Essays on Econometric Models of Volatility

Yushuang Jiang

Thesis submitted in partial fulfilment of the requirements

for the degree of Doctor of Philosophy

ICMA Centre
Henley Business School
University of Reading

Supervisor: Dr. Emese Lazar

June 2021

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Yushuang Jiang

Acknowledgements

First and foremost, I am would like to express my deepest gratitude to my advisor, Dr Emese Lazar, for all the patience, encouragement and continuous guidance she has provided throughout my PhD study. Emese gives me very valuable opinions in research and a very flexible working environment so I could enjoy the research whilst raising two little boys. To me, she is not only a great supervisor but also a good coauthor, mentor, friend. She demonstrates how to be a good supervisor and, more importantly, be a good person. I would never finish this thesis without her help.

I would like to thank the examiners, Dr Alfonso Dufour and Dr Alexandru Badescu, for their precious time to review this thesis, as well as for their valuable comments. I appreciate my friends, and colleagues for being very kind and supportive. Also, my sincere thanks go to ICMA Centre for its generous financial support.

Most importantly, I am very grateful to my parents, my brother for their support, especially during COVID-19 pandemic. They give me an opportunity to do research even the nursery is closed. Finally, I would like to thank Yanfeng and two little cheeky monkeys. Thanks them for being there for me all the time!

Abstract

This thesis contributes to the literature on volatility forecasting, focusing on the VIX index, the VIX futures and the VVIX. It consists of three main chapters.

The first contribution is the introduction of a new VIX forecasting methodology employing both filtered historical simulations and four well-established indices. We examine the forecasting performance of three different GARCH models from 2011-2017. Our empirical results show that this new method outperforms the benchmark model which only uses the VIX index and assumes a normal distribution. Also, our proposed methodology is found to reduce the computational time significantly, compared to the traditional model which uses cross-sectional options prices.

The second contribution is studying the role of the VIX term structure in predicting VIX futures prices. The estimation is carried out under the GJR model, assuming the empirical innovation density under the risk-neutral measure. Several models are employed differing in the data set used, i.e., futures data, or the VIX term structure, or their combinations. We find that the use of the VIX

term structure improves the VIX futures forecasts, especially for the long-term VIX futures or when the VIX level is high. Also, the evidence from the 2020 COVID-19 crisis shows that using both the VIX term structure and the VIX futures provides lower pricing errors compared to using futures data only.

The third contribution is an investigation on the optimal forecasts of the VVIX. This thesis presents a comparison of VVIX forecasts based on three individual models, eight combining methods and two LASSO-type regressions. Our finding is that the simple median combining method gives the lowest forecasting errors across the years among all the methods considered. Moreover, the model selection results of LASSO suggest that instead of daily changes in the VVIX, the changes in monthly VVIX are essential to predict the VVIX.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivations

Stock market volatility plays a critical role in portfolio optimization, asset pricing and risk management. In 1993, the Chicago Board Options Exchange (CBOE) introduced the first volatility index, VIX, which offers a theoretical estimate of the market's future volatility. The VIX index reflects the expected volatility of the $S\&P$ 500 index over the coming 30 days and is calculated from a panel of option prices. The VIX index is often referred as the 'fear gauge' (see Whaley, 2009). In general, high VIX levels reflect the fear that the equity prices will decrease in the future, while low VIX levels mirror greed among the investors and thus increasing the likelihood of a market correction. In addition to the VIX, the CBOE also established a set of volatility indices across different maturities to measure the implied volatility term structure: the CBOE S&P 500 9-day Volatility Index

(VIX9D), the CBOE S&P 500 3-month Volatility Index (VIX3M), the CBOE S&P 500 6-month Volatility Index (VIX6M) and the CBOE S&P 500 1-year Volatility Index (VIX1Y).

Notably, the VIX and the other volatility indices are not easily traded, although theoretically it is possible to replicate a portfolio of the $S\&P$ 500 options in the indices. To enable trading and hedging against changes in volatility, CBOE launched VIX futures on March 26, 2004. Since its creation, VIX futures have attracted a considerable amount of attention in past years given the fact that they are very liquid in the market. The daily trading volume exceeds 200,000 contracts in 2020 and corresponds roughly to 6 billion USD in market value[1]. Also, the VIX futures is a much more convenient hedging tool than S&P500 index options (see Szado, 2009).

To guide and inform the increasing number of investors in VIX derivatives, CBOE published the volatility-of-volatility index, VVIX, on 14 March, 2012. The VVIX reflects the risk-neutral volatility of volatility (vol-of-vol) using the same methodology as the VIX, implied from VIX options instead of S&P 500 options. It measures how market volatility varies in the future rather than measures the volatility itself. The VVIX index has separate dynamics from the VIX and is an important risk factor that affects the level of the VIX and VIX option returns (see Huang et al., 2019a). Moreover, it conveys information to the VIX trading

[1]The data can be found from the CBOE website.

community about the fair values of VIX futures[2].

The focus of this thesis is on forecasting volatility, as an index or a derivative instrument, using several well-established volatility indices. A number of studies suggest that the predictive ability of the option implied volatility, such as the VIX, outperforms traditional time-series volatility models based on historical observations, see, for example, Corrado and Miller (2005), Carr and Wu (2006), Bandi and Perron (2006), etc. Also, forecasting the VIX index is essential for trading strategies based on VIX futures and options either for trading volatility or for hedging purposes (Konstantinidi et al., 2008; Carr and Lee, 2009; Fernandes et al., 2014). The traditional volatility forecasting literature is based on the assumption of normality for return innovations and extracts information from extensive options data. Alternatively, Chapter 2 proposes a faster VIX forecasting method that assumes filtered historical density and uses different volatility indices.

Given the fact that the VIX futures are the most liquid in the volatility futures market as discussed by Konstantinidi and Skiadopoulos (2011a), there are numerous studies examining the VIX futures. Most studies concentrate on developing VIX futures pricing models: either with the underlying, i.e, the VIX index, see, for example, Zhang and Zhu (2006), Zhu and Lian (2012), Xie et al. (2020), etc; or using both the volatility indices and historical futures data, see,

[2]See CBOE VVIX whitepaper for more details. https://ww2.cboe.com/index/dashboard/vvix#vvix-overview

for example, Wang et al. (2017), Huang et al. (2019b). Chapter 3 investigates to what extent the VIX term structure can help to predict the VIX futures prices. More importantly, as in Chapter 2, we assume non-normal return innovations using filtered historical simulation in volatility forecasting.

Furthermore, the accuracy in forecasting the VVIX index is critical to capture the future tendency of both the VIX index and the VIX futures prices (Lin, 2007). Also, incorporating the VVIX into models can significantly enhance the predictive power compared to traditional volatility models (Jeon et al., 2020). However, the existing literature mainly employs the VVIX as a proxy to study the characteristic of the implied volatility-of-volatility, see, for example, Park (2015), Hollstein and Prokopczuk (2018), Huang et al. (2019a). In Chapter 4, we seek to answer a simple question: is there an optimal forecasting method for the VVIX index? It is well-known in the forecasting literature that forecast combinations often outperform individual models (Becker and Clements, 2008; Patton and Sheppard, 2009; Wang et al., 2016). On the other hand, the least absolute shrinkage and selection operator (LASSO) of Tibshirani (1996) is a desirable model when generating financial forecasts as argued by Audrino and Knaus (2016) and Zhang et al. (2019a). Therefore, to answer the above question, we compare the VVIX forecasting performance of thirteen different models across three categories, i.e., individual models, models combinations and LASSO-type models.

Overall, the evolution of the VIX in recent years, alongside the following launched VIX futures and the VVIX index, indicate a strong demand by financial participants for volatility-related products. Therefore, forecasting VIX and its related instruments are of great interest to both academic researchers and financial practitioners. To the best of our knowledge, the literature on forecasting volatility, which employs the volatility indices and assumes filtered historical returns, is limited. The ultimate goal of this thesis is to analyse the forecasting models for the VIX and its related products using the volatility indices.

