return volatility are negatively correlated.

Another model called Sequential Information Arrival Hypothesis (SIAH) introduced by Copeland (1976) is different from MDH. The model assumes that the individuals receive information sequentially and in random order which means that the reactions from traders will not happen in the same time interval. Therefore, the response of each trader to the new information establishes an incomplete equilibrium. SIAH suggests that there should be a lead-lag relation between volume and volatility. Girma & Mougoué (2002) investigated the relation between petroleum futures spread variability, trading volume, and open interest and found that "*contemporaneous (lagged) volume and open interest provide significant explanation for futures spreads volatility when entered separately.*" This finding gives evidence to support the SIAH. Some similar findings are also proposed, such as Darrat et al. (2007) and Chiang et al. (2010). However, there are also findings inconsistent with SIAH. The work by Boubaker and Makram (2011) show that "*in the majority of cases volatility persistence vanish when trading volume is included as an explanatory variable in the conditional variance equation*." They tested the effect of trading volume on the persistence of the time-varying conditional volatility of returns and implies that the MDH explains the autoregressive conditional heteroskedasticity (ARCH) phenomenon better.

Other model like the dispersion of beliefs hypothesis (DBH) proposed by Shalen (1993) and Harris and Raviv (1993) later states that "*price changes and volume are positively correlated, consecutive price changes exhibit negative serial correlation, and volume is positively autocorrelated*." The work by BLUME et al. (1994) also showed how volume, information precision, and price movements relate, and demonstrate how sequences of volume and prices can be informative. The work by Gebka and Wohar (2013) analyzes the causality between past trading volume and index returns in the Pacific Basin countries. The OLS results show a no causal link between volume and returns.

A main topic whether the volume will give useful information when forecasting volatility has been discussed for many years. By the work of Vougas (2007), after examining the relationship between trading volume and returns in Greek stock index futures market, he suggests that there is a significant relationship between lagged volume and absolute returns, while a positive contemporaneous relationship does not hold. Similarly, Le & Zurbruegg, (2010) introduced trading volume into various ARCH frameworks to improve forecasts.

More recently, the work by Kambouroudis & McMillan (2015) suggest that both the VIX and volume do provide some additional forecast power, and this is generally improved when considering both of these series jointly in the model. Aalborg et al. (2019) made a research on the behavior of Bitcoin by using return, volatility and trading volume and suggests that trading volume further improves this volatility model. Similarly, Sapuric et al. (2020) examined the relationship between volume, returns and volatility, using asymmetric models (EGARCH) in 4 time periods/subsamples and show a positive and significant relationship between volume and volatility.

However, some different views are also proposed such as Balcilar et al. (2017). They analyzed the causal relation between trading volume and Bitcoin returns and volatility, over the whole of their respective conditional distributions and showed that volume cannot help predicting the volatility of Bitcoin returns at any point of the conditional distribution.

In general, the investigation of the relationship between the volatility forecasting and volume in financial market has never stopped. Several models and results are proposed. It has obviously become crucial to almost anyone who is involved in the financial markets. In recent years, some new approaches appeared such as machine learning.

Li et al. (2009) using two hybrid models: GARCH-based Support Vector Machine (SVM) and GARCH-based Artificial Neural Networks (ANN) to investigate the trading volume and asset price risk. They state that GARCH-based SVM outperforms GARCH-based ANN for volatility forecast, whereas GARCH-based ANN achieves a better forecast result for the volatility trend. Jiahong Li et al. (2017) also proved that a better volatility forecasting performance can be generated by using deep learning method which adopts the Long Short-Term Memory (LSTM) neural network, incorporates investor sentiment and market factors.

In more recent work by Sebastião & Godinho (2021), they examined the predictability of three major cryptocurrencies—Bitcoin, Ethereum, and Litecoin using machine learning techniques including trading volume as information and suggest that machine learning provides robust techniques for exploring the predictability of cryptocurrencies. Some similar results are pointed out by Christensen et al. (2022) as well. They use several machine learning approaches including Regularization, tree-based algorithms, and neural networks to compare the volatility forecasting results produced by Heterogeneous Auto Regressive (HAR) model and show that the machine learning (ML) algorithms improve the forecasts of realized variance.

5.3 Data

The target of this chapter is to explore the effect of daily trading volume on volatility forecasting. The exercise will investigate the impact of trading volume when trying to improve the volatility forecasting process. Different from previous chapters, since the trading volume was not introduced in the previous chapter, a new dataset will be used in this chapter. Both the daily adjusted closing price and trading volume were selected, the time period is from 5 January 2005 which is the first available trading date in year 2005 to 28 December 2018 which is the final available trading date in year 2018. The time period was selected as same as the previous chapter due to the data access availability and the convenience for comparison across different chapter in the future. The twelve stock indices are selected based on the market capitalization which mentioned in **Chapter 3.** The market capitalization represents the total value of a company's stock, the stock with a high market capitalization means the size of the stock market is large. With the application of market capitalization, it allows investors to size up a stock index. The thesis aims to select the stock index with large size to test the forecast ability of different models. Therefore, the twelve daily stock price indexes are selected. The markets of selected countries in Europe, Australia and America are all developed market since the market in these countries have already experienced a long history of development while the emerging market are mainly located in the Asian area including 3 markets: China, Korea and India. The data comes from different type markets will give a comprehensive result that whether the performance of the methods is suitable or effective when the market changes.

All the price and trading volume data are obtained by the Application Programming Interface (API) finance data from “Yahoo Finance”, while the realized volatility data is obtained by “Oxford Man”.

The whole dataset is separated into two parts: one is from 5 January 2005 to 25 November 2014 (2400 observations or approximate 10 years), it is the training set for the machine. The other is from 26 November 2014 to 28 December 2018 (1000 observations or approximate 4 years), it is the part to check the 1 day rolling window forecasting performance of the machine.

The returns are calculated by the standard method (5.3.1). In order to compare the performance of different methods, the realized volatility/variance of these returns which act as the "true volatility" is obtained by the dataset used in previous chapter.

The index prices are transformed in returns by standard methods which make it more measurable in equation:

$$r_t = \frac{S_t - S_{t-1}}{S_{t-1}} * 100\% \qquad (5.3.1)$$

Some descriptive statistics of the volumes of different stock indices are reported in Table 5.3 All the volume is calculated by the local currency since the exchange rate has been changed rapidly during the sample period. The skewness and kurtosis of all the volume series are calculated and a Jarque-Bera (JB) test is carried out to discover the normality of the volume. The statistics of the JB test gives evidence that there exists significant difference in skewness and kurtosis compared with the normal distribution, which means the normality is rejected for all series which is the similar result of return series in previous chapter.

Table 5.3 Statistics of volume in all sample countries (Million)

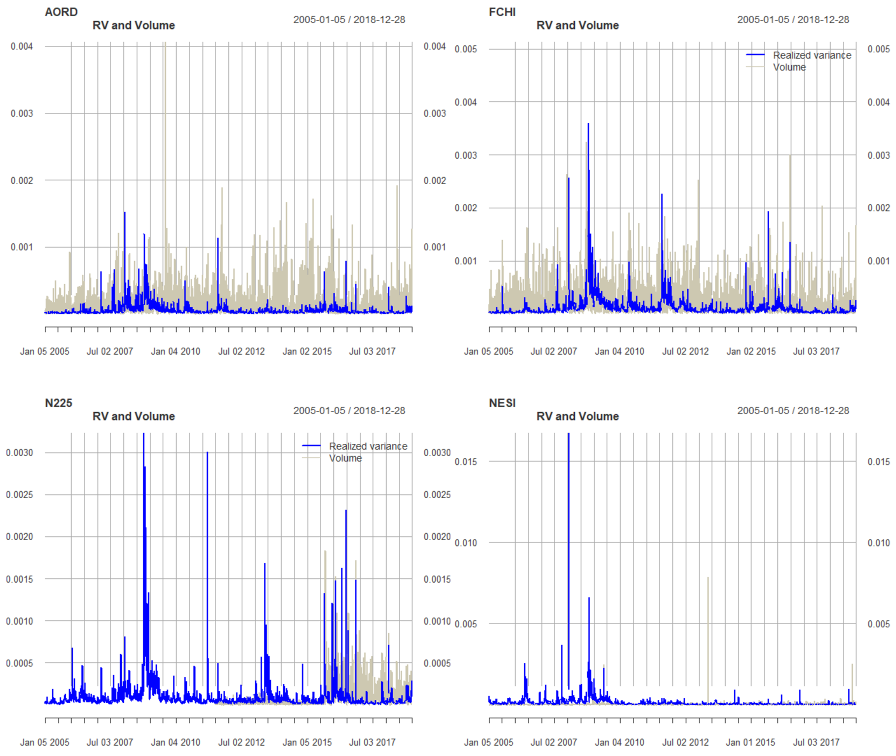
	Mean	Maximum	Minimum	Median	St.Dev	Kurtosis	Skewness	J.B test
USA	1807437	8626995	96259.09	1598311	1142164	132.2279	4.429875	24693
UK	787746.5	3880000	101670	764980	230525.3	21.64267	2.445709	38278.73
Switzerland	60847.72	346770	11170	53815	28206.13	10.02588	2.307962	17837.21
Korea*	382531.2	1210000	136330	360730	128153	2.565272	1.232337	1824.99
Japan	588057.7	3410000	51350	215015	556224.1	0.377456	1.013609	326.3476
India*	170107.4	7990000	9360	154155	168534.5	1369.688	31.17585	27183
HK	1793666	7110000	525970	1660000	645423.8	8.986052	2.221338	7620.227
German	3219770	14190035	44600	2874545	2161348	8.244298	1.666365	13990.58
France	126117.4	531250	11750	118790	49182.6	5.738412	1.585595	6412.051
China*	11645047	49730000	50170	10340000	8198334	3.40213	1.481427	2885.884
Canada	198194.2	858890	15910	191550	64629.12	10.44833	1.830975	17926.95
Australia	909719.6	5640000	100570	847840	333245	13.38832	1.767462	28737.45

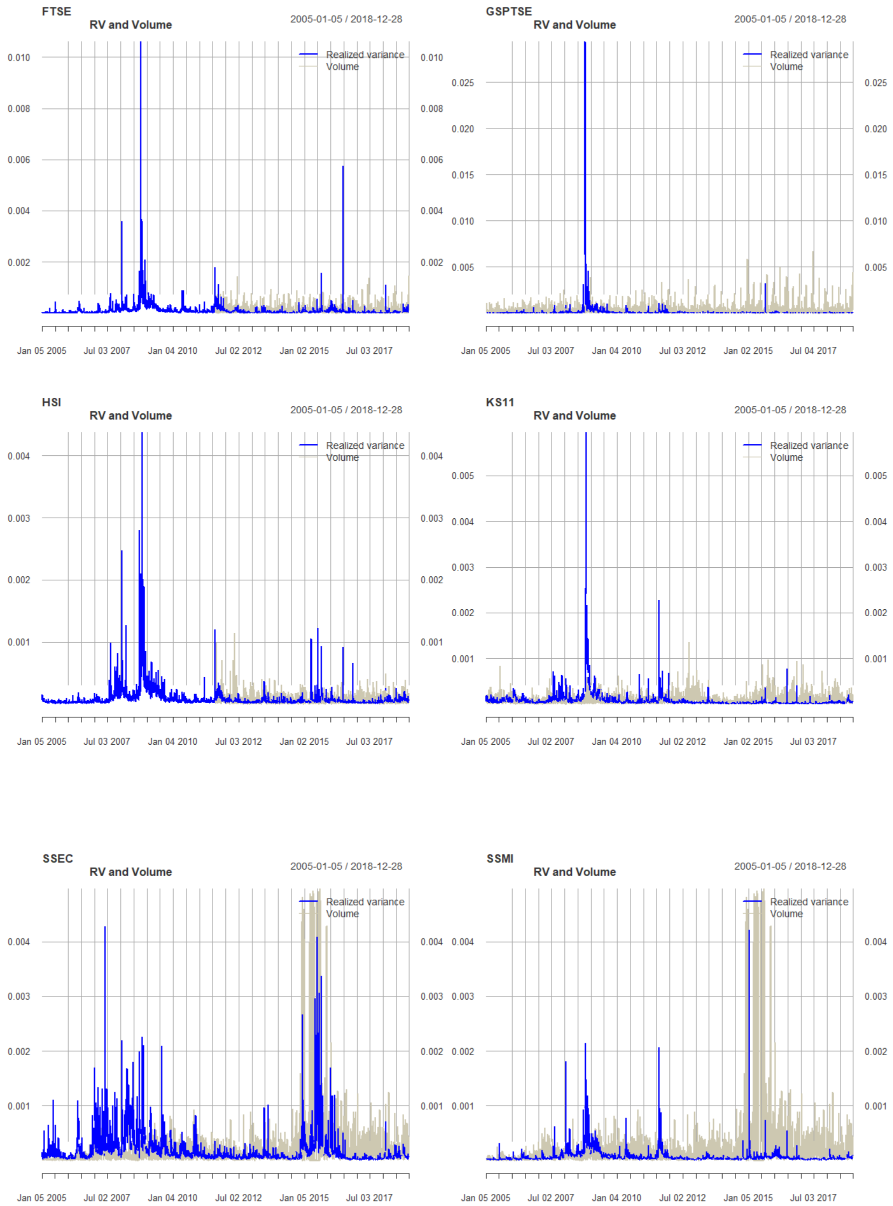
Note: * Emerging/Developing Market; St.Dev: Standard deviation

Since the volume is a large number compared with realized variance, it has been rescaled by the volume changes.

$$Vol_t = \frac{Vol_t - Vol_{t-1}}{Vol_t} \tag{5.3.2}$$

The realized variance and volume difference were plotted for a basic view of the data. It can be seen that the high-volume changes usually indicates a high realized volatility. Similar to the previous chapter, high volume and realized volatility tends to appear during the period of year 2007 and 2008 in all the samples as well as the volume changes. This should be referred to the global financial crisis happening in 2008.





In order to test the stationarity of the trading volume among the 12 stock indices, the Augmented Dickey-Fuller test (ADF Test) by Dickey & Fuller (1979) was introduced. In the test, it assumes a null hypothesis that a unit root is present in a time series sample. The procedure for the ADF test can be written as:

$$\Delta x_t = \alpha + \beta t + \gamma x_{t-1} + \delta_1 \Delta x_{t-1} + \cdots + \delta_p \Delta x_{t-p+1} + \varepsilon_t \qquad (5.3.3)$$

where α is a constant, β is the coefficient on a time trend, p is the lag order of the autoregressive process. If the parameter $\alpha = 0$ and $\beta = 0$, it is referred to a random walk. The unit root test is then carried out under the null hypothesis $\gamma = 0$ against the alternative hypothesis of $\gamma = 0$. A test statistic value will be calculated as follow:

$$DF_{\tau} = \frac{\hat{\gamma}}{SE(\hat{\gamma})} \tag{5.3.4}$$

If the calculated test statistic is less (more negative) than the critical value, then the null hypothesis of $\gamma = 0$ is rejected and no unit root is present.

After testing the stationarity of the trading volume among the 12 stock indices, all the volume data tends to be stationary with the application of Augmented Dickey-Fuller test.

5.4 Methodology

With the application of Mincer-Zarnowitz procedure, the lagged volume will be added into the regression formula to test the forecasting performance of different models.

$$\sigma_t^2 = \alpha + \beta \hat{\sigma}_t^2 + \delta Vol_{t-1} + \varepsilon_t \tag{5.4.1}$$

Four GARCH type models are considered including the symmetric GARCH: standard GARCH model, the asymmetric GARCH: TGARCH by Glosten, Jagannathan and Runkle (1993) and EGARCH by Nelson (1991) models, the long-memory models: CGARCH by Engle and Lee (1999).

Standard *Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models*

The GARCH model developed by Bollerslev (1986) is a generalization of the Autoregressive Conditional Heteroscedasticity (ARCH) model. In GARCH model, the disturbance error term ε_t is under an assumption that it is distributed by zero mean and the conditional variance h_t^2 will change with the time. The conditional variance equation in the simplest case which is the GARCH (1,1) can be written as the form:

$$\sigma_{t+1}^2 = \omega + \alpha\varepsilon_t^2 + \beta\sigma_t^2 \tag{5.4.2}$$

Exponential GARCH (EGARCH) models

A widely used asymmetric GARCH model put forward by Nelson (1991) namely the exponential GARCH (EGARCH) model provides a first explanation for the h_t^2 depends on both the size and the sign of lagged residuals in the return process.

In particular,

$$ln(\sigma_{t+1}^2) = \omega + \beta ln(\sigma_t^2) + \alpha\varphi\left(\frac{\varepsilon_t}{\sqrt{\sigma_t^2}}\right) + \alpha\psi\left|\left(\frac{\varepsilon_t}{\sqrt{\sigma_t^2}}\right)\right| \tag{5.4.3}$$

Threshold-GARCH (TGARCH) models

Another alternative asymmetric model is selected as TGARCH model. The Threshold GARCH (TGARCH) model studied by Glosten, Jagannathan, and Runkle (1993) define the conditional variance as a linear piecewise function.

$$\sigma_{t+1}^2 = \omega + \alpha\varepsilon_t^2 + \delta I_t\varepsilon_t^2 + +\beta\sigma_t^2 \tag{5.4.4}$$

$$I_t = \begin{cases} 1 & \varepsilon_t < 0 \\ 0 & \varepsilon_t > 0 \end{cases} \quad (5.4.5)$$

Component GARCH (CGARCH)

The component GARCH (CGARCH) model following by Engle and Lee (1999) is derived by replacing the constant mean with a time varying long-run component q_t. The conditional variance changes by a long run component which is calculated by itself, autoregressive of the first order. The CGARCH model specification is:

$$\sigma_{t+1}^2 = q_{t+1} + \alpha(\varepsilon_t^2 - q_t) + \beta(\sigma_t^2 - q_t) \quad (5.4.6)$$

$$q_{t+1} = \omega + \gamma q_t + \varphi(\varepsilon_t^2 - \sigma_t) \quad (5.4.7)$$

Heterogenous autoregressive models (HAR model)

Another model which is able to capture the long memory effect is first proposed by Corsi (2003), particularly, the heterogeneous autoregressive (HAR) model. It proves to successfully achieve the purpose of modeling the long-memory behavior of volatility in a very simple and parsimonious way (although not formally belonging to the class of long-memory models) by taking the realized volatility into account. The daily HAR is expressed by

$$\sigma_{t+1} = \beta_0 + \beta_D RV_{t-1,t} + \beta_W RV_{t-5,t} + \beta_M RV_{t-22,t} + \varepsilon_{t.t+1} \quad (5.4.8)$$

Where $RV_{t-1,t}$, $RV_{t-5,t}$, $RV_{t-22,t}$ represents to the 1 day, 5 days and 22 days of the volatility in a time period which can be viewed as "one trading week" and "one trading month" refer to the average realized volatility of 5 days lagged and 22 days lagged. β_0, β_D, β_W, β_M can be estimated with the application of an Ordinary Least Squares (OLS) estimation. The HAR model believes that the latent realized volatility can be observed over time horizons longer than one day. It creates an AR regression of the 1 day, 5 days and 22 days average realized volatility to make forecasting.

The neural networks built on GARCH model

With the hope to improve the forecasting performance of the GARCH model, a hybrid model based on neural networks are introduced. An artificial neural network (ANN) is a network of interconnected elements called neurons. The neurons are used to estimate functions based on the inputs. The neurons are connected with each other by joint mechanism which is consisted of a set of assigned weights. The method can be described as follows:

$$\hat{\sigma}_t = \varphi\left(\sum_{i=1}^{n} \omega_{p_i} x_i + b_p\right) \quad (5.4.9)$$

$\{x_i\}$ is the input data which includes the volatility and the trading volume $\{\sigma_t, vol_t\}$ and $\{\omega_{p_i}\}$ describes the connection weights of neurons. b_p is the bias, $\varphi(\cdot)$ is the activation function and $\hat{\sigma}_t$ is the output of the neuron which is the univariate volatility. In ANN works, multi-layer feed forward (MLP) is a common approach which has three layers: input layer, output layer, and hidden layer. Neuron takes the values of inputs parameters, sums them up with the assigned weights, and adds a bias. With the application of transfer function, the outputs will be displayed. If more the hidden layers are added, the algorithm will turn to the deep learning method.

With the conception of the neural networks, a hybrid GARCH model can be built, particularly here a hybrid GARCH model. In neural networks, the input data can be set as an explanatory variable of financial time series, such as returns, squared returns, trading volumes, etc. Since the target of the GARCH model is to forecast volatility, the input data will be set as the variance generated by univariate GARCH models including the standard GARCH, EGARCH, TGARCH and CGARCH mentioned above and the trading volumes as well. A deep learning model is a complicated neural network with more hidden layers.

5.5 Comparison of forecast performance

As introduced in chapter 3, the Mean Absolute Error (MAE) and Mean Squared Error (MSE) are selected. The Model Confidence Set (MCS) procedure by Hansen (2011) are selected as well to test the equal predictive ability (EPA) at certain confidence level α depending on the loss functions including MAE, MSE, QLIKE and R^2LOG. To retrieve a "actual" variance, the realized variance $\tilde{\sigma}^2_{t+1}$ (calculated by the 5 mins intraday returns) is considered both the in-sample and out-of-sample period, which are believed to be closer to the unobservable covariance. The forecasting values are specified as $\hat{\sigma}^2_{t+1}$.

Mean Absolute Error (MAE) and Mean Squared Error (MSE)

The MAE calculates the average difference of the comparison with equal weight of all individual differences. The MSE calculates the average squared difference between the estimated values and the actual values. The comparison with covariance is similar with volatility. Both comparison measures report the better performance by a lower statistic.

$$MAE = \frac{1}{\delta}\sum_{t=T+1}^{T+\delta}|\tilde{\sigma}^2_{t+1} - \hat{\sigma}^2_{t+1}| \quad (5.5.1)$$

$$MSE = \frac{1}{\delta}\sum_{t=T+1}^{T+\delta}(\tilde{\sigma}^2_{t+1} - \hat{\sigma}^2_{t+1})^2 \quad (5.5.2)$$

Quasi-Likelihood (QLIKE) and R^2LOG loss function

The Quasi-Likelihood (QLIKE) loss function is a test of forecast bias implied by a Gaussian likelihood which has a specification:

$$QLIKE_{t+1} = log(\hat{\sigma}^2_{t+1}) + \frac{\hat{\sigma}^2_{t+1}}{\tilde{\sigma}^2_{t+1}} \quad (5.5.3)$$

The other loss function is the R^2LOG which measure the goodness of fit of the out-of-sample forecasts.

$$R^2LOG_{t+1} = \left[log\left(\frac{\hat{\sigma}^2_{t+1}}{\tilde{\sigma}^2_{t+1}}\right)\right]^2 \quad (5.5.4)$$

Model Confidence Set (MCS)

The statistical method called Model Confidence Set (MCS) are selected to deliver the best performing models with respect to a given criterium since it is hard to point a single model outperforms others by the reason of their statistically equivalent property or not enough information from the data. The MCS procedure will eliminate the worst model at each step if the null hypothesis of equal predictive ability (EPA) is not accepted and the procedure will iterate until all models with EPA hypothesis entered in the "Superior Set Models" (SSM). An optimal SSM set will be created to give a clear view of the selected models.

5.6 Results and analysis of empirical exercise

All the data and models and comparisons are applied with R studio in an R language environment.

5.6.1 The traditional model performance with the consideration of trading volume

Since the trading volume was considered in this exercise, it is essential to know whether there would be any positive or negative effects on forecasting performance. Therefore, several comparisons are carried out. The first exercise is to investigate the forecasting performance of traditional models including GARCH series (GARCH, EGARCH, TGARCH and CGARCH) and HAR model. The exercise marked the GARCH series and HAR model with a star "*" to represent the models with the consideration of trading volume. The models without the star "*" is a normal process without consideration of trading volume. The stock index in 12 countries was selected as same as the previous chapter in order to make the comparison simpler to understand.

In general, the forecasting performance are represented by the "distance" between the forecasting values outputted by the several models and the "realized variance" which calculated by the intraday high frequency data. The distance is calculated by several comparison techniques like MAE, MSE, QLIKE and R^2LOG in this chapter. Therefore, the accuracy of the forecasting can be presented by numbers easily so that it can be compared easily. In this exercise, all of the models are investigated, and the details of "distance" were restored in several tables in the appendix since it is too big to list in the main body of this section.

Several graphs and a report table 5.6.1 which using different comparing techniques are created to examine the performance of the models. (Other table details of all the models can be found in the appendix which mentioned above). Since the patterns of graphs by each certain comparison technique is similar, only GARCH results are presented in this section. (Other graphs can be found in the appendix chapter).

By MAE statistics, the GARCH models with the consideration of volume has a negative effect in 3 countries which is Germany, India and Switzerland which means the accuracy will be lower when considering the trading volume.

By MSE statistics, the models with the consideration of volume perform worse in 3 countries which is German and India and USA while the model with the consideration of trading volume has a better performance.

By QLIKE statistics, the forecasting performance has a better accuracy in 9 countries when considering trading volume, while there still exists 3 countries which has a worse performance including Canada, German and Japan.

With the application of R^2LOG statistics, 3 countries are reported with a worse forecasting performance with the consideration of trading volume which is China, German and United Kingdom.

After producing all the figures of the distance by different comparison techniques, by an overview of all the graphs, it could be found that the lagged trading volume improve the accuracy of the forecasting volatility out of sample since the lines which represent the models with the consideration of volume run below the original models, however, this effect is not very significant. (The two lines are very close.)

5.6.2 The Hybrid models performance

Section 5.6.1 has investigated the performance or GARCH model when considering the trading volume. In this section, another investigation has been conducted in order to enhance the accuracy of forecasting with the application of the hybrid models by neural networks. It is known that the results output by the neural networks will have difference when using different hidden layers. Therefore, the comparison among hybrid models with single hidden layer and serval hidden layers, particularly, in this exercise with 10 hidden layers, will be carried out. The hybrid model with single hidden layer is tailed with "-NN", the hybrid models with 10 hidden layers which is deep learning models are tailed with "-DL" in order to read the table more efficiently.

Table 5.6.2 showed the MAE statistics of all the models in 12 countries. The hybrid model built by GARCH and deep learning are reported to be better in 8 countries while it has a worse forecasting performance in 4 countries which is German, India, Japan and Switzerland. The hybrid model built by EGARCH and deep learning are reported to has a worse performance in China, German, India and Switzerland. The hybrid model built by TGARCH and deep learning are reported to have a better performance in 7 countries while it fail to improve the forecasting accuracy in 5 countries including Canada, China, German, India and Switzerland. The hybrid model built by CGARCH has been reported to be useful in 8 countries and has a worse performance in 4 countries in Canada, China, German and India.

With a total calculation by the whole view of the table, the deep learning model has a better performance by 31 times in 48 cases (64.5%) while they have a worse performance by 17 times in 48 cases. (35.5%)

Table 5.6.3 showed the MSE statistics of all the models in 12 countries. The hybrid model built by GARCH and deep learning are reported to be better in 9 countries and has a worse performance in 3 countries which is Canada, German and Korea. The hybrid model built by EGARCH and TGARCH and deep learning has a better performance in 10 countries while a worse performance is reported in 2 countries which is German and Switzerland. The hybrid model built by CGARCH and deep learning has a better performance in 9 countries and a worse performance in 2 countries which is German, Korea and Switzerland.

By an overview of the table, the hybrid model with more hidden layers has a better performance than the hybrid model with single hidden layer. Particularly, the deep learning model was superior to the original neural networks by 38 times in 48 cases. (79.1%)

Similarly, table 5.6.4 report the QLIKE statistics of the forecasting performance. The Hybrid model build by deep learning and standard GARCH has a better performance in 9 countries, while it was reported a worse performance in 3 countries including Hong Kong, Korea and Switzerland. The deep learning model built with EGARCH are reported to be better in 10 countries except in China and Korea. The hybrid model built by TGARCH and Deep Learning has a better forecasting accuracy in 10 countries out of 12 while it fail to improve the forecasting performance in China and France. The hybrid model built by CGARCH and deep learning also has a better performance in 10 countries out of 12 except in Hong Kong and India.

With an overview of the table, the deep learning model was superior to the original neural networks by 39 times in 48 cases. (81.25%)

Finally, Table 5.6.5 report the R^2LOG statistics of the forecasting performance. It can be seen that the hybrid model built by GARCH and Deep Learning has a better performance in 9 countries out of 12 except in Australia, China and Hong Kong. The hybrid model built by EGARCH and Deep Learning are reported to be better in 8 countries out of 12 while they are not able to improve the forecasting accuracy in China, Hong Kong, Korea and United Kingdom. By looking the performance of hybrid model built by TGARCH and deep learning. It is superior to the original neural networks in 8 countries but it still has a worse performance in Australia, France, German and Hong Kong. When investigating the performance of the hybrid model built by CGARCH and Deep Learning, the model acts to be better in 6 countries including Canada, China, German, Korea, United Kingdom and United States.

By an overview of the whole table, it can be found that the was superior to the original neural networks by 31 times in 48 cases. (64.5%)

After investigating the forecasting performance of the original hybrid models and the new deep learning model with the application of 4 comparison techniques, it can be found that the deep learning method has a positive effect on the forecasting process, the new hybrid model built by deep learning has at least 60% percentage to be superior to the original neural networks.

5.6.3 The forecasting performance among original GARCH series, HAR models and deep learning

Since it is known that the forecasting performance will be improved when considering trading volume by previous exercise, the section will investigate the forecasting performance of GARCH series, HAR model, and the new built hybrid model with deep learning.

The trading volume is considered in all the models in this exercise. The GARCH series with the consideration of trading volume is named as GARCH*, EGARCH*, TGARCH* CGARCH* and the HAR model is named as HAR*. The deep learning built on different type of GARCH models with the consideration of trading volume are recorded as GARCH-DL*, EGARCH-DL*, TGARCH-DL* and CGARCH-DL* which are corresponding to the different GARCH type models for the estimation of volatility.

Table 5.6.6 gives the MAE statistics of the GARCH type models and their extension based on deep learning. The volatility forecasting results which considered the effect of volumes in 12 different countries by using different models are recorded. The realized volatility is calculated by the high frequency 5 mins intraday returns and is obtained from the dataset which is introduced in the previous chapter.

From the table, the best performer can be identified as hybrid model including GARCH-DL*, EGARCH-DL*, TGARCH-DL*, CGARCH-DL* and normal CGARCH* models. The details about which model appears to be the best in which country can be found in the following table.

There is no minimum MAE value reported for GARCH*, EGARCH*, TGARCH*,HAR* which means that these models do not appears to be the best model when using MAE statistics.