1.2 Overview of the Thesis

This thesis discusses forecasting models for the VIX, the VIX futures and the VVIX.

Firstly, this thesis proposes a new method to forecast the VIX which uses filtered historical simulation (FHS) proposed by Barone-Adesi et al. (2008). The non-normality of financial returns is well documented in the literature since Mandelbrot (1963). On the other hand, in a discrete-time setting, the Generalized Autoregressive Conditional Heteroscedasticity (GARCH) model introduced by Engle (1982) and Bollerslev (1986) and its various extensions are very popular in modelling volatility, as they can explain the volatility clustering and are easy to estimate. Barone-Adesi et al. (2008) propose a new method to model the future volatility which captures the non-normality of returns using FHS under the GARCH framework of Glosten et al. (1993) (GJR). However, the estimation of such models uses a large number of cross-sections of option data and thus is computationally intensive. Alternatively, recent papers show that estimating GARCH models using VIX index information improves model performance and saves computational time (e.g. Kanniainen et al., 2014).

In this thesis, we propose a new volatility forecasting approach using the filtered historical simulation and four well-established volatility indices, i.e., VIX9D, VIX, VIX3M and VIX6M. We estimate three different GARCH models: the classic GARCH(1,1) model of Bollerslev (1986), the non-linear asymmetric GARCH model of Engle and Ng (1993) (NAGARCH) and the GJR model by Glosten et al. (1993) in order to capture the leverage effect. Also, we choose the model of Hao and Zhang (2013) as the benchmark, which assumes normal returns and only includes information on the 1-month VIX index. As robustness checks, we perform the following: 1) examine our results across different forecasting horizons; 2) consider alternative weights in the optimisation function; 3) calculate pricing error statistics using different weighting approaches; and 4) extend the sample period to include the 2008 financial crisis. In addition, we compare the computational time of the proposed approach with options-based calibration as in Barone-Adesi et al. (2008).

Our empirical analysis shows that the proposed estimation method outperforms the model of Hao and Zhang (2013) both in-sample and out-of-sample. The NAGARCH model under the new method is superior to all the other models for both one-week-ahead and four-week-ahead VIX forecasts, while the GJR model based on the new method dominates for one-day-ahead forecasts. Additionally, the use of volatility indices significantly reduces the computational burden compared to the option-based pricing method.

Secondly, this thesis develops a VIX futures evaluation model using volatility indices that fills the gap between the VIX futures pricing literature and the VIX term structure literature. We employ the GJR model, which is calibrated from the data using filtered historical simulation, to model the daily volatility. To explore the effects of the VIX term structure on the performance of VIX futures pricing models, we examine the forecasting performance of including data on the VIX term structure. Given the two sources of information, i.e., the VIX futures data and the VIX term structure (VIX9D, VIX, VIX3M, VIX6M and VIX1Y), seven estimation methods are presented that differ in terms of the data used. We find that the out-of-sample performance of the models that use the VIX term structure and the VIX futures is not significantly different from the model that uses futures data, but provides significant outperformance compared to the models which are based on the VIX term structure.

Also, we perform the model confidence set procedure of Hansen et al. (2011) for a detailed comparison of the pricing performance based on different time to maturity and the levels of the VIX index. An impressive finding is that the use of the VIX term structure improves the VIX futures forecasting when the VIX level is higher than 15 or with a maturity longer than 120 days. Meanwhile, the evidence of the 2020 COVID-19 pandemic confirms that the addition of the VIX term structure lowers the pricing errors when the market is volatile.

Furthermore, we apply the model of Xie et al. (2020) as the benchmark which employs the GJR model assuming normally distributed innovations and is estimated using the VIX index. Our empirical results suggest that the proposed pricing models based on the filtered historical simulations significantly outperform the benchmark both in-sample and out-of-sample.

Thirdly, given the importance of the VVIX, this thesis endeavours to answer the following question: is there an optimal forecasting method for the VVIX? To answer this question, we employ three common models used in volatility forecasting: the linear regression, autoregressive–moving-average (ARMA) model and the heterogeneous autoregressive (HAR) model of Corsi (2009). Then eight popular combining methods are implemented based on these three individual models to generate the VVIX forecasts (see Rapach et al., 2010; Hsiao and Wan, 2014). Also, we consider the original LASSO proposed by Tibshirani (1996) and the elastic net of Zou and Hastie (2005) in the comparison, which results in thirteen forecasting models in total.

Among all the models we consider, our empirical analysis shows that the median combining method performs the best by providing the lowest squared errors of the forecasts over the full sample period. Importantly, the results on

LASSO-type models reveal that the daily changes in average monthly VVIX play an important role in the forecasting of VVIX .

1.3 Original Contributions

The forecasting of the volatility of financial time series plays a critical role in asset allocation and risk management. With the growing uses of the VIX and its related products, forecasting the volatility indices becomes essential but challenging. This thesis, which contains three main chapters, contributes to forecasting the VIX index, the VIX futures and the VVIX, respectively.

(1) Our original contributions in forecasting the VIX index include:

- We propose the use of GARCH models with filtered historical simulations in the VIX forecasting literature to capture the non-normal features of returns data.
- We allow for flexible change of measure in the model, i.e., different parameters under the physical and risk-neutral volatility process.
- We compare three different models based on GARCH specifications: original GARCH, GJR GARCH and NAGARCH.
- Instead of using cross-sectional options data, we consider another forward-looking information in our estimation, which are the CBOE volatility indices, i.e., VIX9D, VIX, VIX3M and VIX6M.

- We demonstrate that the computational time of proposed method is reduced significantly compared to option-based models.

(2) Secondly, our original contributions to the literature on forecasting the VIX futures prices include:

- We develop a model estimation method for VIX futures prices by incorporating the volatility indices.
- We show that the addition of the VIX term structure improves the forecasting performance, especially for the long-term VIX futures or when the level of the VIX is high.
- Differently from the majority of literature which assumes a normal distribution for the returns, we apply the empirical innovation density extracted from historical returns.
- We take the VVIX term structure into account when modelling the VIX futures prices.

(3) Thirdly, our original contributions to the literature on forecasting the VVIX include:

- We examine the daily behaviour of the VVIX time series.
- We compare thirteen different forecasting models/methods, which belong to three categories: individual models, forecast combinations and LASSO-type models.

- We analyse the model selection results of LASSO-type models.

1.4 Outline of the Thesis

The rest of this thesis is organized as follows: Chapter 2 introduces a new VIX forecast method using GARCH models based on the filtered historical simulation; Chapter 3 investigates the effects of the VIX term structure on the performance of VIX futures pricing models; Chapter 4 presents the forecasting performance across several different forecasting methods/models. Chapter 5 summarises our main findings and discusses further research that builds on the findings presented in this thesis.

For a better reading experience, we make each chapter self-contained. As such, we (re)introduce variables and abbreviations in each chapter. Whenever possible, we attempt to follow consistent notations throughout this thesis.

Notes

[1]Please see Clemen (1989), Clements and Hendry (2004) and Timmermann (2006) for reviews of forecast combinations.

[2]Konstantinidi and Skiadopoulos (2011b) suggest that the slope of yield curve has predicative power for the VIX futures market, hence we also take this variable into consideration. We examine the yield curve slope within different maturities; however, the estimated regression shows that the information on the yield curve does not explain the daily changes in the VVIX index.

[3]We compare the two regressions, i.e, the regression using VVIX and the regression using the changes. Interestingly, all the coefficients in the regression using changes are significantly different from zero at 5%, while only the coefficient of lagged daily VVIX is significant in the regression using the VVIX index.

[4]In this study, we use the trading day count convention. Hence the weekly and monthly VVIX levels are calculated as the average values over the past 5 and 22 days, respectively.

[5]We also perform the max-min normalisation to scale the features; the results are similar and available on request.

[6]See Patton (2011a) for a range of loss functions which are employed in the literature of volatility forecast evaluation.

[7]Patton (2011a) shows that, among all the loss functions, only MSE and QLIKE are robust to the noise in the volatility proxy.

Chapter 2

Forecasting VIX using filtered historical simulation

2.1 Introduction

There is substantial empirical research showing that volatility clustering plays an important role in modelling financial time series, such as equity returns. The Generalized Autoregressive Conditional Heteroscedasticity (GARCH) framework introduced by Engle (1982) and Bollerslev (1986) allows the volatility to be time-varying – initially assuming normally distributed innovations. However, the non-normality of the return innovations is well documented in the finance literature since Mandelbrot (1963). Consequently, GARCH models with non-normal innovations (assuming more flexible distributions such as the student's t or the generalised error distribution) gained popularity - see, for example, Bollerslev (1987) and Nelson (1991). Other approaches can be found in Christoffersen et al. (2006), Stentoft (2008) and Christoffersen et al. (2009). The recent option pricing literature captures the non-normality of returns by employing filtered historical simulation (FHS) as in Barone-Adesi et al. (2008), where the empirical innovation density is extracted from historical index returns, and these methods can be used in volatility forecasting. Nonetheless, the estimation of such models uses cross-sectional option prices and is computationally intensive.

In this chapter, we propose an alternative, faster approach to forecast volatility, which uses volatility indices information. However, our approach is based on not only the 1-month VIX index, but the VIX indices at all available maturities (9 days, 1 month, 3 months and 6 months)[1], and employs filtered historical returns. VIX9D, VIX3M and VIX6M measure the expected annualised volatility in the coming days as well as the VIX index, although with different maturities, they are informative for future VIX values. We provide evidence that our approach outperforms the Normal-VIX model of Hao and Zhang (2013) both in-sample and out-of-sample and leads to a significant reduction of computational time when compared with the model of Barone-Adesi et al. (2008).

The traditional way to estimate GARCH parameters is via maximum like-

[1] The new indices (VIX9D, VIX3M and VIX6M) are derived by applying the VIX algorithm to options on the S & P 500 Index, and use SPX options with expiration dates that bracket a different period of time, e.g., the VIX9D is calculated using two 'near-term' option contracts in which one has maturity less than 9 days and the other has maturity more than 9 days.

lihood estimation (MLE) using equity returns which produces estimates under the physical measure. In order to price options, non-linear least-squares (NLS), based on option prices, are more desirable than using historical returns (see, for example, Christoffersen and Jacobs, 2004; Christoffersen et al., 2013) since option prices contain forward-looking information. However, as pointed out by Duan and Yeh (2010) and Kanniainen et al. (2014), estimating GARCH models using a large amount of cross-sections of option data increases the computational burden.

Several recent papers focus on using VIX index information to estimate GARCH models. The VIX index, introduced by the Chicago Board Options Exchange (CBOE) in 1993, reflects investor fear levels and market sentiment on a day-by-day basis, showing the risk-neutral expected annualised volatility of the S&P 500 over the next 30 days. Therefore, the risk-neutral GARCH parameters are estimated based on the information provided by the VIX index. For example, Hao and Zhang (2013) estimate GARCH models by proposing a joint likelihood function using both returns and the VIX. Their work is carried out under the locally risk-neutral valuation relationship proposed by Duan (1995). Kanniainen et al. (2014) suggest that calculating spot volatilities with VIX data, rather than from returns, improves the performance of GARCH option pricing. Also, they point out that a joint maximum likelihood function using returns and the VIX generates better estimates than a maximum likelihood function based on returns only in terms of option pricing errors. Liu et al. (2015) calibrate three different types of GARCH models on the VIX index of the previous trading day. They show that their estimates produce reasonable one-day out-of-sample VIX forecasts. Wang et al. (2017) propose a closed-form formula for pricing VIX futures based on the Heston and Nandi (2000) GARCH model, where the parameters are estimated using both the VIX and VIX futures prices. Also, several studies use GARCH estimates to forecast VIX as an extended application of GARCH pricing models; see, for instance, Barone-Adesi et al. (2008) and Byun and Min (2013). Other related articles include Kambouroudis and McMillan (2016) who consider VIX as an exogenous variable within a selection of GARCH models, and Huang et al. (2019b) who estimate the extended leverage heterogeneous autoregressive gamma (LHARG) model of Majewski et al. (2015) using both the VIX term structure and the VIX futures.

However, the current literature on GARCH option pricing using CBOE VIX considers only normally distributed returns. In the approach presented in this chapter we not only use filtered historical innovations, but also four volatility indices to estimate GARCH models. Following Barone-Adesi et al. (2008), we allow the volatility parameters to be different under the physical and the risk-neutral measures. Byun and Min (2013) point out that using the same values for the one-day-ahead conditional volatility under both measures, as in Barone-Adesi et al. (2008), will lead to unstable estimated parameters. Therefore, in this chapter, following Byun and Min (2013), we consider the one-day-ahead volatility to be different under the two measures. Instead of using cross-sectional option prices leading to time-consuming estimations, our estimation is based on VIX data that reduces estimation time significantly. This is in line with Kanniainen et al. (2014) who point out that the joint estimation with returns and VIX saves computational time, especially for non-affine GARCH models, which do not have closed-form solutions of option prices. We compare the forecasting performance of our proposed model with the Normal-VIX model of Hao and Zhang (2013). Also, we compare our model with the FHS-options model of Barone-Adesi et al. (2008) from a computational burden perspective.

To our knowledge, this is the first study in which the four well-established VIX indices are used in volatility modelling based on GARCH. As such, from a VIX forecasting perspective, our method improves on the traditional GARCH models in three different ways. First, the empirical distribution of innovations captures excess skewness, kurtosis, and other non-normal features of return data. Second, the flexible change of measure (different parameters for the risk-neutral and physical volatility processes) induces better pricing performance both in-sample and out-of-sample. Third, we consider forward-looking information in our estimation, but instead of option prices we use the CBOE volatility indices (VIX9D, VIX, VIX3M and VIX6M) in order to significantly reduce computational time when compared to the FHS-options method of Barone-Adesi et al. (2008).

The remainder of this chapter is organized as follows. Section 2.2 presents the new estimation method that uses the filtered historical simulation and the CBOE volatility indices. Section 2.3 provides the empirical results and analysis, Section 2.4 details a series of robustness checks, and Section 2.5 concludes the study.

2.2 The models

In this section, we introduce the different GARCH model estimations we investigate in this chapter. We first discuss two competing approaches: the model of Barone-Adesi et al. (2008) (the FHS-options method, hereafter) and the one of Hao and Zhang (2013) (the Normal-VIX method, hereafter). The FHS-options method is used to estimate model parameters assuming non-normal innovations and uses option prices, while the Normal-VIX method combines normal innovations with the CBOE VIX information. Subsequently, motivated by the benchmark models, we propose a new approach to estimate GARCH models using non-normal innovations and volatility indices. To show the relationship between the daily conditional variance and the volatility indices, we explain the CBOE volatility indices in a discrete-time setting.

2.2.1 The FHS-options method

It is a well-established fact that returns have fat left tails, which refers to negative skewness and leptokurtosis. Barone-Adesi et al. (2008) employ the filtered historical simulation to accommodate for these nonstandard features of the return innovations by using the empirical innovation density. Also, they use the GJR GARCH model of Glosten et al. (1993) (GJR, hereafter) to account for the leverage effect, i.e., negative returns having more impact on the volatility than positive returns.