Best performer

CGARCH*	**GARCH-DL***	**EGARCH-DL***	**TGARCH-DL***	**CGARCH-DL***
Canada	Australia	China	Germany	HK
India	France			Japan
UK	Korea			Switzerland
USA				

The second-best performing models can be found by the second minimum MAE values reported in each row which are reported as GARCH*, CGARCH*, GARCH-DL*, TGARCH-DL*, CGARCH-DL* and HAR* model. A summary of the performance is recorded in following table.

2nd Best performer

GARCH*	**CGARCH***	**GARCH-DL***	**TGARCH-DL***	**CGARCH-DL***	**HAR***
Canada	Korea	Germany	China	Australia	USA
India		HK	Japan	France	UK
			Switzerland		

The third-best performing models can be observed by the third minimum MAE values in each row. The results are in a wide dispersion. The TGARCH-NN model comes to the first in 3 countries. The GARCH, CGARCH and CGARCH-NN appeared to be the 3rd best performer in 2 countries, respectively, followed by HAR model in 1 country.

3rd Best performer

GARCH*	**TGARCH***	**GARCH-DL***	**EGARCH-DL***	**TGARCH-DL***	**CGARCH-DL***	**HAR**
Korea	UK	China	Australia	France	Germany	Canada
USA		Switzerland	HK			India
			Japan			

Finally, Table 5.5.1 also reported the worst performing models of volatility forecasting by considering the maximum MAE statistics in each row. Here, the normal EGARCH* appears to be the worst performer with a maximum MAE value in 4 countries.

Worst Performer

GARCH*	**EGARCH***	**TGARCH***	**EGARCH-DL***	**CGARCH-DL***
Japan	Australia	Korea	India	Canada
China	HK		USA	
	Germany		UK	
	France			
	Switzerland			

By observing the whole performance of the selected models, the best models can be identified as the new model built on the deep learning since the best performer are recorded the most times in the series of new models.

Since the MAE methods are the simplest way like an “naïve method” to compare the error which takes the difference with an equal weight, the MSE measure is carried out to identify if there exist some outlier predictions with huge errors.

Similarly, Table 5.6.7 reports the MSE statistics for the volatility forecasting results between 12 countries by using 9 different models with the consideration of volume. The MSE will put larger weight on the errors due to the squaring part of the function which will enlarge the outlier prediction error.

Unlike using MAE statistics, the TGARCH-DL* and CGARCH-DL* model comes to be the first with a minimum MSE value in 3 out of 12 countries, respectively. The rest are GARCH-DL* and EGARCH-DL* in 2 countries each. The CGARCH* become to be the best performer in India.

Best performer

CGARCH*	**GARCH-DL***	**EGARCH-DL***	**TGARCH-DL***	**CGARCH-DL***
India	France	Australia	German	Canada
	Korea	China	Japan	HK
			Switzerland	UK

The second-best model can be identified as the TGARCH-DL* model with a second minimum MSE value in 4 countries. The rest are EGARCH-DL* in 3 countries, CGARCH-

DL* in 2 countries. The CGARCH* HAR* and GARCH-DL* model are reported with a second minimum MSE value in 1 country each.

2^{nd} Best performer

CGARCH*	HAR*	GARCH-DL*	EGARCH-DL*	TGARCH-DL*	CGARCH-DL*
Korea	India	Switzerland	France	Canada	Australia
			German	China	Japan
			USA	HK	
				UK	

The third-best model is recorded as CGARCH-DL* in 4 out of 12 countries. The rest are the GARCH-DL* in 3 countries and EGARCH-NN in 2 countries. The GARCH*, HAR*, and TGARCH-DL* model comes to the last with a third-minimum MSE value in 1 country each.

3^{rd} Best performer

GARCH*	HAR*	GARCH-DL*	EGARCH-DL*	TGARCH-DL*	CGARCH-DL*
India	Canada	HK	Korea	Australia	China
		Japan	UK		France
		USA			German
					Switzerland

Like the results with using MAE statistics, the normal EGARCH* comes to be the worst performing model again with the highest MSE statistics in 4 out of 12 sample countries. Likewise, The GARCH* model are reported as the worst performer for 4 times as well. The TGARCH*, CGARCH* EGARCH-DL* and TGARCH-DL* appeared to be the worst performer in 1 country each.

Worst performer

GARCH*	**EGARCH***	**TGARCH***	**CGARCH***	**EGARCH-DL***	**TGARCH-DL***
China	Australia	Korea	Switzerland	Canada	India
Japan	France				
UK	German				
USA	HK				

Loss of the variance

Model Confidence Set procedure (MCS) is an iteration of a sequence of statistic tests which permits to construct a set of "superior" models, the "Superior Set Models" (SSM), under the null hypothesis of equal predictive ability (EPA) is not rejected at certain confidence level α based on a loss function. It is introduced in chapter 3 and the MCS are applied as well to check the predictive ability. A direct comparison based on the loss function QLIKE and R^2LOG are taken out. The details of the "Superior Set Models" (SSM) will be displayed as a conclusion in next section.

The QLIKE and R^2LOG are selected to check if there is any significant difference. The loss function specification for covariance is specified as:

$$QLIKE_{t+1} = log(\hat{\sigma}_{t+1}^2) + \frac{\hat{\sigma}_{t+1}^2}{\tilde{\sigma}_{t+1}^2} \quad (5.6.1)$$

$$R^2LOG_{t+1} = \left[log\left(\frac{\hat{\sigma}_{t+1}^2}{\tilde{\sigma}_{t+1}^2}\right)\right]^2 \quad (5.6.2)$$

where $\tilde{\sigma}_{t+1}^2$ refers to the "actual" variance and evaluated covariance is specified as $\hat{\sigma}_{t+1}^2$.

The R^2LOG statistics prefers a value closer to zero which indicates that the distance between the "actual" variance and forecasting variance is small enough. The value is no doubt non-negative due to the squaring part of the function so that a lower R^2LOG statistics is preferred when comparing different models.

With the application of the QLIKE loss function in equation (5.5.1), it is hard to say that a lower statistic indicates a better performance. In order to make it more appropriate to compare directly, a specification is defined as:

$$QLIKE_{t+1} = \underbrace{\left(log(\tilde{\sigma}_{t+1}^2) - log(\hat{\sigma}_{t+1}^2)\right)}_{Part\ 1} + \underbrace{\left(\frac{\hat{\sigma}_{t+1}^2}{\tilde{\sigma}_{t+1}^2} - 1\right)}_{Part\ 2} \quad (5.6.3)$$

These two parts of the QLIKE statistic measures the log difference between the "actual" variance and forecasting variance and their rates. The two parts tend to be zero if the forecasting values tend to be same as the "actual" value so that a QLIKE statistic closer to zero will be preferred when comparing different models.

In Table 5.6.8 the QLIKE statistics based on equation (5.6.3) are reported.

The HAR* and GARCH-DL* comes to be the first with a minimum QLIKE value reported in 3 out of 12 countries followed by normal GARCH* and TGARCH-DL* in 2 countries. The rest are EGARCH-DL*, CGARCH-DL* in 1 country, respectively.

Best performer

GARCH*	**HAR***	**GARCH-DL***	**EGARCH-DL***	**TGARCH-DL***	**CGARCH-DL***
China	Canada	Australia	France	German	HK
India	UK	Japan		Switzerland	
	USA	Korea			

The second-best performing models can be found by the second minimum QLIKE statistics reported in each row. The CGARCH* appears to be the first in 3 country samples followed by EGARCH-DL* and CGARCH-DL* model in 2 countries, respectively. The TGARCH*, GARCH-DL* and TGARCH-DL* appeared to be the 2^{nd} best performer in 1 country each.

2nd Best performer

TGARCH*	CGARCH*	HAR*	GARCH-DL*	EGARCH-DL*	TAGRCH-DL*	CGARCH-DL*
USA	Canada Korea UK	India	France	HK Switzerland	China	German Japan

The third-best performing models can be observed by the third minimum QLIKE values in each row. There are 8 models recorded as the 3rd best performer by QLIKE statistics. Their performance in details can be found in the following table.

3rd Best performer

GARCH*	TGARCH*	CGARCH*	HAR*	GARCH-DL*	EGARCH-DL*	TGARCH-DL*	CGARCH-DL*
Canada	UK	USA	China	HK	Australia German	Japan	France India Switzerland Korea

Finally, Table 5.5.3 reported the worst performing models of the variance forecasting by considering the maximum QLIKE statistics in each row. The EGARCH* appears to be the worst performer with a maximum QLIKE value in 4 countries.

Worst performer

EGARCH*	TGARCH*	HAR*	GARCH-DL*	TGARCH-DL*	CGARCH-DL*
Australia	HK	France	Canada	India	China
German	Switzerland			USA	UK
Japan					
Korea					

The overall conclusion of Table 5.5.3 generated by the whole performance of the selected models are similar with the measures applied above. The hybrid GARCH models built on deep learning perform better than others when considering trading volume.

Another measure of the loss function based on the MCS procedure mentioned above is the R^2LOG. Unlike the QLIKE measure it only concerns about the LOG difference between the "actual" volatility and the forecasting values. Table 5.6.9 reports the R^2LOG statistics of the loss function for the variance forecasting models in 12 country samples. A lower value is preferred when selecting a better performance model.

The TGARCH-DL* appears to be the best performer with a minimum R^2LOG value reported in 4 countries. The rest are CGARCH* in 3 countries and GARCH -DL*in 2 countries. The TGARCH*, EGARCH-DL* and CGARCH-DL* model appeared to be the best performer in 1 country, respectively.

Best performer

TGARCH*	CGARCH*	GARCH-DL*	EGARCH-DL*	TGARCH-DL*	CGARCH-DL*
USA	Canada	Korea	Australia	Switzerland	HK
	India	France		Japan	
	UK			Germany	
				China	

The second-best model is identified as the GARCH-DL* and HAR models in 3 countries, respectively followed by EGARCH-DL* and CGARCH-DL* with a minimum R^2LOG value in 2 countries each. The TGARCH-DL* and CGARCH* comes to be the last with a minimum R^2LOG value in 1 country each.

2nd Best performer

CGARCH*	HAR*	GARCH-DL*	EGARCH-DL*	TGARCH-DL*	CGARCH-DL*
USA	UK	Australia	China	HK	France
	India	Japan	German		Korea
	Canada	Switzerland			

The third-best model is recorded as EGARCH-DL* in 4 countries. The GARCH* and EGARCH* appeared to be the 3rd best model in 2 countries, respectively. The TGARCH*, GARCH-DL*, TGARCH-DL* and CGARCH-DL* model has a 3rd minimum R^2LOG value in 1 country each.

3rd Best performer

GARCH*	EGARCH*	TGARCH*	GARCH-DL*	EGARCH-DL*	TGARCH-DL*	CGARCH-DL*
India	USA	UK	Germany	Switzerland	Korea	Australia
Canada	China			Japan		
				HK		
				France		

The EGARCH* appears to be the worst performer again with a maximum R^2LOG value in 4 countries. The GARCH-DL* appeared to be the worst performer in 3 countries. The EGARCH-DL* was identified as the worst performer in 2 countries. The GARCH*, TGARCH* and TGARCH-DL* model also has a maximum R^2LOG value in 1 country, respectively.

Worst performer

GARCH*	EGARCH*	TGARCH*	GARCH-DL*	EGARCH-DL*	TGARCH-DL*
Japan	Switzerland	Korea	USA	UK	France
	HK		China	India	
	Germany		Canada		
	Australia				

Again, in attempt to look at the overall performance of the models using R^2LOG loss function in Table 5.5.4, the conclusion is similar to the results from other comparison techniques. The Deep Learning built on GARCH models appears to be a better performer, while the EGARCH* model comes to be the worst performer in 4 comparison techniques.

Table 5.6.1 Distance by Several Comparison techniques of GARCH model

	MAE		MSE		QLIKE		R^2LOG	
	GARCH	GARCH*	GARCH	GARCH*	GARCH	GARCH*	GARCH	GARCH*
USA	0.48441	0.433262	1.73871	**1.854363**	1.52444	1.44897	1.21002	1.08771
UK	0.53067	0.457283	3.925	3.210285	1.20461	1.17662	0.65939	**0.70442**
Switzerland	0.61921	**0.633365**	2.76779	2.254365	1.31612	1.29566	0.80376	0.75885
Korea	0.43238	0.363363	0.32819	0.324052	1.55011	1.44654	1.11389	1.01224
Japan	1.27871	1.216334	4.082	3.58608	1.99759	**2.15442**	2.04339	1.96775
India	0.52288	**0.533344**	0.56489	**0.571637**	1.48918	1.30227	1.05888	0.99156
HK	0.88333	0.812553	1.47589	1.357819	1.60804	1.42605	1.23678	1.04556
German	0.77622	**0.816021**	1.53319	**1.61685**	1.31212	**1.41887**	0.90173	**1.05227**
France	0.72936	0.663718	1.60642	1.535426	1.27441	1.1263	0.85304	0.83224
China	1.69475	1.55213	11.05978	10.82537	1.51456	1.35165	1.01867	**1.1456**
Canada	0.32501	0.285009	1.13559	1.125622	1.28333	**1.33622**	0.83209	0.80229
Australia	0.39618	0.333536	0.34851	0.321085	1.36356	1.2586	0.89984	0.76504

Note: The number with bold font means that the forecasting performance will be worse with the trading volume

Table 5.6.2 MAE statistics*10^-4

	GARCH-NN	**GARCH-DL**	**EGARCH-NN**	**EGARCH-DL**	**TGARCH-NN**	**TGARCH-DL**	**CGARCH-NN**	**CGARCH-DL**
USA	0.56867	0.553524	0.583088	0.564663	0.586784	0.543256	0.669822	0.58423
UK	0.479911	0.526624	0.559181	0.544566	0.607348	0.524253	0.597738	0.526651
Switzerland	0.340995	**0.366074**	0.367934	**0.386236**	0.35633	**0.364535**	0.35295	0.344116
Korea	0.234947	0.212244	0.409152	0.364534	0.413798	0.363407	0.390624	0.342524
Japan	0.536713	**0.615406**	0.617639	0.522165	0.464612	0.434545	0.48137	0.444336
India	0.576668	**0.815135**	0.591347	**0.834635**	0.594245	**0.835254**	0.416697	**0.576034**
HK	0.474598	0.415564	0.495551	0.435033	0.490315	0.430245	0.373379	0.315428
German	0.62425	**0.62443**	0.619171	**0.622306**	0.571142	**0.592336**	0.626765	**0.662322**
France	0.40506	0.385913	0.53708	0.506797	0.82505	0.485203	0.64887	0.455837
China	1.373458	1.233183	1.118296	**1.154663**	1.121503	**1.155554**	1.224335	**1.330514**
Canada	0.505941	0.48101	0.499293	0.430204	0.465875	**0.48326**	0.464788	**0.485583**
Australia	0.221345	0.198361	0.234185	0.213654	0.233383	0.212323	0.22838	0.205261