Barone-Adesi et al. (2008) assume that in each period under the physical measure the asset return is assumed to follow the asymmetric GJR model below:

$$\begin{aligned} \ln(S_t/S_{t-1}) &= \mu + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t \\ \sigma_t^2 &= \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma I_{t-1}\varepsilon_{t-1}^2, \end{aligned} \tag{2.2.1}$$

where

$$I_{t-1} = \begin{cases} 1, & \varepsilon_{t-1} < 0 \\ 0, & \varepsilon_{t-1} \geq 0. \end{cases}$$

S_t is the stock price at time t, μ is the expected return, and σ_t^2 is the conditional variance of the log returns $\ln(S_t/S_{t-1})$, where $z_t \mid \mathcal{F}_{t-1} \sim F(0,1)$, and $\mathcal{F}_t$ is the information set up to time t. F is some unknown distribution function with zero mean and unit variance, which we estimate using the empirical distribution function. $\gamma > 0$ captures the asymmetric response of volatility to positive and negative returns.

On the other hand, under the risk-neutral measure the stock process is assumed to follow:

$$\ln(S_i/S_{i-1}) = \mu^* + \varepsilon_i, \quad \varepsilon_i = \sigma_i z_i$$
$$\sigma_i^2 = \omega^* + \alpha^* \varepsilon_{i-1}^2 + \beta^* \sigma_{i-1}^2 + \gamma^* I_{i-1} \varepsilon_{i-1}^2, \tag{2.2.2}$$

The notation used is the same as in Barone-Adesi et al. (2008): μ^* is the risk-neutral drift which ensures that the expected stock return equals the risk-free rate, and z_i is assumed to follow the same distribution function $F(0,1)$ as under the physical measure for $i > t$. Under the risk-neutral measure the volatility dynamics also follow an asymmetric GJR process. Differently from the traditional GARCH estimation procedure which specifies the change of probability measure from $\mathbb{P}$ to $\mathbb{Q}$, this method directly calibrates a new set of risk-neutral parameters using $S\&P$ 500 index options.

2.2.2 The Normal-VIX method

Hao and Zhang (2013) use the information of CBOE VIX to GARCH model estimation. They calculate the squared VIX as a risk-neutral expectation of the arithmetic average variance over the next 21 trading days under Duan (1995)'s locally risk-neutral valuation relationship (LRNVR) framework[2]. The estimation is then carried out within a set of GARCH model specifications using both the

[2]More details about the relationship between CBOE VIX and the daily conditional variance are discussed in Section 2.2.3.

returns and the VIX. The GJR model defined under the LRNVR is[3]

$$
\begin{aligned}
\text{Physical measure:} \quad & \ln(S_t/S_{t-1}) = r_t + \lambda\sigma_t - \frac{1}{2}\sigma_t^2 + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t \\
& \sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma I_{t-1}\varepsilon_{t-1}^2 \\
\text{Risk-neutral measure:} \quad & \ln(S_t/S_{t-1}) = r_t - \frac{1}{2}\sigma_t^2 + \xi_t, \quad \xi_t = \sigma_t z_t \\
& \sigma_t^2 = \omega + \alpha(\xi_{t-1} - \lambda\sigma_{t-1})^2 + \beta\sigma_{t-1}^2 + \gamma I_{t-1}(\xi_{t-1} - \lambda\sigma_{t-1})^2
\end{aligned}
\tag{2.2.3}
$$

where r_t is the risk-free rate at time t, λ is the risk premium, $z_t \mid \mathcal{F}_{t-1} \sim N(0,1)$, and $\{\omega, \alpha, \beta, \gamma\}$ are the GJR parameters.

The implied VIX at time t is a linear function of the conditional variance in the next period under the LRNVR:

$$
Vix_t = A + B\sigma_{t+1}^2, \tag{2.2.4}
$$

where

$$
\begin{aligned}
Vix_t &= (VIX_t/100)^2/252, \\
A &= \frac{\omega}{1-\eta}(1-B), \\
B &= \frac{1-\eta^n}{n(1-\eta)}, \\
\eta &= \alpha(1+\lambda^2) + \beta + \gamma S.
\end{aligned}
\tag{2.2.5}
$$

[3]See Hao and Zhang (2013) for other GARCH specifications under the LRNVR.

If $z_t = \xi_t/\sigma_t$ follows i.i.d.$N(0,1)$, then $S = [\frac{\lambda}{\sqrt{2\pi}}e^{-\frac{\lambda^2}{2}} + (1+\lambda^2)N(\lambda)]$. Hao and Zhang (2013) propose a joint log-likelihood estimation using the CBOE VIX and the returns.

2.2.3 CBOE volatility indices

In this section, we briefly describe the CBOE volatility indices which measure the market expectation of volatility implied by option prices. The CBOE VIX, the first introduced volatility index, is often referred to as the "market fear gauge" (see Whaley, 2009). Since its creation, it has become the standard measure of volatility risk for practitioners. Nowadays, the investors are able to trade volatility via VIX derivatives as the VIX itself is not a tradable asset (see Mencía and Sentana, 2013). This chapter focuses on volatility indices calculated from S&P 500 options data, i.e., VIX, the CBOE short-term volatility index (VIX9D), the CBOE 3-month volatility index (VIX3M) and the CBOE mid-term volatility index (VIX6M).

According to Carr and Madan (1998) and Demeterfi et al. (1999), the VIX index is calculated from out-of-the-money (OTM) S&P 500 index options (put and call) using the formula[4],

$$\sigma^2 = \frac{2}{T}\sum_i \frac{\triangle K_i}{K_i^2} e^{RT} Q(K_i) - \frac{1}{T}[\frac{F}{K_0} - 1]^2, \tag{2.2.6}$$

where T is 30 days, F denotes the implied forward index level derived from index option prices by using the put-call parity. K_i is the strike price of the ith OTM option, $\triangle K_i$ is the interval between strike prices, and K_0 is the first strike that is below the forward index level F. $Q(K_i)$ is the midpoint of the bid-ask spread of each option with strike K_i. Then VIX is defined as $\sigma \times 100$. VIX^2 represents the S&P 500 30-day variance swap rate. This can be interpreted as the expectation of the integrated variance of the following 30 days under the risk-neutral measure. Formally, in a discrete-time setting, at time t we have:

$$VIX_t = 100 * \sqrt{\frac{\tau}{T} * \sum_{k=1}^{30} E^Q[\sigma^2_{t+k} \mid \mathcal{F}_t]}. \tag{2.2.7}$$

where $E^Q[\cdot]$ is the expectation under the risk-neutral measure. When applying the calendar day count convention, $\tau = 365$ is the annualising parameter and $T = 30$ is the number of calendar days in a month.[5] Then, VIX9D, VIX3M and VIX6M are calculated in a similar way to VIX, except that the VIX represents a constant 30 calendar days ahead volatility, whereas VIX9D, VIX3M and VIX6M

[4]The original VIX index, proposed by Whaley (1993), was the implied volatility of the at-the-money (ATM) S&P 100 options. In 2003, CBOE introduces the new VIX index which is based on the S&P 500 options, and the old VIX is then renamed as VXO.

[5]When trading day count convention is used, $\tau = 252$ and $T = 22$.

measure the implied volatility of the $S\&P$ 500 options for the next nine days, three months and six months, respectively.

2.2.4 The FHS-VI method

In this section, we propose a new approach to estimate GARCH models using the filtered historical returns and volatility indices; we investigate three different GARCH models. We employ the classic GARCH(1,1) model of Bollerslev (1986) (GARCH, hereafter), the nonlinear asymmetric GARCH model of Engle and Ng (1993) (NAGARCH, hereafter) and the GJR model by Glosten et al. (1993) in order to capture the leverage effect.

The specification of asset returns is the same in all three models we investigate. Under the physical measure $\mathbb{P}$, the logarithm of returns follows the dynamic:

$$\ln(S_t/S_{t-1}) = \mu_t - \kappa_t + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t \tag{2.2.8}$$

where S_t is the stock price at time t, μ_t is the expected return, σ_t is the conditional volatility of the log return $\ln(S_t/S_{t-1})$, $z_t \mid \mathcal{F}_{t-1} \sim F(0,1)$, $\mathcal{F}_{t-1}$ is the information set up to time $t-1$. F is some unknown distribution function with zero mean and unit variance, which we estimate using the empirical distribution function. κ_t is the mean correction factor defined as:

$$\kappa_t = \ln(E_{t-1}[\exp\{\varepsilon_t\}]) \tag{2.2.9}$$

We have:

$$E_{t-1}[S_t/S_{t-1}] = E_{t-1}[\exp\{\mu_t - \kappa_t + \varepsilon_t\}] = \exp\{\mu_t\}. \tag{2.2.10}$$

Motivated by Christoffersen and Jacobs (2004), the conditional variance dynamics of the three GARCH models are nested in the general form below:

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + g(\varepsilon_{t-1}) \tag{2.2.11}$$

The different GARCH models have different expressions for the innovation function g:

$$\begin{aligned} \text{GARCH:} \quad & g(\varepsilon_{t-1}) = \alpha\varepsilon_{t-1}^2 \\ \text{NAGARCH:} \quad & g(\varepsilon_{t-1}) = \alpha(\varepsilon_{t-1} - \theta\sigma_{t-1})^2 \\ \text{GJR:} \quad & g(\varepsilon_{t-1}) = [\alpha + \gamma I(\varepsilon_{t-1} < 0)]\varepsilon_{t-1}^2 \end{aligned} \tag{2.2.12}$$

For the NAGARCH and GJR models, a positive θ and γ ensure an asymmetric response of the volatility to positive and negative returns, i.e., negative returns increase future volatility by a larger amount than positive returns of the same magnitude.

When assuming that the return innovations are normally distributed, the GARCH models are often estimated by the maximum likelihood estimation (MLE) method. Bollerslev and Wooldridge (1992) demonstrate that this method yields consistent estimates, even when the normality assumption is violated. The estimation procedure is then called quasi-maximum likelihood estimation (QMLE). Under the physical measure, we perform QMLE using the historical log-returns $\{R_t = \ln(S_t/S_{t-1}); t = 1, 2, ..., n\}$. The estimates are obtained by maximising the following log-likelihood function for the GARCH models in equation (2.2.11):

$$\ln L_R = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\left\{\ln(\sigma_t^2) + \frac{(R_t - \mu_t + \kappa_t)^2}{\sigma_t^2}\right\} \tag{2.2.13}$$

Given the estimates, the spot variance σ_t^2 is updated according to the return dynamics.

Under the risk-neutral measure we have that:

$$\ln(S_t/S_{t-1}) = r_t - \kappa_t^* + \varepsilon_t^*, \quad \varepsilon_t^* = \sigma_t^* z_t^* \tag{2.2.14}$$

where r_t is the risk-free rate at time t which is same as in the LRNVR framework of equation (2.2.3), and κ_t^* is the mean correction factor under the risk-neutral measure:

$$\kappa_t^* = \ln(E_{t-1}^Q[\exp\{\varepsilon_t^*\}]) \tag{2.2.15}$$

so that

$$E_{t-1}[S_t/S_{t-1}] = E_{t-1}[\exp\{r_t - \kappa_t^* + \varepsilon_t^*\}] = \exp\{r_t\}. \tag{2.2.16}$$

The conditional variance dynamics are as follows:

$$\sigma_t^{*2} = \omega^* + \beta^* \sigma_{t-1}^{*2} + g^*(\varepsilon_{t-1}^*) \tag{2.2.17}$$

where for the different models we have:

$$\begin{aligned} \text{GARCH:} \quad & g^*(\varepsilon_{t-1}^*) = \alpha^* \varepsilon_{t-1}^{*2} \\ \text{NAGARCH:} \quad & g^*(\varepsilon_{t-1}^*) = \alpha^* (\varepsilon_{t-1}^* - \theta^* \sigma_{t-1}^*)^2 \\ \text{GJR:} \quad & g^*(\varepsilon_{t-1}^*) = [\alpha^* + \gamma^* I(\varepsilon_{t-1}^* < 0)] \varepsilon_{t-1}^{*2}. \end{aligned} \tag{2.2.18}$$

To distinguish from the spot variance under the physical measure σ_t^2, the risk-neutral variance is denoted by σ_t^{*2}. Whilst Barone-Adesi et al. (2008) assume that the spot variance is the same under the physical and risk-neutral measures, Byun and Min (2013) show that a model provides more accurate pricing performance by allowing the risk-neutral spot variance to be different from the physical one. Also, Kanniainen et al. (2014) demonstrate that extracting the spot volatility from the VIX index can improve on the model's performance compared with calculating spot volatility using the series of the underlying asset returns. The difference is driven by the conditional skewness and excess kurtosis as shown in Christoffersen et al. (2009). For a given predetermined sequence $\{\nu_t\}$, they define the Radon-Nikodym derivative as follows:

$$\frac{dQ}{dP} \mid \mathcal{F}_t = \exp\Big(- \sum_{i=1}^{t} (\nu_i \varepsilon_i + \Psi_i(\nu_i)) \Big) \tag{2.2.19}$$

where $\mathcal{F}_t$ is the information set up to time t, $\Psi_t(u)$ is the logarithm of the moment generating function:

$$E_{t-1}[\exp(-u\varepsilon_t)] \equiv \exp(\Psi_t(u)). \tag{2.2.20}$$

The mean correction factor κ_t in equation (2.2.8) thus can be viewed as $\Psi_t(-1)$. The authors then demonstrate the existence of an equivalent martingale measure and show that:

$$\sigma_t^{*2} \approx \sigma_t^2 - skew_t \sigma_t^3 \nu_t + \frac{kurt_t}{2} \sigma_t^4 \nu_t^2 \tag{2.2.21}$$

where ν_t is an approximation of the modified Sharpe ratio:

$$\nu_t \approx \frac{\mu_t - r_t}{\sigma_t^2} + \frac{1}{2} - \frac{\kappa_t}{\sigma_t^2} \tag{2.2.22}$$

Therefore, with a negative skewness and positive excess kurtosis, the risk-neutral conditional variance is greater than the conditional variance under the physical measure[6]. In this chapter, we allow σ_t^{*2} to be different from σ_t^2 by estimating the risk-neutral spot variance σ_t^{*2} from the information on the volatility index.

Specifically, following Byun and Min (2013), we allow the risk-neutral spot

[6]See Christoffersen et al. (2009) for more details.

variance to be different from the physical one. In addition, a new set of risk-neutral parameters are calibrated by using information on the CBOE volatility indices directly.[7] Since the distribution of the future return innovations cannot be derived analytically, Monte Carlo simulations are used in the computation of the GARCH conditional variance. Estimates are then found by minimising the mean squared error between the prices given by the model and the market prices. The estimation process is discussed in the next section.

2.2.5 Estimation using the FHS-VI method

This section introduces a new approach to calibrate the GARCH models to the information provided by the volatility indices. The calibration is based on the filtered historical simulation method introduced by Barone-Adesi et al. (2008). They estimate the GJR model by minimising the errors between the simulated option prices and the $S\&P$ 500 option prices. To ensure better pricing performance, they calibrate the GJR model to option prices of a large sample size of three years, i.e., 29,211 OTM call and put options in total. This requires intensive computation and is time-consuming. Hao and Zhang (2013) and Kanniainen et al. (2014) show that using information on CBOE VIX can improve the pricing performance of GARCH models whilst avoiding costly computations. Here we

[7]Barone-Adesi et al. (2008) show that the flexible change of measure achieves a better pricing performance than other competing GARCH option pricing models, such as the *ad hoc* Black-Scholes model introduced by Dumas et al. (1998), the Heston and Nandi (2000) GARCH model, and the GARCH model with inverse Gaussian innovations of Christoffersen et al. (2006).

propose a new extension, calibrating model parameters assuming filtered historical returns and using CBOE volatility indices, which reduces the computational burden significantly.