Table 5.6.3 MSE statistics*10^-8

	GARCH-NN	**GARCH-DL**	**EGARCH-NN**	**EGARCH-DL**	**TGARCH-NN**	**TGARCH-DL**	**CGARCH-NN**	**CGARCH-DL**
USA	1.51509	1.252654	1.4905	1.24352	1.47796	1.243253	1.52203	1.342509
UK	3.35137	2.256575	3.33733	2.325577	3.33479	2.433654	3.32971	2.36555
Switzerland	0.43491	0.431661	0.45052	**0.46752**	0.41406	**0.436503**	0.41595	**0.45251**
Korea	0.16698 a	**0.167856**	0.29294	0.286553	0.29674	0.295275	0.2851	**0.293448**
Japan	1.65084	1.253436	1.50502	1.352347	1.22938	1.075599	1.29482	1.124516
India	1.15968	0.856729	1.20112	0.886608	1.20607	0.890176	0.81878	0.61094
HK	0.79192	0.730853	0.77488	0.716116	0.77098	0.712743	0.54767	0.519625
German	0.82028	**0.858342**	0.74986	**0.765352**	0.6757	**0.713526**	0.79255	**0.82236**
France	0.53877	0.424565	0.78453	0.465226	0.7518	0.68453	0.65052	0.609189
China	12.05234	9.876209	10.24947	8.470542	10.24578	8.365514	12.14279	8.680605
Canada	1.32261	**1.132422**	1.32722	1.154256	1.3252	1.003563	1.33363	1.014615
Australia	0.15324	0.152524	0.15439	0.153403	0.15407	0.153265	0.15574	0.155352

Table 5.6.4 QLIKE statistics*10^4

	GARCH-NN	**GARCH-DL**	**EGARCH-NN**	**EGARCH-DL**	**TGARCH-NN**	**TGARCH-DL**	**CGARCH-NN**	**CARCH-DL**
USA	1.96871	1.896091	1.95389	1.785345	1.94353	1.877436	1.9767	1.829962
UK	1.40993	1.253698	1.43061	1.375804	1.42672	1.277906	1.41474	1.368981
Switzerland	1.09763	**1.276341**	1.09051	0.919693	1.07623	0.908697	1.07081	1.032524
Korea	1.07508	**1.227569**	1.4839	**1.627544**	1.49608	1.466237	1.44918	1.421213
Japan	1.4692	1.1519	1.47088	1.373792	1.2271	1.15439	1.25361	1.178249
India	1.80398	1.782786	1.81484	1.790388	1.82244	1.777708	1.50198	**1.553386**
HK	1.09433	**1.17867**	1.1263	1.108722	1.11923	1.102076	0.977	**1.06838**
German	1.26127	1.240819	1.25491	1.234714	1.17834	1.161206	1.27638	1.255325
France	1.02615	1.01505	1.21185	1.12617	1.18269	**1.2566**	1.14861	1.12355
China	1.65396	1.652406	1.41565	**1.547459**	1.41514	**1.44702**	1.66194	1.496455
Canada	1.86666	1.835328	1.83435	1.807541	1.86157	1.646178	1.86313	1.647348
Australia	1.02758	1.020685	1.08064	1.06048	1.08211	1.061583	1.06211	1.046583

Table 5.6.5 R^2LOG statistics*10^4

	GARCH-NN	**GARCH-DL**	**EGARCH-NN**	**EGARCH-DL**	**TGARCH-NN**	**TGARCH-DL**	**CGARCH-NN**	**CARCH-DL**
USA	2.22134	2.11335	2.18026	1.98002	2.16455	2.06751	2.24107	2.05732
UK	1.03004	0.97068	1.06117	**1.10057**	1.05683	1.00234	1.04023	0.98997
Switzerland	0.57566	0.51096	0.56453	0.51469	0.55512	0.50744	0.55016	**0.63162**
Korea	0.54897	0.50211	1.02045	**1.18263**	1.04311	1.03139	0.98795	0.97843
Japan	1.10977	0.98233	1.09838	0.99354	0.73133	0.7082	0.76111	**1.235**
India	1.63095	1.56167	1.64896	1.57427	1.66746	1.56922	1.09698	**1.26989**
HK	0.56651	**0.68252**	0.61308	**0.6263**	0.60575	**0.61941**	0.4032	**0.52901**
German	0.85422	0.85005	0.82134	0.81849	0.70896	**0.7106**	0.87833	0.8732
France	0.46766	0.44236	0.76045	0.66554	0.72457	**0.88527**	0.66612	**0.66752**
China	1.23875	**1.29533**	0.80039	**0.82134**	0.7998	0.71783	1.245	1.18375
Canada	1.87878	1.84575	1.83242	1.80588	1.88031	1.66023	1.87814	1.65861
Australia	0.5003	**0.52523**	0.57666	0.4825	0.57741	**0.58306**	0.54616	**0.55962**

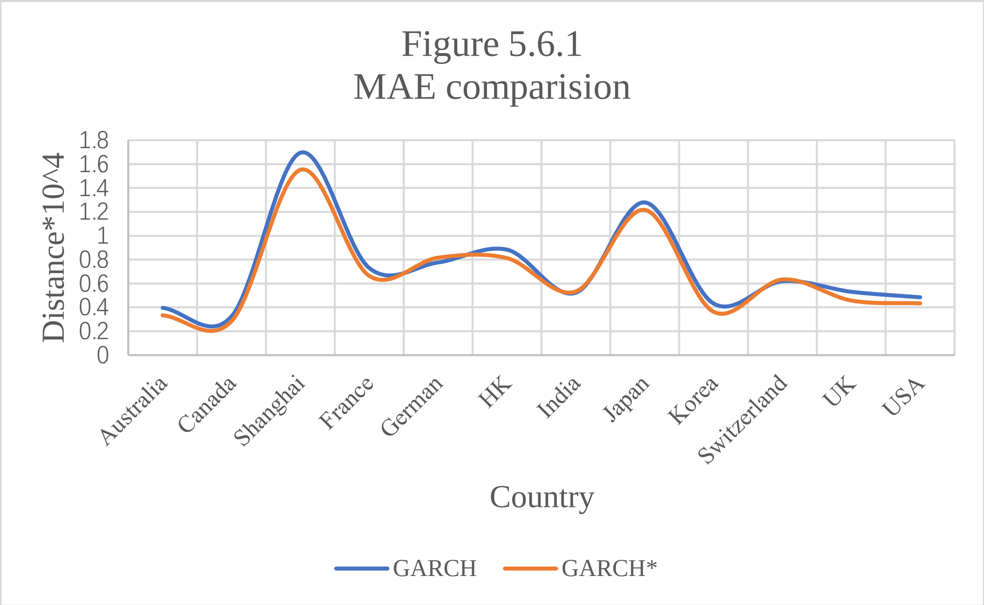


Figure 5.6.1
MAE comparision
Distance*10^4
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Australia
Canada
Shanghai
France
German
HK
India
Japan
Korea
Switzerland
UK
USA
Country
GARCH
GARCH*

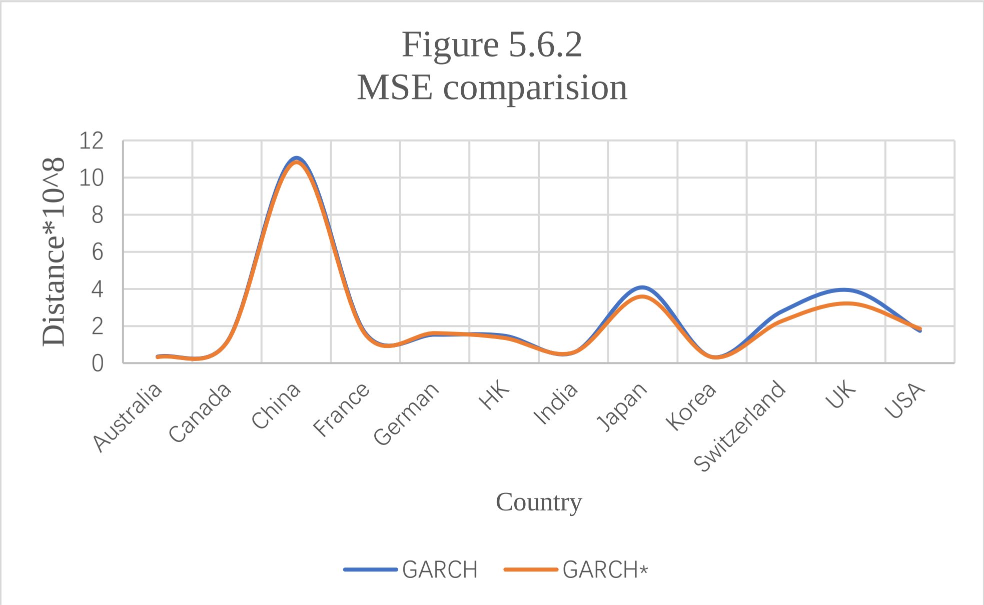


Figure 5.6.2
MSE comparision
Distance*10^8
12
10
8
6
4
2
0
Australia
Canada
China
France
German
HK
India
Japan
Korea
Switzerland
UK
USA
Country
GARCH
GARCH*

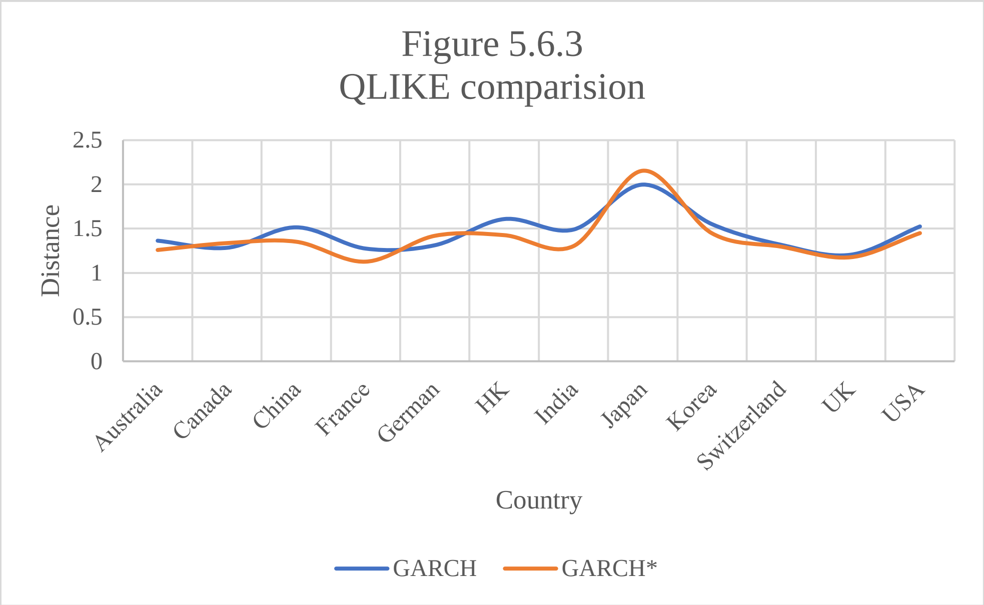


Figure 5.6.3
QLIKE comparision
Distance
2.5
2
1.5
1
0.5
0
Australia
Canada
China
France
German
HK
India
Japan
Korea
Switzerland
UK
USA
Country
GARCH
GARCH*

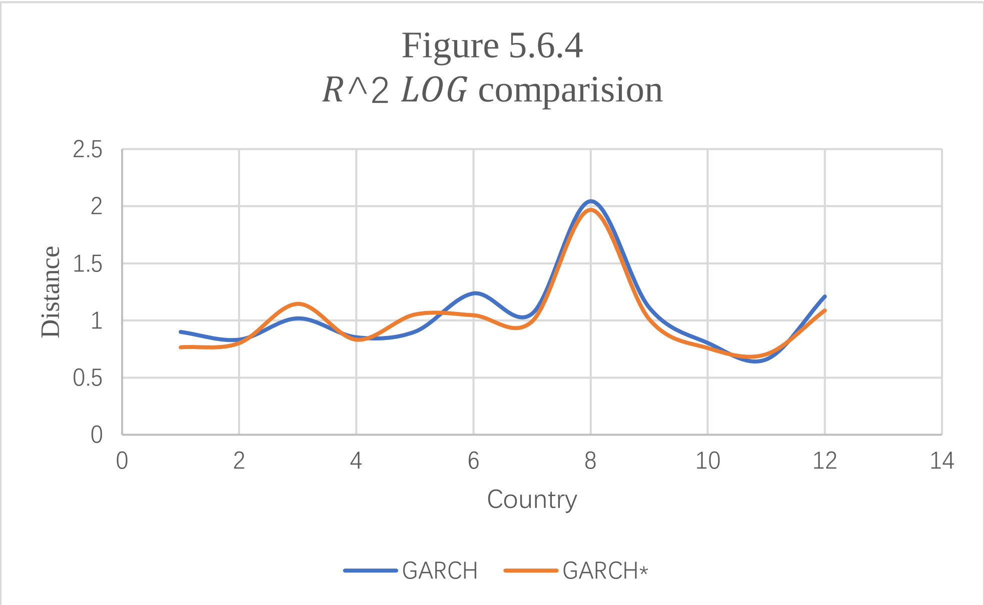


Figure 5.6.4
R^2 LOG comparision
Distance
2.5
2
1.5
1
0.5
0
0
2
4
6
8
10
12
14
Country
GARCH
GARCH*

Table 5.6.6 MAE statistics for volatility forecasting models in all samples

	GARCH*	EGARCH*	TGARCH*	CGARCH*	HAR*	GARCH-DL*	EGARCH-DL*	TGARCH-DL*	CGARCH-DL*
USA	0.43326 c	0.45318	0.43405	0.4125 a	0.42125 b	0.54776	0.58994 *	0.55641	0.57332
UK	0.45728	0.45327	0.45251 c	0.42531 a	0.45068 b	0.51661	0.53215 *	0.51557	0.52461
Switzerland	0.63337	0.6652 *	0.64435	0.56155	0.5864	0.35964 c	0.37256	0.35158 b	0.33689 a
Korea	0.36336 c	0.43432	0.45437 *	0.35315 b	0.42653	0.20512 a	0.38512	0.40157	0.39225
Japan	1.21633 *	1.15255	1.1464	1.12365	1.12176	0.65211	0.53412 c	0.44621 b	0.43019 a
India	0.53334 b	0.63221	0.64234	0.49324 a	0.53514 c	0.80119	0.82569 *	0.82451	0.60775
HK	0.81255	0.86303 *	0.85519	0.74537	0.8235	0.40351 b	0.42118 c	0.43012	0.37556 a
German	0.81602	0.93166 *	0.91414	0.76203	0.78363	0.63554 b	0.66512	0.60778 a	0.65441 c
France	0.66372	0.73758 *	0.68842	0.5927	0.643	0.40551 a	0.48594	0.47112 c	0.4478 b
Shanghai	1.55213 *	1.52359	1.54358	1.50556	1.52313	1.21324 c	1.14662 a	1.15121 b	1.30151
Canada	0.28501 b	0.31856	0.31356	0.25323 a	0.28324 c	0.46552	0.44113	0.47795	0.49541 *
Australia	0.33354	0.35332 *	0.35183	0.30264	0.32521	0.18745 a	0.20926 c	0.21312	0.19954 b

Note: a: Best performer, b: Second best performer, c: Third Best performer, *: Worst performer