The estimation procedure is:

1. Under the physical measure, the GARCH models are estimated on each Wednesday which is least likely to be a holiday or affected by the weekend effect. The GARCH parameters $\{\omega, \alpha, \beta, (\gamma), (\theta)\}$ are estimated by maximising the log-likelihood function in equation (2.2.13) with 3,500 historical returns (daily)[8][9]. Thus, the return innovations $\{\hat{z}_t\}$ are acquired. We repeat this estimation every week.

2. Under the risk-neutral measure, a daily variance series is simulated for the next 6 months using the variance dynamics of equation (2.2.17)[10]. The GARCH parameters are initialized with $\{\hat{\omega}, \hat{\alpha}, \hat{\beta}, (\hat{\gamma}), (\hat{\theta})\}$ which are the model estimates obtained under the physical measure in the step 1. The spot variance here is an unknown parameter in the calibration procedure[11]. The conditional variance of the following 6 months $\{\sigma^{*2}_{t+1}, \sigma^{*2}_{t+2} ..., \sigma^{*2}_{t+126}\}$[12] are

[8] To be aligned with the model of Barone-Adesi et al. (2008), we also use 3,500 historical returns to estimate the GJR GARCH model under the physical measure. Moreover, Bollerslev and Wooldridge (1992) point out that a large sample size will ensure the consistency of the quasi-maximum likelihood estimation.

[9] We use the unconditional variance as the initial variance in the estimation. $\mu_t - \kappa_t$ is also estimated in this step.

[10] Liu et al. (2015) propose a closed-form solution under this framework.

[11] Byun and Min (2013) show that, instead of just simply improving the goodness of fit, the estimated spot variance can be treated as the true spot variance under the risk-neutral measure.

[12] In this study, we use the trading day count convention, as the innovation distribution is estimated with trading days returns.

then updated by each day drawing an observation from the past innovations of $\{\hat{z}_t\}$.

3. N simulated sample paths are generated by repeating the procedure in step 2. The expectation of the risk-neutral conditional variance for the following ith day can be computed as: $E_t^Q[\sigma_{t+i}^{*2}] = \frac{1}{N}\sum_{n=1}^{N}\sigma_{t+i}^{*(n)}$, where $\sigma_{t+i}^{*(n)}$ is the simulated conditional variance at time $t+i$ in the nth sample path and N is the total number of simulated paths. In this chapter, we use $N = 50,000$ paths[13].

4. According to the definition of VIX and equation (2.2.7), the GARCH model implied VIX (model VIX, hereafter) under the trading day count convention can be calculated as:

$$VIX_t^{model} = 100 * \sqrt{\frac{252}{22} * \sum_{i=1}^{22} E_t^Q[\sigma_{t+i}^{*2}]} \tag{2.2.23}$$

Similarly:

$$VIX9D_t^{model} = 100 * \sqrt{\frac{252}{7} * \sum_{i=1}^{7} E_t^Q[\sigma_{t+i}^{*2}]} \tag{2.2.24}$$

$$VIX3M_t^{model} = 100 * \sqrt{\frac{252}{63} * \sum_{i=1}^{63} E_t^Q[\sigma_{t+i}^{*2}]} \tag{2.2.25}$$

[13] Both Barone-Adesi et al. (2008) and Byun and Min (2013) calibrate risk-neutral GARCH parameters using a cross-section of option prices, producing 20,000 and 50,000 simulation paths, respectively. More simulation paths could also be considered.

$$VIX6M_t^{model} = 100 * \sqrt{\frac{252}{126} * \sum_{i=1}^{126} E_t^Q[\sigma_{t+i}^{*2}]} \tag{2.2.26}$$

5. The optimisation is then achieved by minimising the root mean square error (RMSE) between the model volatility index and the market volatility index[14]:

$$\sqrt{\sum_{k=1}^{4} \left[w_k * \left(VI^{(k)market} - VI^{(k)model} \right)^2 \right]} \tag{2.2.27}$$

   with $VI^{(k)market}$ denoting the market prices of VIX, VIX9D, VIX3M and VIX6M, respectively, $VI^{(k)model}$ standing for the GARCH model implied volatility index produced in step 4, and here we use $w_k = 0.25$ representing equal weights for each index.

2.2.6 Model evaluation

To measure the quality of fit for the pricing models in-sample, we calculate several measures: the mean of absolute errors (MAE) and the root mean squared error (RMSE). These are defined as:

$$MAE = \frac{1}{M} \sum_{i=1}^{M} \sum_{k=1}^{4} \left[w_k * |VI_i^{(k)market} - VI_i^{(k)model}| \right] \tag{2.2.28}$$

[14]The alternative function, e.g., the relative measure, also could be considered in the optimization.

$$RMSE = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{4}\left[w_k * \left(VI_i^{(k)market} - VI_i^{(k)model}\right)^2\right]} \qquad (2.2.29)$$

where $w_k = 0.25$ is the weight of each index assuming equal weighting, M is the number of total observations in a year, $VI_i^{(k)market}$ and $VI_i^{(k)model}$ refer to the market price and the model price of different volatility indices, respectively.

We use four different volatility indices to estimate the models, while the benchmark model only uses the VIX index. Minimising the errors between the market prices and model prices will place a greater weight on the volatility index with a higher value. Therefore, we also report the MAE in relative terms (MAE%), i.e., the percentage of MAE compared to the average market price; and the RMSE in relative terms (RMSE%), i.e., the percentage of RMSE compared to the average market price.

Patton (2011b) recommends the use of two loss functions, i.e., MSE and QLIKE, as these are the only ones that are robust to noise in the volatility proxy. Hence, we also report QLIKE values, which are defined as (we use $w_k = 0.25$):

$$QLIKE = \frac{1}{M}\sum_{i=1}^{M}\sum_{k=1}^{4}\left[w_k * \left(\frac{{VI_i^{(k)market}}^2}{{VI_i^{(k)model}}^2} - \log(\frac{{VI_i^{(k)market}}^2}{{VI_i^{(k)model}}^2}) - 1\right)\right]. \qquad (2.2.30)$$

To compare our approach with the Normal-VIX model, we also assess the out-of-sample pricing performance in the following way: for each Wednesday in our sample period, the in-sample parameter estimates from Section 2.3.2 are used to forecast the VIX index for the following Wednesday. For out-of-sample comparison, we use the mean squared error (MSE) to evaluate the forecasting accuracy of six GARCH models, as follows:

$$MSE = \frac{1}{M}\sum_{i=1}^{M}\left(VIX_i^{model} - VIX_i^{market}\right)^2 \tag{2.2.31}$$

where VIX_i^{model} is the one week ahead VIX produced by the models, and VIX_i^{market} is the corresponding market price of the CBOE VIX.

Smaller forecasting errors indicate the predictive superiority of a given model. However, one may want to know whether a model has statistically significant superior forecasting ability. To address this, we use the approach proposed by Diebold and Mariano (1995) to test the equal accuracy of two different forecasting models. Since we estimate our models on a finite window of data, in our case, the DM test coincides with the test of Giacomini and White (2006), which applies to nested models. The two sets of forecast errors are defined as $e_{1,t}$ and $e_{2,t}$, respectively. The function $g(\cdot)$ is a loss function which typically is the squared error loss, i.e., $e_{1,t}^2$ and $e_{2,t}^2$ or absolute error loss $|e_{1,t}|$ and $|e_{2,t}|$. Then the loss differential between the two forecasts is $d_t = g(e_{1,t}) - g(e_{2,t})$. Therefore, the null hypothesis of equal forecast accuracy can be expressed as on expectation of zero for the loss differential $E[d_t] = 0$. Under fairly weak conditions, the DM test statistic:

$$DM = \frac{\bar{d}}{\sqrt{2\pi\hat{f}_d(0)/T}} \tag{2.2.32}$$

has an asymptotic standard normal distribution under the null hypothesis, where T is the number of total observations; $\bar{d}$ is the sample mean of the loss differential $\bar{d} = \frac{1}{T}\sum_{t=1}^{T} d_t$ and $2\pi\hat{f}_d(0)$ is a consistent estimator of the asymptotic variance. In this chapter, the DM test is calculated based on the MSE of the different GARCH models.