Table 5.6.7 MSE statistics for volatility forecasting models in all samples

	GARCH*	EGARCH*	TGARCH*	CGARCH*	HAR*	GARCH-DL*	EGARCH-DL*	TGARCH-DL*	CGARCH-DL*
USA	1.85436 *	1.69152	1.65236	1.54236	1.46541	1.21641 c	1.18597 b	1.12664 a	1.67554
UK	3.21028 *	2.99354	3.15236	3.08390	2.95054	2.65178	2.58664 c	2.53448 b	2.42116 a
Switzerland	2.25437	2.50264	2.54625	2.5623 *	2.27252	0.45469 b	0.47665	0.44227 a	0.46123 c
Korea	0.32405	0.37466	0.38862 *	0.30655 b	0.33728	0.15760 a	0.31255 c	0.32612	0.32422
Japan	3.58608 *	3.32427	3.32421	3.48154	3.57455	1.51224 c	1.67215	1.23145 a	1.35698 b
India	0.57164 c	0.80356	0.78001	0.52402 a	0.57001 b	0.84121	0.85547	0.87717 *	0.76521
HK	1.35782	1.53769 *	1.46671	1.16783	1.36413	0.70551 c	0.70689	0.70476 b	0.50112 a
German	1.61685	2.21313 *	1.86632	1.37633	1.53669	0.95445	0.80621 b	0.75341 a	0.82452 c
France	1.53543	1.64889 *	1.35618	1.16506	1.31653	0.48556 a	0.50876 b	0.70221	0.65418 c
China	10.8254 *	10.33122	10.42657	10.45756	10.34665	10.52440	8.32551 a	8.35442 b	8.55194 c
Canada	1.12562	1.11071	1.10412	1.05530	1.05403 c	1.14205	1.14668 *	1.05261 b	0.9953 a
Australia	0.32108	0.35537 *	0.34253	0.27661	0.27262	0.16274	0.15215 a	0.15655 c	0.15422 b

Note: a: Best performer, b: Second best performer, c: Third Best performer, *: Worst performer

Table 5.6.8 QLIKE statistics for volatility forecasting models in all samples

	GARCH *	EGARCH*	TGARCH*	CGARCH*	HAR*	GARCH-DL*	EGARCH-DL*	TGARCH-DL*	CARCH-DL*
USA	1.44897	1.39479	1.36229 b	1.37041 c	1.35771 a	1.75443	1.74561	1.81694 *	1.79950
UK	1.17662	1.16203	1.14281 c	1.11112 b	0.99751 a	1.20615	1.24986	1.21774	1.31264 *
Switzerland	1.29566	1.33515	1.35344 *	1.23610	1.25673	1.25664	0.90879 b	0.88441 a	0.99593 c
Korea	1.44654	1.4965 *	1.43279	1.37384 b	1.44261	1.20546 a	1.60227	1.41549	1.40779 c
Japan	2.15442	2.19819 *	2.16624	2.18852	2.17642	1.51028 a	1.86255	1.60427 c	1.5279 b
India	1.30227 a	1.45602	1.47888	1.37651	1.31581 b	1.67527	1.66845	1.69197 *	1.39413 c
HK	1.42605	1.46555	1.46597 *	1.38871	1.41442	1.15226 c	1.08997 b	1.10112	0.99876 a
German	1.41887	1.43531 *	1.42695	1.32653	1.30551	1.18564	1.17664 c	1.14251 a	1.17641 b
France	1.12630	1.18663	1.16987	1.14227	1.22118 *	1.11656 b	1.11579 a	1.28626	1.12391 c
China	1.35165 a	1.47258	1.47654	1.49551	1.44128 c	1.56979	1.58509	1.38867 b	1.63663 *
Canada	1.33622 c	1.34876	1.35697	1.32116 b	1.29551 a	1.67444 *	1.64435	1.40102	1.46116
Australia	1.25860	1.38595 *	1.36880	1.24665	1.28014	1.00231 a	1.05239 c	1.05947	1.04774 b

Note: a: Best performer, b: Second best performer, c: Third Best performer, *: Worst performer

Table 5.6.9 R^2LOG statistics forecasting models in all samples

	GARCH*	EGARCH*	TGARCH*	CGARCH*	HAR*	GARCH-DL*	EGARCH-DL*	TGARCH-DL*	CGARCH-DL*
USA	1.08771	1.03291 c	0.98572 a	1.01781 b	1.07637	1.98518 *	1.94113	1.97519	1.96051
UK	0.70442	0.68714	0.67538 c	0.61731 a	0.64831 b	0.99541	1.15624 *	1.08351	1.05244
Switzerland	0.75885	0.78861 *	0.74891	0.67541	0.69914	0.48118 b	0.48265 c	0.47987 a	0.55752
Korea	1.01224	1.00215	1.13994 *	0.97142	1.01174	0.47971 a	1.05252	0.9575 c	0.90184 b
Japan	1.96775 *	1.87541	1.85431	1.88843	1.94621	0.95197 b	0.96651 c	0.65705 a	1.18665
India	0.99156 c	1.01821	1.12497	0.94333 a	0.98844 b	1.50641	1.51777 *	1.49813	1.12250
HK	1.04556	1.08194 *	1.07824	0.98666	1.03866	0.65994	0.60612 c	0.59874 b	0.49651 a
German	1.05227	1.07552 *	1.06814	0.95669	1.01570	0.90177 c	0.88579 b	0.81897 a	1.03251
France	0.83224	0.82315	0.77654	0.73229	0.80127	0.43186 a	0.65418 c	0.84115 *	0.64271 b
China	1.14560	1.03271 c	1.05612	1.15211	1.08741	1.30152 *	0.95741 b	0.89467 a	1.21761
Canada	0.80229 c	0.84991	0.86134	0.74213 a	0.79874 b	1.79634 *	1.73225	1.58510	1.54398
Australia	0.76504	0.82157 *	0.81994	0.71475	0.72715	0.51374 b	0.46227 a	0.55667	0.51379 c

Note: a: Best performer, b: Second best performer, c: Third Best performer, *: Worst performer

5.7 Discussion and Findings

In order to determine the best and worst performance models, a table which summary the empirical results are created (Table 5.7). It sums the times of each model to be the best performer or worst performer based on 4 measures of evaluating the forecast among GARCH type models, their extension based on deep learning in the 12 sample countries.

A strong suggestion can be made that the deep learning built on a GARCH type model will improve the forecast performance. Both deep learning built on normal GARCH and asymmetric GARCH (TGARCH) models has an overall better performance. The deep learning built on a normal GARCH model acted as the best model for 10 times and the deep learning built on TGARCH* model acted as the best model for 11 times as well. Others like normal CGARCH* model outperforms 8 times the rest of the specifications. Its extension CGARCH-DL* also has been recorded as the best model for 8 times. Moreover, the HAR* has been identified as the best model for 3 times, GARCH* for 2 times and TGARCH* for 1 time. However, other original GARCH* models haven't been counted for the best model in the whole comparison. The EGARCH* has been counted by 17 times of 48 cases as the worst performer. By an overview of the table, it gives the evidence that the deep learning methods will improve the forecasting performance as well when considering the volumes.

Model Confidence Set

As mentioned in the comparing technique section (section 5.6), a "Superior Set Models" (SSM) will be generated after the Model Confidence Set procedure (MCS) to test the equal predictive ability (EPA) hypothesis. The models which enter the SSM set will be assumed to have an equal predictive ability under a confidence level. With the direct comparison of the measure of MAE, MSE, QLIKE and R^2LOG statistics, although some basic results are reported in section 5.6, it is still necessary to have a look about the forecasting ability of each model.

Table 5.7.2 count the times of the entrance of different models using the Superior Set of Models based on loss function of at a 95% confidence level, particularly, the MAE, MSE, QLIKE and R^2LOG techniques. The different values in each column represent the number of models that enter the Superior Set Model at the end of the MCS procedure, when the null hypothesis of equal predictive ability (EPA) is not rejected at the 95% confidence level.

The deep learning built on TGARCH* model rank the first by "surviving" from the procedure of 41 followed by GARCH-DL* of 38 times. The rest are the standard CGARCH* and its extension CGARCH-DL* with 30 times both. The last place is the traditional EGARCH* with no doubt since it appeared to be the worst performer in all the 4 comparing techniques.

Take an overview of the Superior Set of Models, it can be found that the deep learning built on a standard GARCH or asymmetric GARCH (TGARCH) outperforms other models when considering trading volume.

Table 5.7 Summary of best/worst performer

Measure	***MAE***		***MSE***		***QLIKE***		***R2LOG***		***Total***	
Performance	**Best**	**Worst**	**Best**	**Worst**	**Best**	**Worst**	**Best**	**Worst**	**Best**	**Worst**
CGARCH-DL*	3	1	3	0	1	2	1	0	8	3
TGARCH-DL*	1	0	4	1	2	2	4	1	**11**	4
EGARCH-DL*	1	3	2	1	1	0	1	2	5	6
GARCH-DL*	3	0	2	0	3	1	2	3	**10**	4
HAR*	0	0	0	0	3	1	0	0	3	1
CGARCH*	4	0	1	1	0	0	3	0	8	1
TGARCH*	0	1	0	1	0	2	1	1	1	5
EGARCH*	0	5	0	4	0	4	0	4	0	17
GARCH*	0	2	0	4	2	0	0	1	2	7

Table 5.7.2 Number of models that belong to SSM					
Models/Loss function	MAE	MSE	QLIKE	***R2LOG***	Total
GARCH*	3	3	9	3	18
EGARCH*	3	3	4	4	14
TGARCH*	3	3	3	6	15
CGARCH*	11	6	3	10	30
HAR*	3	4	9	4	20
GARCH-DL*	10	8	11	9	38
EGARCH-DL*	5	8	6	5	24
TGARCH-DL*	7	12	10	12	41

5.8 Conclusion and Implication

After investigating the forecasting performance of GARCH series models, HAR models and several hybrid-built models with deep learning by using 1-step ahead volatility and taking an overview of all these tables both in section 5.6 and section 5.7, some conclusions can be generated.

The first conclusion is that the trading volume has a positive effect on variance forecasting with the application of GARCH model or HAR model. The second conclusion can be found that the deep learning method is an effective way to enhance the performance of original neural networks. The third conclusion is that with the consideration of trading volume, the new built hybrid models by deep learning still has a better forecasting performance than the original GARCH series or HAR model. Since there is very few paper explored the effect of trading volume on volatility forecasting, most of them only talk about the performance of 1 or 2 models using very limited stock index, this exercise provides a more comprehensive view of the volatility forecasting across 12 countries and 9 models when considering trading volume. With the four comparison techniques, the results become more solid rather than using 1 or 2 comparison methods.

This exercise also provides a new approach which is deep learning (neural networks with multi hidden layers) rather than original neural networks to build the hybrid models, the results suggests that the new approach is still better than the original GARCH models when considering the trading volume and this new machine learning method also outperforms the original neural networks with single hidden layers.

The wide comparison among GARCH series models, HAR models and hybrid models also gives some empirical suggestions that a hybrid model is an effective way to enhance the volatility forecasting performance. Moreover, after comparing the forecasting performance with the previous chapter, it could be found that the GARCH models with consideration of the trading volume will has a better performance than these without consideration of the trading volume which contributed to the finds of Liu et al. (2020). Similarly, the machine trained by trading volume has a better performance when forecasting volatility rather than the machine trained without volume which is a similar result of the finding by Zhu et al. (2008 The results showed a positive effect of trading volume, which gives evidence that the trading volume do provide additional forecasting power to the volatility forecasting when using both traditional GARCH models and hybrid GARCH models.

The results can be addressed for further research on univariate volatility forecasting, studies can start from the results and apply a hybrid model with deep learning to forecast volatility with the consideration of the trading volume. It provides a view about the effect of trading volume on univariate volatility forecasting by both traditional GARCH genres and the hybrid models built by neural networks and deep learnings. It also gives a valuable insight for improving stock volatility predictions. The investors can use the results to consider more about the trading volume on their investment portfolios and take action in advance to avoid risk and make profits.

6. Summary and Conclusion

6.1 Summary

In **Chapter 2**, which is the literature review chapter, some definitions and related works about the measurement and forecasting of volatility and covariance are reviewed for the next chapters. Although several models or methods have been proposed and proved to be efficient or superior for univariate volatility or multivariate covariance forecasting, there still existed a debate which model can be considered as the “best” performer. Moreover, with the development of machine learning methods, more and more studies and research tend to produce forecasts with machine learning method. However, some of the studies focused on very limited models and make comparisons in a small range of models, and there still no general conclusion on which model has the best performance. This thesis enters the debate on exploring the most appropriate and efficient model or methods to get better volatility or covariance forecasting performance.

With the aim of finding a better performer on univariate volatility forecasting, a volatility forecasting empirical exercise has been carried out in **Chapter 3**. Several traditional volatility models have been selected including standard GARCH, EGARCH, TGARCH, CGARCH and HAR models.

Since the GARCH models mainly focus on the past conditional variance and squared returns using maximum likelihood method, a hybrid model built with GARCH and neural networks is introduced. Neural networks have the ability to concerns more explanatory variables of the volatility by formulating them with an activation function including both linear and nonlinear function. These hybrid models are built with four different GARCH type models, and a comparison among the wide range of volatility models is carried out.

The results show that the proposed hybrid models built with GARCH genres and neural networks outperform overall the traditional models, which means that the hybrid GARCH model based on neural networks will improve the forecasting performance of a traditional GARCH model. After ranking the models by performance, the hybrid model built with TGARCH and neural networks comes to the first, followed by the hybrid model built by standard GARCH and neural networks. The third place was taken by the traditional long memory model CGARCH.

Neural networks, as machine learning methods, needs plenty of related data to train the machine in order to get a more accurate performance. Since the proposed models are trained by lots of factors which are estimated from traditional GARCH model, there would be a question whether the factors come from different GARCH type models will have effect of the forecasting performance. Following the questions, another comparison is carried out, we first pick the best performer in traditional GARCH series and combined it with neural networks in order to investigate the forecasting performance of it. However, the results showed that there is no strong evidence to support that a better forecasting performance in GARCH type will yield in a success of a corresponding hybrid GARCH model based on neural networks. It means that although one of the GARCH models has been identified as the best performer in GARCH series, there do not exist empirical evidence to support that the model built with neural networks will has a better performance again.

The results of the chapter are important to anyone who is involved in the volatility modelling and forecasting. It confirms the effect of machine learning methodologies. Since some of the literature concerned with volatility models only selected very limited models and assets, our empirical results make a new approach to compare four GARCH type models and four hybrid GARCH in twelve countries stock indices with five comparison measures, which gives a wider and more comprehensive comparison.

With the aim for improving the forecast ability of different models, **Chapter 4** investigates the multivariate covariance models and builds a model with neural networks to improve forecasting ability. Since lots of paper and studies put emphasis on the covariance modelling with the application of traditional models like VEC, BEKK, CCC/DCC process in the past, our empirical exercise aims to explore the covariance forecasting by a hybrid-built model based on neural networks. Moreover, with the help of the neural networks, our new built model will not be restricted to heavy parameters or high dimension estimation. The DCC process is selected from the MGARCH series as a traditional methodology to model the covariance. Several different GARCH type models are chosen in the first step of DCC process including standard GARCH, EGARCH, TGARCH and CGARCH. Similar to the previous chapter, four corresponding hybrid models built by neural networks are proposed. A wide comparison among the different approaches will be created. The forecasting power will be explored by using both DCC GARCH methods and several hybrid-built models based on neural networks.

The results showed that the proposed hybrid models are superior to the traditional DCC GARCH models. It gives evidence to support that the neural networks still have positive effect of improving forecasting accuracy on covariance models. After ranking all the models by their performance, the neural networks built on a TGARCH DCC model acted as the best model and other hybrid models built on standard GARCH DCC and EGARCH DCC also outperform the rest of the specifications.

In this chapter, we used the DCC GARCH model as a competitor since it is one of the most widely used models in covariance forecasting and the procedure of DCC is simple to perform and related factors and parameters are easy to estimate when the dataset is in a large scale, which will make the procedure of training machine to be efficient. Moreover, after investigating the performance of traditional DCC methods in the exercises individually, the results show that there is no direct evidence to support that good fitness in the first step of DCC process will lead to a more accurate forecasting covariance.