The DM test is only used for pairwise testing of two models. In order to test whether a particular forecasting model significantly outperforms a set of competing models, we employ the superior predictive ability (SPA) test proposed by Hansen (2005). This test uses the loss differential defined as $d_{k,t} = g(e_{0,t}) - g(e_{k,t})$, where $g(e_{0,t})$ and $g(e_{k,t})$ are the values of the loss function $g(\cdot)$ at time t for the base model and m competing models, for $k = 1, 2, ..., m$. The null hypothesis that the base model is not outperformed by its competitors can be written as $\max\limits_{k=1,...,m} E[d_{k,t}] \leq 0$. Then the statistic for the SPA test is calculated as:

$$T^{SPA} = \max_{k=1,...,m} \frac{T^{1/2}\bar{d}_k}{\hat{\omega}_k} \tag{2.2.33}$$

where $\bar{d}_k$ is the sample mean of the loss function for model k, $\bar{d}_k = \frac{1}{T}\sum_{t=1}^{T} d_{k,t}$ and $\hat{\omega}_k^2$ is a consistent estimator of $\omega_k^2 = var(T^{1/2}\bar{d}_k)$. The distribution and the p-value of T^{SPA} can be obtained by using a stationary bootstrap procedure as in Hansen

(2005). The higher the p-value, the less likely that the null hypothesis is rejected, which means that the base model has superior forecasting ability compared to the set of competing models.

2.3 Empirical analysis

2.3.1 Data

The CBOE volatility indices used in this chapter are the VIX, VIX9D, VIX3M and VIX6M, downloaded from the CBOE website. Since the VIX9D data is available from 2 January 2011, our sample data is from 2 January 2011 to 29 December 2017.[15] The VIX information for the same period is also used to estimate the Normal-VIX model. The three months Treasury bill rate is used as the risk-free rate which is downloaded from the U.S. Department of Treasury website. In addition, to compare our approach with the FHS-options method, we use European options on the S&P 500 index from 2 January 2002 to 30 December 2017, downloaded from OptionMetrics.[16]

[15] The starting dates of VIX, VIX3M and VIX6M are 2 January 2004, 4 December 2007 and 7 January 2008, respectively.

[16] We follow the same criteria of Barone-Adesi et al. (2008) to sort data: (1) only use the out-of-the-money European options since they are more actively traded than in-the-money options. (2) choose options which mature in more than 10 days and less than 360 days. (3) only include options which cost more than \$0.05. (4) options with implied volatility value larger than 70% are excluded. This yields a sample of 882,009 observations in total. To compare with the FHS-options model, we choose the same start date as in Barone-Adesi et al. (2008).

Figure 2.3.1: The dynamics of the CBOE volatility indices between 03 January, 2011 and 29 December, 2017

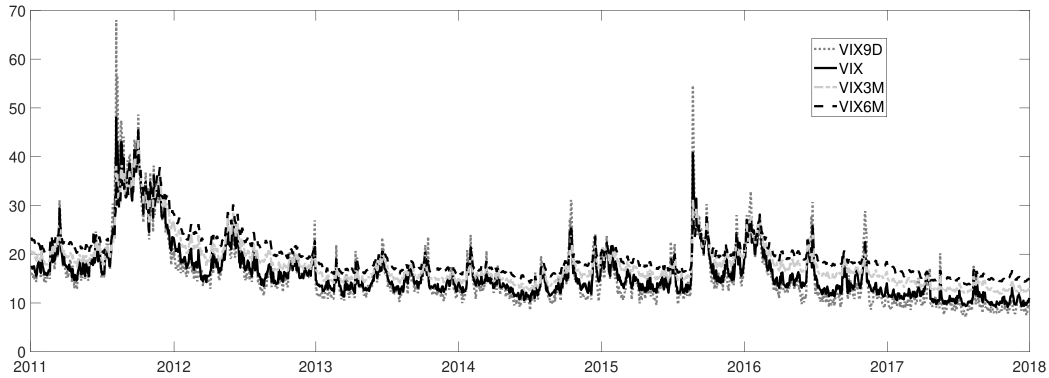


Figure 2.3.1 shows the dynamics of the four CBOE volatility indices during the sample period. We observe that the four indices experience the same pattern of fluctuations, i.e., a sharp increase and then drop in 2011-2012 and 2015-2016. Furthermore, for most of the days in the sample, VIX6M has the highest values while VIX9D has the lowest values among the indices. The difference in the price pattern can be explained as longer maturity means more volatility due to the uncertainty in the future.

2.3.2 In-sample model comparison

In this section, we carry out the estimation of the different GARCH models using different methods described in Section 2.2. Then we compare the in-sample performance of the GARCH, GJR and NAGARCH models under the FHS-VI and the Normal-VIX frameworks.

We first discuss the estimation results of the GARCH models using different

Table 2.3.1: Parameter estimate statistics obtained using different methods for 2017

P measure	$\omega \times 10^6$		β		α		γ		θ		Ann. vol.	
Model	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.	Mean	Std.dev.
Panel A. FHS-VI												
GARCH	2.119	0.134	0.877	0.001	0.101	0.001	-	-	-	-	0.088	0.014
GJR	2.267	0.110	0.886	0.002	0.000	0.000	0.176	0.002	-	-	0.089	0.015
NAGARCH	2.300	0.122	0.767	0.005	0.074	0.001	-	-	1.407	0.035	0.088	0.016
Q measure	$\omega^* \times 10^6$		β^*		α^*		γ^*		θ^*		Ann. vol.	
Model	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev
Panel B. FHS-VI												
GARCH	0.657	0.732	0.919	0.024	0.075	0.023	-	-	-	-	0.134	0.012
GJR	0.458	0.619	0.831	0.065	0.078	0.023	0.174	0.105	-	-	0.126	0.011
NAGARCH	0.721	0.744	0.877	0.062	0.058	0.025	-	-	1.096	0.427	0.156	0.016
Panel C. FHS-options												
GJR	2.265	1.272	0.871	0.048	0.009	0.020	0.173	0.055	-	-	0.147	0.013
Panel D. Normal-VIX												
GARCH	1.588	0.061	0.947	0.001	0.047	0.001	-	-	-	-	0.108	0.006
GJR	1.341	0.078	0.961	0.002	0.000	0.002	0.062	0.004	-	-	0.111	0.007
NAGARCH	1.624	0.073	0.932	0.001	0.038	0.001	-	-	0.779	0.040	0.110	0.007

This table presents the parameter estimate statistics obtained using FHS-VI, FHS-options and Normal-VIX for the year 2017. The models are estimated on each Wednesday of the year. Panel A reports the FHS-VI parameter estimate statistics under physical measure. Panel B, C and D present the parameter estimate statistics obtained using the FHS-VI, FHS-options and Normal-VIX under the risk-neutral measure, respectively. Ann. vol. is the annualized volatility.

methods. Table 2.3.1 reports the statistics (mean and standard deviation) of the parameter estimates obtained using volatility indices-, options- and VIX-based estimation procedures, i.e., FHS-VI, FHS-options and Normal-VIX, for the year 2017. For the GJR and NAGARCH models, estimates of γ and θ larger than zero show that negative returns affect the conditional variance more than positive returns, i.e. evidence of leverage effect. The table also presents the annualised volatilities implied by the models. The difference between the annualised conditional volatilities under physical and risk-neutral measures captures the volatility risk premium (VRP). When VRP is negative, i.e., the risk-neutral volatility is higher than the physical volatility as shown in Table 2.3.1[17], then investors demand a premium to bear the risks in future realised volatilities. This finding is in line with a number of empirical studies documenting a negative VRP, including Carr and Wu (2009), Bollerslev et al. (2011) and Bekaert and Hoerova (2014).