The final empirical chapter aims to investigate the effects of trading volumes on volatility forecasting. After realizing the usefulness of the neural networks in previous chapter, we introduced a neural network with more hidden layers which could be called as a Deep Learning method (DL) in order to improve the forecasting ability of original neural networks. Several traditional models including standard GARCH, EGARCH, TGARCH, CGARCH and HAR models are selected and a new deep learning model with ten hidden layers is built with four different GARCH models. A comparison among the wide range of volatility models is carried out with the consideration of trading volume.

The empirical results show that the proposed hybrid models built with GARCH and deep learning methods outperform overall the traditional models, which means that the hybrid GARCH model based on deep learnings will improve the forecasting performance of a traditional GARCH model. Moreover, after comparing the forecasting performance with the previous chapter, it could be found that the models with consideration of the trading volume will have a better performance than these without consideration of the trading volume.

Similarly,

The machine trained by trading volume has a better performance when forecasting volatility rather than the machine trained without volume. The results showed a positive effect of trading volume, which gives evidence that the trading volume provides additional forecasting power to the volatility forecasting. Furthermore, it shows that the deep learning do has improved the forecasting ability than the normal neural networks.

6.2 Conclusions

This thesis aims to fill some gaps and add some knowledge to the literature on the univariate volatility and multivariate covariance forecasting by several empirical exercises. With the application of neural networks and deep learning methods, several hybrid models were introduced to improve the accuracy of volatility and covariance forecasting.

There are four main contributions. First, a hybrid model which combined the neural networks and univariate GARCH genre of models were introduced. After a wide comparison among these models, the empirical results revealed that the hybrid model are superior to the original GARCH models and HAR models. The results is useful for future research, hybrid models can be applied directly when forecasting univariate volatility. For economics, the policymakers can benefit from the results to formulate their policies to avoid risk. The investors can use appropriate models in this empirical chapter to forecast more recent volatility to avoid risk and loss or to revise their portfolio to make more profits. Second, a multivariate hybrid model which built with the DCC GARCH models and neural networks is introduced. This hybrid model is built from the idea of the univariate model on the previous findings. After investigating the out-of- sample forecasting performance among the multivariate models and DCC GARCH models, the findings showed that the hybrid models are still preferred rather than original DCC GARCH models. The results are able to provide some suggestions for market managers on risk control, especially for the portfolios containing multivariate assets in different countries. For economics, the investors can use to appropriate models in this empirical chapter to forecast the covariance and investigate the co-movements of different assets across the world which is useful for them to observe risk and revise their portfolio. Third, the trading volume is considered in empirical exercise, the effect of the trading volume are addressed in the GARCH models and previous hybrid models. Finally, the original neural networks are improved by a deep learning model which has more hidden layers than the previous neural networks. The forecasting ability of a hybrid model which combined the univariate GARCH genre and deep learning are investigated. After the comparison of the hybrid-built models and GARCH genres with the consideration of trading volume, the proposed models are superior as well. The results can be addressed for further research on univariate volatility forecasting. Studies can start from the results and apply a hybrid model with deep learning to forecast volatility with the consideration of the trading volume. It provides a view about the effect of trading volume on univariate volatility forecasting by both traditional GARCH genres and the hybrid models built by neural networks and deep learnings. Our findings highlight the importance of the trading volume in forecasting the volatility. It also gives a valuable insight for improving stock volatility predictions. The investors can use the results to consider more about the trading volume on their investment portfolios and take action in advance to avoid risks and make profits.

In addition to these four main contributions mentioned above, some other properties of the models are examined as well. The data selected by the thesis are stock indices from twelve countries and the four different comparison techniques are selected. Since some of the exist literature investigated very few samples in their research, this thesis investigated the forecasting performance among different models to provide a more comprehensive view of the models in different countries. With the application of four comparison techniques among all the exercise, the thesis aimed to give more objective results, since the findings by one or two comparison techniques may cause bias. Moreover, the deep learning methods are introduced to replace the original neural networks in order to enhance the forecast ability of the previous models. After the investigation of the hybrid models with deep learning, the results revealed that the hybrid models built with deep learning are preferred rather than the hybrid models built with neural networks.

The forecasting of conditional volatility and covariance are addressed in lots of literature and it is still a hot topic in the finance area. There still exist lots of unsolved problems which need to be investigated further. The forecasting performance can be affected by the data frequency, the model selection, the time period of the data, the sample size, therefore it is still meaningful to do further research on these aspects.

In this thesis, all the empirical investigations are based on a hybrid model which is built with neural networks. Although deep learning is a different method, it can be identified as a modern neural network with more hidden layers. For the further development of the traditional models, more machine learning methods like SVM can be a potential research topic. Moreover, the data used throughout all the chapters are intraday returns or realized volatility, and the forecasting results are the one-step ahead forecast. In further research, a further ahead could be investigated and the forecast ability of the model could be addressed with different ahead forecasts.

7.BIBLIOGRAOHY AND REFERENCES

Aalborg, H.A., Molnár, P. and de Vries, J.E. (2019) 'What can explain the price, volatility and trading volume of Bitcoin?', *Finance Research Letters*, 29, pp. 255–265. doi:10.1016/j.frl.2018.08.010.

Aggarwal, R. and Mougoue, M. (2010) 'Trading Volume and Exchange Rate Volatility: Evidence for the sequential arrival of Information Hypothesis', *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.1758960.

Aielli, G.P. (2011) 'Dynamic conditional correlation: On properties and estimation', *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.1507743.

Akgiray, V. (1989) 'Conditional Heteroscedasticity in Time Series of Stock Returns: Evidence and Forecasts', *The Journal of Business*, 62(1), p. 55. doi:10.1086/296451.

Álvarez, E. *et al.* (2025) 'Comprehensive analysis of the crypto-assets market through multivariate analysis, clustering, and wavelet decomposition', *Physica A: Statistical Mechanics and its Applications*, p. 130330. doi:10.1016/j.physa.2024.130330.

Amirshahi, B. and Lahmiri, S. (2023) 'Hybrid deep learning and garch-family models for forecasting volatility of Cryptocurrencies', *Machine Learning with Applications*, 12, p. 100465. doi:10.1016/j.mlwa.2023.100465.

Ampountolas, A. (2021) 'Modeling and forecasting daily hotel demand: A comparison based on SARIMAX, neural networks, and GARCH models', *Forecasting*, 3(3), pp. 580–595.

doi:10.3390/forecast3030037.

Andersen, T.G. (1996) ‘Return volatility and trading volume: An information flow interpretation of stochastic volatility’, *The Journal of Finance*, 51(1), pp. 169–204. doi:10.1111/j.1540-6261.1996.tb05206.x.

Andersen, T.G. and Benzoni, L. (2008) ‘Realized Volatility’, *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.1092203.

Andersen, T.G. and Bollerslev, T. (1998) ‘Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts’, *International Economic Review*, 39(4), p. 885. doi:10.2307/2527343.

Andersen, T.G. *et al.* (2001) ‘The Distribution of Realized Exchange Rate Volatility’, *Journal of the American Statistical Association*, 96(453), pp. 42–55. doi:10.1198/016214501750332965.

Aras, S. (2021) ‘Stacking hybrid GARCH models for forecasting bitcoin volatility’, *Expert Systems with Applications*, 174, p. 114747. doi:10.1016/j.eswa.2021.114747.

Arı, Y. (2022) ‘Chasing volatility of USD/try foreign exchange rate: the comparison of Carr, EWMA, and Garch models’, *Ekoist: Journal of Econometrics and Statistics*, 0(37), pp. 107–127. doi:10.26650/ekoist.2022.37.1113670.

Ayele, A.W., Gabreyohannes, E. and Tesfay, Y.Y. (2017) ‘Macroeconomic determinants of volatility for the gold price in Ethiopia: The application of Garch and EWMA volatility models’, *Global Business Review*, 18(2), pp. 308–326. doi:10.1177/0972150916668601.

Baillie, R.T., Bollerslev, T. and Mikkelsen, H.O. (1996) ‘Fractionally integrated generalized autoregressive conditional heteroskedasticity’, *Journal of Econometrics*, 74(1), pp. 3–30. doi:10.1016/s0304-4076(95)01749-6.

Balaban, E., Bayar, A. and Faff, R.W. (2002) ‘Forecasting Stock Market Volatility: Evidence From Fourteen Countries’, *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.339520.

Balcilar, M. *et al.* (2017a) ‘Can volume predict Bitcoin returns and volatility? A quantiles-based approach’, *Economic Modelling*, 64, pp. 74–81. doi:10.1016/j.econmod.2017.03.019.

Balcilar, M. *et al.* (2017b) ‘Can volume predict Bitcoin returns and volatility? A quantiles-based approach’, *Economic Modelling*, 64, pp. 74–81. doi:10.1016/j.econmod.2017.03.019.

Banumathy, K. and Azhagaiah, R. (2015) ‘Modelling Stock Market Volatility: Evidence from India’, *Managing Global Transitions, University of Primorska, Faculty of Management Koper*, 13(1), pp. 27–41.

Basher, S.A. and Sadorsky, P. (2016) ‘Hedging emerging market stock prices with oil, gold, Vix, and bonds: A comparison between DCC, ADCC and go-GARCH’, *Energy Economics*, 54, pp. 235–247. doi:10.1016/j.eneco.2015.11.022.

Becketti, S. and Roberts, D.J. (1990) ‘Will increased regulation of stock index futures reduce stock market volatility?’, *Economic Review*, 75, pp. 33–36.

Bera, A.K. and Higgins, M.L. (1993) ‘ARCH MODELS: PROPERTIES, ESTIMATION AND TESTING’, *Journal of Economic Surveys*, 7(4), pp. 305–366. doi:10.1111/j.1467-6419.1993.tb00170.x.

Bergsli, L.Ø. *et al.* (2022) ‘Forecasting volatility of bitcoin’, *Research in International Business and Finance*, 59, p. 101540. doi:10.1016/j.ribaf.2021.101540.

Bessembinder, H. and Seguin, P.J. (1992) ‘Futures-trading activity and stock price volatility’, *The Journal of Finance*, 47(5), pp. 2015–2034. doi:10.1111/j.1540-6261.1992.tb04695.x.

Bhowmik, R. and Wang, S. (2020) ‘Stock market volatility and return analysis: A systematic literature review’, *Entropy*, 22(5), p. 522. doi:10.3390/e22050522.

Bildirici, M. and Ersin, Ö.Ö. (2009a) ‘Improving forecasts of GARCH family models with the Artificial Neural Networks: An application to the daily returns in Istanbul Stock Exchange’, *Expert Systems with Applications*, 36(4), pp. 7355–7362. doi:10.1016/j.eswa.2008.09.051.

Bildirici, M. and Ersin, Ö.Ö. (2009b) ‘Improving forecasts of GARCH family models with the artificial neural networks: An application to the daily returns in Istanbul Stock Exchange’, *Expert Systems with Applications*, 36(4), pp. 7355–7362. doi:10.1016/j.eswa.2008.09.051.

Bishop, C.M. (1995) *Neural networks for pattern recognition*. Oxford: Oxford University Press.

Black, F. and Scholes, M. (1973) ‘The Pricing of Options and Corporate Liabilities’, *Journal of Political Economy*, 81(3), pp. 637–654. doi:10.1086/260062.

Blume, L., Easley, D. and O’Hara, M. (1994) ‘Market statistics and technical analysis: The role of volume’, *The Journal of Finance*, 49(1), p. 153. doi:10.2307/2329139.

Bollerslev, T. (1986) ‘Generalized autoregressive conditional heteroskedasticity’, *Journal of Econometrics*, 31(3), pp. 307–327. doi:10.1016/0304-4076(86)90063-1.

Bollerslev, T. (1990) 'Modelling the coherence in short-run nominal exchange rates: A multivariate generalized arch model', *The Review of Economics and Statistics*, 72(3), p. 498. doi:10.2307/2109358.

Boubaker, A. and Makram, B. (2011) 'The empirical relationship between stock returns volatility and trading volume: Evidence on the tunis stock market', *International Journal of Management Science and Engineering Management*, 6(5), pp. 374–381. doi:10.1080/17509653.2011.10671186.

Brailsford, T.J. and Faff, R.W. (1996) 'An evaluation of volatility forecasting techniques', *Journal of Banking & Finance*, 20(3), pp. 419–438. doi:10.1016/0378-4266(95)00015-1.

Braun, P.A., Nelson, D.B. and Sunier, A.M. (1995) 'Good News, Bad News, Volatility, and Betas', *The Journal of Finance*, 50(5), pp. 1575–1603. doi:10.1111/j.1540-6261.1995.tb05189.x.

Cai, X., Lai, G. and Lin, X. (2012) 'Forecasting large scale conditional volatility and covariance using neural network on GPU', *The Journal of Supercomputing*, 63(2), pp. 490–507. doi:10.1007/s11227-012-0827-1.

Campbell, J.Y. and Shiller, R.J. (1988) 'Stock Prices, Earnings, and Expected Dividends', *The Journal of Finance*, 43(3), pp. 661–676. doi:10.1111/j.1540-6261.1988.tb04598.x.

Cao, L. and Tay, F.E.H. (2001) 'Financial Forecasting Using Support Vector Machines', *Neural Computing & Applications*, 10(2), pp. 184–192. doi:10.1007/s005210170010.

Caporin, M. and McAleer, M. (2008) 'Scalar Bekk and indirect DCC', *Journal of Forecasting*, 27(6), pp. 537–549. doi:10.1002/for.1074.

Cappiello, L., Engle, R.F. and Sheppard, K.K. (2003) ‘Asymmetric Dynamics in the correlations of global equity and Bond Returns’, *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.376223.

Carlson, M.A. (2007) ‘A brief history of the 1987 stock market crash with a discussion of the Federal Reserve Response’, *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.982615.

Floros, C & Vougas, D 2007, 'Trading volume and returns relationship in Greek stock index futures market: GARCH vs. GMM', International Research Journal of Finance and Economics, no. 12, pp. 98-115. <http://www.eurojournals.com/irjfe_12_06.pdf>

Chan, W.H. and Maheu, J.M. (2002) ‘Conditional jump dynamics in stock market returns’, *Journal of Business & Economic Statistics*, 20(3), pp. 377–389. doi:10.1198/073500102288618513.

Chen, C.W.S., Watanabe, T. and Lin, E.M.H. (2023) ‘Bayesian estimation of realized GARCH-type models with application to financial tail risk management’, *Econometrics and Statistics*, 28, pp. 30–46. doi:10.1016/j.ecosta.2021.03.006.

Chen, S., Härdle, W.K. and Jeong, K. (2009) ‘Forecasting volatility with support vector machine-based GARCH model’, *Journal of Forecasting*, p. n/a-n/a. doi:10.1002/for.1134.

Chen, X. (2023) ‘Comparing various GARCH-type models in the estimation and forecasts of volatility of S&P 500 returns during Global Finance Crisis of 2008 and covid-19 financial crisis’, *SHS Web of Conferences*, 169, p. 01077. doi:10.1051/shsconf/202316901077.

Chiang, T.C., Qiao, Z. and Wong, W. (2010) ‘New evidence on the relation between return volatility and Trading Volume’, *Journal of Forecasting*, 29(5), pp. 502–515. doi:10.1002/for.1151.

Christensen, K., Siggaard, M. and Veliyev, B. (2022) ‘A machine learning approach to volatility forecasting’, *Journal of Financial Econometrics*, 21(5), pp. 1680–1727. doi:10.1093/jjfinec/nbac020.

Christie, A. (1982) ‘The stochastic behavior of common stock variances Value, leverage and interest rate effects’, *Journal of Financial Economics*, 10(4), pp. 407–432. doi:10.1016/0304-405x(82)90018-6.

Clark, P.K. (1973) ‘A subordinated stochastic process model with finite variance for speculative prices’, *Econometrica*, 41(1), p. 135. doi:10.2307/1913889.

Clements, A. and Preve, D.P.A. (2021) ‘A practical guide to harnessing the har volatility model’, *Journal of Banking & Finance*, 133, p. 106285. doi:10.1016/j.jbankfin.2021.106285.

Cootner, P.H. (1970) *The random character of stock market prices*. Cambridge, Mass: MIT Pr.

Copeland, T.E. (1976) ‘A model of asset trading under the Assumption of Sequential Information Arrival’, *The Journal of Finance*, 31(4), p. 1149. doi:10.2307/2326280.

Corsi, F. (2004) ‘A Simple Long Memory Model of Realized Volatility’, *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.626064.

Cover, T. and Hart, P. (1967a) ‘Nearest neighbor Pattern Classification’, *IEEE Transactions on Information Theory*, 13(1), pp. 21–27. doi:10.1109/tit.1967.1053964.

Cover, T. and Hart, P. (1967b) ‘Nearest neighbor pattern classification’, *IEEE Transactions on Information Theory*, 13(1), pp. 21–27. doi:10.1109/tit.1967.1053964.

Darrat, A.F. and Rahman, S. (1995) ‘Has futures trading activity caused stock price volatility?’, *Journal of Futures Markets*, 15(5), pp. 537–557. doi:10.1002/fut.3990150503.

Darrat, A.F., Zhong, M. and Cheng, L.T.W. (2007) ‘Intraday volume and volatility relations with and without public news’, *Journal of Banking & Finance*, 31(9), pp. 2711–2729. doi:10.1016/j.jbankfin.2006.11.019.

Dickey, D.A. and Fuller, W.A. (1979) ‘Distribution of the estimators for autoregressive time series with a unit root’, *Journal of the American Statistical Association*, 74(366), p. 427. doi:10.2307/2286348.

Diebold, F.X., Gunther, T.A. and Tay, A.S. (1998) ‘Evaluating Density Forecasts with Applications to Financial Risk Management’, *International Economic Review*, 39(4), p. 863. doi:10.2307/2527342.

Ding, Z. and Engle, R.F. (2008) *Large scale conditional covariance matrix modeling, estimation and testing*, *SSRN*. Available at: https://ssrn.com/abstract=1294569 (Accessed: 14 January 2025).

Ding, Z., Granger, C.W.J. and Engle, R.F. (1993) ‘A long memory property of stock market returns and a new model’, *Journal of Empirical Finance*, 1(1), pp. 83–106. doi:10.1016/0927-5398(93)90006-d.

Dixit, J.K. and Agrawal, V. (2019) ‘Foresight for stock market volatility – a study in the Indian perspective’, *foresight*, 22(1), pp. 1–13. doi:10.1108/fs-05-2019-0040.

Dixit, J.K., Agrawal, V. and Agarwal, S. (2022) ‘Measuring stock market volatility - a study in India’s perspective’, *International Journal of Economic Policy in Emerging Economies*, 16(1), p. 1. doi:10.1504/ijepee.2022.124655.

Donaldson, R.G. and Kamstra, M. (1997) ‘An artificial neural network-GARCH model for international stock return volatility’, *Journal of Empirical Finance*, 4(1), pp. 17–46. doi:10.1016/s0927-5398(96)00011-4.

Duffie, D. (2019) ‘Prone to fail: The Pre-Crisis Financial System’, *Journal of Economic Perspectives*, 33(1), pp. 81–106. doi:10.1257/jep.33.1.81.

Echaust, K. and Just, M. (2020) ‘Value at risk estimation using the GARCH-EVT approach with optimal tail selection’, *Mathematics*, 8(1), p. 114. doi:10.3390/math8010114.

Engle, R. (2002) ‘Dynamic Conditional Correlation’, *Journal of Business & Economic Statistics*, 20(3), pp. 339–350. doi:10.1198/073500102288618487.

Engle, R.F. (1982) ‘Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation’, *Econometrica*, 50(4), p. 987. doi:10.2307/1912773.

Engle, R.F. and Kroner, K.F. (1995) ‘Multivariate simultaneous generalized arch’, *Econometric Theory*, 11(1), pp. 122–150. doi:10.1017/s0266466600009063.

Engle, R.F. and Lee, G. g. (1999) ‘A long-run and short-run component model of stock return volatility’, *Cointigration, Causality, and forecating*, pp. 475–497. doi:10.1093/oso/9780198296836.003.0020.

Epps, T.W. and Epps, M.L. (1976) ‘The stochastic dependence of security price changes and transaction volumes: Implications for the mixture-of-distributions hypothesis’, *Econometrica*, 44(2), p. 305. doi:10.2307/1912726.

Eraker, B., Johannes, M. and Polson, N. (2003) ‘The Impact of Jumps in Volatility and Returns’, *The Journal of Finance*, 58(3), pp. 1269–1300. doi:10.1111/1540-6261.00566.

Fama, E.F. (1965) 'The Behavior of Stock-Market Prices', *The Journal of Business*, 38(1), p. 34. doi:10.1086/294743.

Fan, L. *et al.* (2023) 'Forecasting stock volatility during the stock market crash period: The role of hawkes process', *Finance Research Letters*, 55, p. 103839. doi:10.1016/j.frl.2023.103839.

Fang, Y., Yu, P.L.H. and Tang, Y. (2021) *CNN-based realized covariance matrix forecasting*, *arXiv.org*. Available at: https://arxiv.org/abs/2107.10602 (Accessed: 14 January 2025).

Farooq, U. *et al.* (2022) 'How do market capitalization and intellectual capital determine industrial investment?', *Borsa Istanbul Review*, 22(4), pp. 828–837. doi:10.1016/j.bir.2022.05.002.

Fałdziński, M., Fiszeder, P. and Orzeszko, W. (2020) 'Forecasting volatility of energy commodities: Comparison of GARCH models with support vector regression', *Energies*, 14(1), p. 6. doi:10.3390/en14010006.

Figlewski, S. (1997) 'Forecasting Volatility', *Financial Markets, Institutions and Instruments*, 6(1), pp. 1–88. doi:10.1111/1468-0416.00009.

Fiszeder, P., Fałdziński, M. and Molnár, P. (2019) 'Range-based DCC models for covariance and value-at-risk forecasting', *Journal of Empirical Finance*, 54, pp. 58–76. doi:10.1016/j.jempfin.2019.08.004.

Franses, P.H. and Van Dijk, D. (1996) 'Forecasting stock market volatility using (non-linear)

Garch models', *Journal of Forecasting*, 15(3), pp. 229–235. doi:10.1002/(sici)1099-131x(199604)15:3<229::aid-for620>3.0.co;2-3.

French, K.R., Schwert, G.W. and Stambaugh, R.F. (1987) 'Expected stock returns and volatility', *Journal of Financial Economics*, 19(1), pp. 3–29. doi:10.1016/0304-405x(87)90026-2.

Gabriel, A.S. (2012) 'Evaluating the Forecasting Performance of GARCH Models. Evidence from Romania', *Procedia - Social and Behavioral Sciences*, 62, pp. 1006–1010. doi:10.1016/j.sbspro.2012.09.171.

Galai, D. and Masulis, R.W. (1976) 'The option pricing model and the risk factor of stock', *Journal of Financial Economics*, 3(1–2), pp. 53–81. doi:10.1016/0304-405x(76)90020-9.

Gavrishchaka, V.V. and Banerjee, S. (2006) 'Support Vector Machine as an Efficient Framework for Stock Market Volatility Forecasting', *Computational Management Science*, 3(2), pp. 147–160. doi:10.1007/s10287-005-0005-5.

Gavrishchaka, V.V. and Ganguli, S.B. (2003) 'Volatility forecasting from multiscale and high-dimensional market data', *Neurocomputing*, 55(1–2), pp. 285–305. doi:10.1016/s0925-2312(03)00381-3.

Gebka, B. and Wohar, M.E. (2013) 'Causality between trading volume and returns: Evidence from quantile regressions', *International Review of Economics & Finance*, 27, pp. 144–159. doi:10.1016/j.iref.2012.09.009.

Gençay, R. *et al.* (2010) 'Asymmetry of information flow between volatilities across time scales', *Quantitative Finance*, 10(8), pp. 895–915. doi:10.1080/14697680903460143.

Geske, R. (1979) ‘The valuation of compound options’, *Journal of Financial Economics*, 7(1), pp. 63–81. doi:10.1016/0304-405x(79)90022-9.

Girma, P.B. and Mougoué, M. (2002) ‘An empirical examination of the relation between futures spreads volatility, volume, and open interest’, *Journal of Futures Markets*, 22(11), pp. 1083–1102. doi:10.1002/fut.10047.

Global Financial Markets Crash on Black Monday (2019) *Goldman Sachs*. Available at: https://www.goldmansachs.com/our-firm/history/moments/1987-black-monday (Accessed: 07 January 2025).

Glosten, L.R., Jagannathan, R. and 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), pp. 1779–1801. doi:10.1111/j.1540-6261.1993.tb05128.x.

Guldimann, T. (1980) *History of VaR*, *Value at Risk: Monte Carlo Simulation*. Available at: https://valueatrisk.weebly.com/history-of-var.html (Accessed: 17 September 2020).

Hansen, P.R. and Lunde, A. (2005) ‘A forecast comparison of volatility models: does anything beat a GARCH(1,1)?’, *Journal of Applied Econometrics*, 20(7), pp. 873–889. doi:10.1002/jae.800.

Hansen, P.R., Lunde, A. and Nason, J.M. (2010) ‘The model confidence set’, *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.522382.

Harris, L. and Gurel, E. (1986) ‘Price and volume effects associated with changes in the S&P 500 list: New evidence for the existence of price pressures’, *The Journal of Finance*, 41(4), pp. 815–829. doi:10.1111/j.1540-6261.1986.tb04550.x.

Harris, M. and Raviv, A. (1993) ‘Differences of opinion make a horse race’, *Review of Financial Studies*, 6(3), pp. 473–506. doi:10.1093/rfs/5.3.473.

Haykin, S.S. (1999) *Neural networks*. New York: Macmillan College.

Haykin, S.S. (2009) *Neural networks and learning machines*. Upper Saddle River: Prentice Hall.

Hochreiter, S. and Schmidhuber, J. (1997) ‘Long Short-Term Memory’, *Neural Computation*, 9(8), pp. 1735–1780. doi:10.1162/neco.1997.9.8.1735.

Hornik, K., Stinchcombe, M. and White, H. (1989) ‘Multilayer feedforward networks are universal approximators’, *Neural Networks*, 2(5), pp. 359–366. doi:10.1016/0893-6080(89)90020-8.

Hu, M.Y. and Tsoukalas, C. (1999) ‘Combining conditional volatility forecasts using neural networks: an application to the EMS exchange rates’, *Journal of International Financial Markets, Institutions and Money*, 9(4), pp. 407–422. doi:10.1016/s1042-4431(99)00015-3.

Huang, Y. (2009) ‘Advances in artificial neural networks – methodological development and application’, *Algorithms*, 2(3), pp. 973–1007. doi:10.3390/algor2030973.

Huang, Y., Su, W. and Li, X. (2010) ‘Comparison of BEKK GARCH and DCC GARCH MODELS: An empirical study’, *Lecture Notes in Computer Science*, pp. 99–110. doi:10.1007/978-3-642-17313-4_10.

Huang, Z., Liu, H. and Wang, T. (2016) ‘Modeling long memory volatility using realized measures of volatility: A realized Har Garch model’, *Economic Modelling*, 52, pp. 812–821.

doi:10.1016/j.econmod.2015.10.018.

Hyup Roh, T. (2007a) ‘Forecasting the volatility of stock price index’, *Expert Systems with Applications*, 33(4), pp. 916–922. doi:10.1016/j.eswa.2006.08.001.

Hyup Roh, T. (2007b) ‘Forecasting the volatility of stock price index’, *Expert Systems with Applications*, 33(4), pp. 916–922. doi:10.1016/j.eswa.2006.08.001.

Jiahong Li, Hui Bu and Junjie Wu (2017) ‘Sentiment-aware stock market prediction: A deep learning method’, *2017 International Conference on Service Systems and Service Management*, pp. 1–6. doi:10.1109/icsssm.2017.7996306.

Jones, C.M., Kaul, G. and Lipson, M.L. (1994) ‘Transactions, Volume, and Volatility’, *Review of Financial Studies*, 7(4), pp. 631–651. doi:10.1093/rfs/7.4.631.

Kambouroudis, D.S. (2012) *Essays on volatility forecasting*. Available at: https://hdl.handle.net/10023/3191 (Accessed: 13 January 2025).

Kambouroudis, D.S. and McMillan, D.G. (2015) ‘Does Vix or volume improve GARCH volatility forecasts?’, *Applied Economics*, 48(13), pp. 1210–1228. doi:10.1080/00036846.2015.1096004.

Karpoff, J.M. (1987) ‘The relation between price changes and Trading Volume: A survey’, *The Journal of Financial and Quantitative Analysis*, 22(1), p. 109. doi:10.2307/2330874.

Kenton, W. (2020) *What is the GARCH process? how it’s used in different forms*, *Investopedia*. Available at: https://www.investopedia.com/terms/g/generalalizedautogregressiveconditionalheteroskedasti city.asp (Accessed: 19 January 2025).

Kim, S.W. and Rogers, J.H. (1995) ‘International stock price spillovers and market liberalization: Evidence from Korea, Japan, and the United States’, *Journal of Empirical Finance*, 2(2), pp. 117–133. doi:10.1016/0927-5398(94)00013-7.

Klaassen, F. (2002) ‘Improving GARCH volatility forecasts with regime-switching GARCH’, *Empirical Economics*, 27(2), pp. 363–394. doi:10.1007/s001810100100.

Kristjanpoller, W., Fadic, A. and Minutolo, M.C. (2014) ‘Volatility forecast using hybrid Neural Network models’, *Expert Systems with Applications*, 41(5), pp. 2437–2442. doi:10.1016/j.eswa.2013.09.043.

Kroner, K.F. (1996) *Creating and using volatility forecasts*. [San Francisco, CA].

Le, V. and Zurbruegg, R. (2010) ‘The role of trading volume in volatility forecasting’, *Journal of International Financial Markets, Institutions and Money*, 20(5), pp. 533–555. doi:10.1016/j.intfin.2010.07.003.

Ledoit, O. and Wolf, M. (2004) ‘A well-conditioned estimator for large-dimensional covariance matrices’, *Journal of Multivariate Analysis*, 88(2), pp. 365–411. doi:10.1016/s0047-259x(03)00096-4.

Lee, W. and Chun, H. (2016) ‘A deep learning analysis of the Chinese Yuan’s volatility in the onshore and offshore markets’, *Journal of the Korean Data and Information Science Society*, 27(2), pp. 327–335. doi:10.7465/jkdi.2016.27.2.327.

LeRoy, S.F. and Porter, R.D. (1981) ‘The Present-Value Relation: Tests Based on Implied

Variance Bounds', *Econometrica*, 49(3), p. 555. doi:10.2307/1911512.

Li, N. *et al.* (2009) 'Network environment and financial risk using machine learning and sentiment analysis', *Human and Ecological Risk Assessment: An International Journal*, 15(2), pp. 227–252. doi:10.1080/10807030902761056.

Liu, H.-C. and Hung, J.-C. (2010) 'Forecasting S&P-100 stock index volatility: The role of volatility asymmetry and distributional assumption in GARCH models', *Expert Systems with Applications*, 37(7), pp. 4928–4934. doi:10.1016/j.eswa.2009.12.022.

Liu, M., Lee, C. and Choo, W. (2020) 'An empirical study on the role of trading volume and data frequency in volatility forecasting', *Journal of Forecasting*, 40(5), pp. 792–816. doi:10.1002/for.2739.

Liu, W.K. and So, M.K. (2020) 'A GARCH model with Artificial Neural Networks', *Information*, 11(10), p. 489. doi:10.3390/info11100489.

Liu, Y.A. and Pan, M.-S. (1997) 'Mean and Volatility Spillover Effects in the U.S. and Pacific–Basin Stock Markets', *Multinational Finance Journal*, 1(1), pp. 47–62. doi:10.17578/1-1-3.

Mademlis, D.K. and Dritsakis, N. (2021) 'Volatility forecasting using hybrid GARCH neural network models: The case of the italian stock market', *International Journal of Economics and Financial Issues*, 11(1), pp. 49–60. doi:10.32479/ijefi.10842.

Mandelbrot, B. (1963) 'The Variation of Certain Speculative Prices', *The Journal of Business*, 36(4), p. 394. doi:10.1086/294632.

Markowitz, H. (1952) 'Portfolio Selection', *The Journal of Finance*, 7(1), p. 77.

doi:10.2307/2975974.

Mastro, D. (2014) 'Forecasting realized volatility: Arch-type models vs. the Har-RV model', *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.2519107.

McAleer, M. and Medeiros, M.C. (2008) 'Realized volatility: A Review', *Econometric Reviews*, 27(1–3), pp. 10–45. doi:10.1080/07474930701853509.

McCulloch, W.S. and Pitts, W. (1943) 'A logical calculus of the ideas immanent in nervous activity', *Bulletin of mathematical biophysics*, 5, pp. 115–133.

McMillan, D.G. and Speight, A.E.H. (2004) 'Daily volatility forecasts: reassessing the performance of GARCH models', *Journal of Forecasting*, 23(6), pp. 449–460. doi:10.1002/for.926.

Merton, R.C. (1974) 'ON THE PRICING OF CORPORATE DEBT: THE RISK STRUCTURE OF INTEREST RATES*', *The Journal of Finance*, 29(2), pp. 449–470. doi:10.1111/j.1540-6261.1974.tb03058.x.

Merton, R.C. (1980) 'On estimating the expected return on the market', *Journal of Financial Economics*, 8(4), pp. 323–361. doi:10.1016/0304-405x(80)90007-0.

Michańków, J., Kwiatkowski, Ł. and Morajda, J. (2023) 'Combining deep learning and GARCH models for financial volatility and risk forecasting', *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.4589950.

Miron, D. and Tudor, C. (2010) 'Asymmetric Conditional Volatility Models: Empirical Estimation and Comparison of Forecasting Accuracy', *Journal for Economic Forecasting,* *Institute for Economic Forecasting*, 0(3).

Mishkin, F.S. (1999) ‘International experiences with different monetary policy regimes). Any views expressed in this paper are those of the author only and not those of Columbia University or the National Bureau of Economic Research.’, *Journal of Monetary Economics*, 43(3), pp. 579–605. doi:10.1016/s0304-3932(99)00006-9.

Nelson, D.B. (1991) ‘Conditional Heteroskedasticity in Asset Returns: A New Approach’, *Econometrica*, 59(2), p. 347. doi:10.2307/2938260.

Ni, J. and Xu, Y. (2021) ‘Forecasting the dynamic correlation of stock indices based on Deep Learning Method’, *Computational Economics*, 61(1), pp. 35–55. doi:10.1007/s10614-021-10198-3.

Niu, H. and Wang, J. (2013) ‘Volatility clustering and long memory of financial time series and financial price model’, *Digital Signal Processing*, 23(2), pp. 489–498. doi:10.1016/j.dsp.2012.11.004.

Nugroho, D.B. *et al.* (2019) ‘Empirical performance of GARCH, GARCH-M, Gjr-GARCH and log-GARCH models for returns volatility’, *Journal of Physics: Conference Series*, 1307(1), p. 012003. doi:10.1088/1742-6596/1307/1/012003.

Oskay, B. *et al.* (2017) ‘Advancing Financial Risk Estimates by Incorporating GARCH Into Value at Risk’, *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.3137587.

Pagan, A.R. and Schwert, G.W. (1990) ‘Alternative models for conditional stock volatility’, *Journal of Econometrics*, 45(1–2), pp. 267–290. doi:10.1016/0304-4076(90)90101-x.

Papadimitriou, T., Gogas, P. and Athanasiou, A.F. (2020) ‘Forecasting S&P 500 spikes: An SVM approach’, *Digital Finance*, 2(3–4), pp. 241–258. doi:10.1007/s42521-020-00024-0.

Patton, A.J. (2005) ‘Volatility Forecast Evaluation and Comparison Using Imperfect Volatility Proxies’, *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.675322.

Patton, A.J. (2011) ‘Volatility forecast comparison using imperfect volatility proxies’, *Journal of Econometrics*, 160(1), pp. 246–256. doi:10.1016/j.jeconom.2010.03.034.

Poon, S.-H. and Granger, C.W.J. (2003) ‘Forecasting Volatility in Financial Markets: A Review’, *Journal of Economic Literature*, 41(2), pp. 478–539. doi:10.1257/002205103765762743.

Pérez-cruz, F., Afonso-rodríguez, J.A. and Giner, J. (2003) ‘Estimating GARCH models using support vector machines*’, *Quantitative Finance*, 3(3), pp. 163–172. doi:10.1088/1469-7688/3/3/302.

Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986) ‘Learning representations by back-propagating errors’, *Nature*, 323(6088), pp. 533–536. doi:10.1038/323533a0.

Sapuric, S., Kokkinaki, A. and Georgiou, I. (2020) ‘The relationship between Bitcoin returns, volatility and volume: Asymmetric Garch Modeling’, *Journal of Enterprise Information Management*, 35(6), pp. 1506–1521. doi:10.1108/jeim-10-2018-0228.

Schwert, G.W. (1989) ‘Why Does Stock Market Volatility Change Over Time?’, *The Journal of Finance*, 44(5), pp. 1115–1153. doi:10.1111/j.1540-6261.1989.tb02647.x.

Sebastião, H. and Godinho, P. (2021) ‘Forecasting and trading cryptocurrencies with machine learning under changing market conditions', *Financial Innovation*, 7(1). doi:10.1186/s40854-020-00217-x.

Sentana, E. (1995) 'Quadratic ARCH Models', *The Review of Economic Studies*, 62(4), pp. 639–661. doi:10.2307/2298081.

Shalen, C.T. (1993) 'Volume, volatility, and the dispersion of beliefs', *Review of Financial Studies*, 6(2), pp. 405–434. doi:10.1093/rfs/6.2.405.

Sharma, S., Aggarwal, V. and Yadav, M.P. (2021) 'Comparison of linear and non-linear GARCH models for forecasting volatility of select emerging countries', *Journal of Advances in Management Research*, 18(4), pp. 526–547. doi:10.1108/jamr-07-2020-0152.

Sheikh, A.Z. and Qiao, H. (2009) 'Non-normality of market returns: *a framework for Asset Allocation Decision making*', *The Journal of Alternative Investments*, 12(3), pp. 8–35. doi:10.3905/jai.2010.12.3.008.

Shiller, R.J. (1979) 'The Volatility of Long-Term Interest Rates and Expectations Models of the Term Structure', *Journal of Political Economy*, 87(6), pp. 1190–1219. doi:10.1086/260832.

Silvennoinen, A. and Terasvirta, T. (2008) 'Multivariate garch models', *SSRN Electronic Journal* [Preprint]. doi:10.2139/ssrn.1148139.

Srivastava, N. *et al.* (2014) 'Dropout: a simple way to prevent neural networks from overfitting', *The Journal of Machine Learning Research*, 15(1).

Stevenson, A. (2020) *Froth returns to China's stock market, echoing the 2015 crisis*, *The New York Times*. Available at: https://www.nytimes.com/2020/07/22/business/china-stock-market.html (Accessed: 07 January 2025).

Sturesson, E. and Wennström, A. (2023) *Volatility forecasting - A comparative study of different forecasting models.*, *Handle Proxy*. Available at: https://hdl.handle.net/2077/77548 (Accessed: 19 January 2025).

Sun, H. and Yu, B. (2019) ‘Forecasting financial returns volatility: A GARCH-SVR model’, *Computational Economics*, 55(2), pp. 451–471. doi:10.1007/s10614-019-09896-w.

Takahashi, M., Watanabe, T. and Omori, Y. (2024) ‘Forecasting daily volatility of stock price index using daily returns and realized volatility’, *Econometrics and Statistics*, 32, pp. 34–56. doi:10.1016/j.ecosta.2021.08.002.

Tauchen, G.E. and Pitts, M. (1983) ‘The price variability-volume relationship on speculative markets’, *Econometrica*, 51(2), p. 485. doi:10.2307/1912002.

Taylor, S.J. (1987) ‘Forecasting the volatility of currency exchange rates’, *International Journal of Forecasting*, 3(1), pp. 159–170. doi:10.1016/0169-2070(87)90085-9.

Tsay, R.S. (2005) ‘Analysis of Financial Time Series’, *Wiley Series in Probability and Statistics* [Preprint]. doi:10.1002/0471746193.

Tu, J.V. (1996a) ‘Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes’, *Journal of Clinical Epidemiology*, 49(11), pp. 1225–1231. doi:10.1016/s0895-4356(96)00002-9.

Van der Weide, R. (2002) ‘Go-GARCH: A multivariate generalized orthogonal Garch model’, *Journal of Applied Econometrics*, 17(5), pp. 549–564. doi:10.1002/jae.688.

Vinod, N. and Geoffrey, H. (2010) ‘Rectified Linear Units Improve Restricted Boltzmann Machines Vinod Nair’, *Proceedings of ICML*, 27(0), pp. 807–814.

Werner, G.A. and Hanka, L. (2016) ‘Tuning an artificial neural network to increase the efficiency of a fingerprint matching algorithm’, *2016 IEEE 14th International Symposium on Applied Machine Intelligence and Informatics (SAMI)*, pp. 105–109. doi:10.1109/sami.2016.7422990.

Winter, G. (2019) ‘Machine learning in Healthcare’, *British Journal of Healthcare Management*, 25(2), pp. 100–101. doi:10.12968/bjhc.2019.25.2.100.

Xiong, R., Nichols, E.P. and Shen, Y. (2016) *Deep Learning Stock Volatility with Google Domestic Trends*, *arXiv.org*. Available at: https://arxiv.org/abs/1512.04916 (Accessed: 18 September 2020).

Xu, Y. *et al.* (2024) ‘Liquidity and realized volatility prediction in Chinese stock market: A time-varying transitional dynamic perspective’, *International Review of Economics & Finance*, 89, pp. 543–560. doi:10.1016/j.iref.2023.07.083.

Zahid, M. *et al.* (2022) ‘Modeling and forecasting the realized volatility of bitcoin using realized Har-garch-type models with jumps and inverse leverage effect’, *Sains Malaysiana*, 51(3), pp. 929–942. doi:10.17576/jsm-2022-5103-25.

Zakoian, J.-M. (1994) ‘Threshold heteroskedastic models’, *Journal of Economic Dynamics and Control*, 18(5), pp. 931–955. doi:10.1016/0165-1889(94)90039-6.

Zhu, X. *et al.* (2008) ‘Predicting stock index increments by Neural Networks: The role of trading volume under Different Horizons’, *Expert Systems with Applications*, 34(4), pp. 3043–3054. doi:10.1016/j.eswa.2007.06.023.

8. Appendix

MAE statistics for models with/without consideration of trading volume

	GARCH	GARCH*	EGARCH	EGARCH*	TGARCH	TGARCH*	CGARCH	CGARCH*	HAR	HAR*
USA	0.48441	0.43326	0.51465	0.45318	0.49894	0.43405	0.44456	0.41250	0.46697	0.42125
UK	0.53067	0.45728	0.52384	0.45327	0.51752	0.45251	0.48433	0.42531	0.48942	0.45068
Switzerland	0.61921	0.63337	0.67593	0.66520	0.62578	0.64435	0.55404	0.56155	0.58010	0.58640
Korea	0.43238	0.36336	0.47351	0.43432	0.49277	0.45437	0.39475	0.35315	0.45634	0.42653
Japan	1.27871	1.21633	1.22068	1.15255	1.20895	1.14640	1.19354	1.12365	1.25448	1.12176
India	0.52288	0.53334	0.61883	0.63221	0.62880	0.64234	0.48621	0.49324	0.52174	0.53514
HK	0.88333	0.81255	0.95334	0.86303	0.92955	0.85519	0.81122	0.74537	0.89501	0.82350
German	0.77622	0.81602	0.89640	0.93166	0.86955	0.91414	0.69921	0.76203	0.74994	0.78363
France	0.72936	0.66372	0.81053	0.73758	0.75651	0.68842	0.65132	0.59270	0.70659	0.64300
Shanghai	1.69475	1.55213	1.66042	1.52359	1.68216	1.54358	1.63767	1.50556	1.65563	1.52313
Canada	0.32501	0.28501	0.36212	0.31856	0.36086	0.31356	0.29920	0.25323	0.32267	0.28324
Australia	0.39618	0.33354	0.41877	0.35332	0.41015	0.35183	0.35537	0.30264	0.37577	0.32521

MSE statistics for models with/without consideration of trading volume

	GARCH	GARCH*	EGARCH	EGARCH*	TGARCH	TGARCH*	CGARCH	CGARCH*	HAR	HAR*
USA	1.73871	1.85436	1.74640	1.69152	1.73445	1.65236	1.66423	1.54236	1.58341	1.46541
UK	3.92500	3.21029	3.89797	2.99354	3.89980	3.15236	3.83667	3.08390	3.66762	2.95054
Switzerland	2.76779	2.25437	3.17322	2.50264	2.58144	2.54625	2.52300	2.56230	2.04765	2.27252
Korea	0.32819	0.32405	0.40103	0.37466	0.41606	0.38862	0.29151	0.30655	0.35112	0.33728
Japan	4.08200	3.58608	3.54615	3.32427	3.52607	3.32421	3.68236	3.48154	3.76580	3.57455
India	0.56489	0.57164	0.78016	0.80356	0.75729	0.78001	0.51944	0.52402	0.56014	0.57001
HK	1.47589	1.35782	1.67140	1.53769	1.59425	1.46671	1.26938	1.16783	1.48275	1.36413
German	1.53319	1.61685	1.90869	2.21313	1.79847	1.86632	1.32796	1.37633	1.33551	1.53669
France	1.60642	1.53543	1.90909	1.64890	1.66952	1.35618	1.34897	1.16506	1.37202	1.31653
China	11.05978	10.82537	10.68908	10.33122	10.95257	10.42657	10.47770	10.45756	10.76963	10.34665
Canada	1.13559	1.12562	1.14506	1.11071	1.13827	1.10412	1.09928	1.05530	1.08457	1.05403
Australia	0.34851	0.32108	0.36462	0.35537	0.35119	0.34253	0.30285	0.27661	0.29552	0.27262