To evaluate how well the different models estimate the volatility process, Table 2.3.2 reports the in-sample pricing errors. By looking at the pricing errors by years, the FHS-VI method outperforms the Normal-VIX method in fitting the volatility indices, regardless of the model or the measurement of fit. This is not surprising as the FHS-VI method employs the empirical innovation distribution and the flexible change of measure, which enhance the model's flexibility to fit the volatility indices. Notably, the GJR model under the FHS-VI framework yields

[17]In this thesis, the VRP is calculated as the annualised volatility under the physical measure minus the annualised volatility under the risk-neutral measure, thus a constant.

Table 2.3.2: In-sample pricing errors

		FHS-VI						Normal-VIX					
Year	Model	MAE	RMSE	MAE%	RMSE%	QLIKE	p-value	MAE	RMSE	MAE%	RMSE%	QLIKE	p-value
2011	GARCH	0.328	0.585	1.257	1.840	0.001	0.182	2.562	3.489	9.878	12.578	0.040	0.094
	GJR	**0.288**	**0.566**	**1.117**	**1.771**	**0.001**	0.930	2.452	3.259	9.744	12.128	0.036	0.343
	NAGARCH	0.912	1.522	3.562	5.758	0.008	0.000	2.559	3.456	9.950	12.439	0.041	0.018
2012	GARCH	0.267	0.385	1.393	2.046	0.001	0.013	1.824	2.323	9.790	12.034	0.033	0.272
	GJR	**0.208**	**0.339**	**1.116**	**1.820**	**0.001**	0.001	1.736	2.167	9.222	10.956	0.028	0.091
	NAGARCH	0.539	0.798	2.698	3.927	0.003	0.000	1.598	2.036	8.609	10.651	0.025	0.187
2013	GARCH	0.202	**0.293**	1.325	**1.893**	**0.001**	0.230	1.891	2.158	13.613	15.749	0.038	0.000
	GJR	**0.201**	0.295	**1.323**	1.920	0.001	0.060	1.696	1.976	12.164	14.297	0.032	0.000
	NAGARCH	0.430	0.707	2.704	4.245	0.004	0.000	1.463	1.680	10.419	12.046	0.024	0.000
2014	GARCH	0.187	0.341	1.174	1.853	0.001	0.091	2.076	2.356	14.968	16.593	0.046	0.000
	GJR	**0.176**	**0.331**	**1.090**	**1.733**	**0.001**	0.047	2.111	2.391	15.457	17.364	0.049	0.000
	NAGARCH	0.314	0.640	1.856	3.283	0.003	0.002	1.823	2.090	13.281	14.955	0.038	0.000
2015	GARCH	0.308	0.446	1.684	2.299	0.001	0.120	2.097	2.500	13.012	15.397	0.038	0.000
	GJR	**0.273**	**0.419**	**1.500**	**2.152**	**0.001**	0.008	2.077	2.459	12.989	15.492	0.038	0.000
	NAGARCH	0.665	1.056	3.484	5.138	0.006	0.000	2.084	2.398	13.236	15.489	0.037	0.000
2016	GARCH	0.287	0.500	1.600	2.485	0.001	0.690	2.269	2.945	14.556	18.738	0.062	0.031
	GJR	**0.268**	**0.487**	**1.500**	**2.409**	**0.001**	0.015	2.175	2.840	14.190	18.685	0.058	0.010
	NAGARCH	0.529	0.934	2.906	4.625	0.005	0.000	2.008	2.524	13.046	16.288	0.047	0.004
2017	GARCH	0.187	0.343	1.456	2.468	0.001	0.001	1.671	1.868	15.308	17.189	0.047	0.000
	GJR	**0.166**	**0.318**	**1.312**	**2.270**	**0.001**	0.003	1.621	1.874	14.846	17.222	0.047	0.000
	NAGARCH	0.216	0.472	1.640	2.969	0.002	0.014	1.742	1.930	16.049	17.999	0.049	0.000
Overall	GARCH	0.252	0.425	1.413	2.144	0.001	0.000	2.057	2.570	13.025	15.638	0.043	0.000
	GJR	**0.226**	**0.405**	**1.280**	**2.026**	**0.001**	0.000	1.983	2.468	12.670	15.407	0.041	0.000
	NAGARCH	0.515	0.932	2.693	4.376	0.005	0.000	1.899	2.367	12.098	14.487	0.037	0.000

This table presents the in-sample pricing errors. MAE is the average absolute error between the market price and the model price; RMSE is the square root of the averaged squared pricing error; MAE% and RMSE% are in relative terms expressed in percentages. QLIKE is given in equation (2.2.30). The p-value is for the null hypothesis that the pricing errors have zero mean. Numbers in bold are the lowest values among different models for a given year.

the best results across the models considering the pricing errors over the years. Following Hao and Zhang (2013), we test whether the pricing errors have zero mean and in the last column for each model of Table 2 we present the p-values of this t-test. Consistent with Hao and Zhang (2013), the model prices implied by the Normal-VIX method are significantly different from the market prices for all three GARCH models we investigate. A visual presentation of the fit of the different GARCH models to the CBOE VIX, using different estimation methods, can be found in the Supplementary Appendix. This is largely similar to Figure 2.3.2, which shows the out-of-sample VIX forecasts for different models.

2.3.3 Out-of-sample model comparison

To test how the FHS-VI method fits the volatility indices out-of-sample, we generate one-week-ahead volatility forecasts of the GARCH models using different estimation methods. Table 2.3.3 shows the out-of-sample pricing errors using the various measures. Importantly, the out-of-sample results confirm that across the years the FHS-VI method has smaller pricing errors than the Normal-VIX method.

To offer a fair comparison of the two methods (FHS-VI and Normal-VIX), Table 2.3.4 summarises the forecast mean squared errors based only on the CBOE VIX. In all the years considered, the NAGARCH model estimated using the FHS-VI method dominates. To determine whether the forecasts produced by the two

Table 2.3.3: Out-of-sample pricing errors

		FHS-VI					Normal-VIX				
Year	Model	MAE	RMSE	MAE%	RMSE%	QLIKE	MAE	RMSE	MAE%	RMSE%	QLIKE
2011	GARCH	2.620	4.387	9.784	14.177	0.057	3.551	5.383	13.204	17.143	0.123
	GJR	2.629	4.396	9.836	14.238	0.057	3.275	4.979	12.433	16.178	0.104
	NAGARCH	**2.554**	**4.314**	**9.398**	**13.524**	**0.053**	3.555	5.309	13.227	16.994	0.114
2012	GARCH	1.801	2.329	9.767	13.340	0.033	2.184	2.844	11.439	14.070	0.056
	GJR	1.809	2.336	9.804	13.357	0.033	2.180	2.730	11.489	13.613	0.051
	NAGARCH	**1.686**	**2.203**	**8.983**	**12.168**	**0.026**	2.057	2.569	10.945	13.097	0.044
2013	GARCH	1.214	1.624	8.219	11.117	0.025	1.749	2.050	12.269	14.216	0.037
	GJR	1.228	1.646	8.335	11.322	0.026	1.710	2.010	11.979	13.912	0.035
	NAGARCH	**1.228**	**1.622**	**8.124**	**10.661**	**0.023**	1.635	1.916	11.394	13.120	0.032
2014	GARCH	1.547	2.621	9.773	15.276	0.052	2.040	2.648	13.929	16.251	0.063
	GJR	1.537	2.617	9.717	15.324	0.051	2.215	2.746	15.410	17.707	0.066
	NAGARCH	**1.459**	**2.529**	**9.151**	**14.632**	**0.048**	2.098	2.621	14.558	16.816	0.060
2015	GARCH	1.785	2.977	9.852	14.521	0.058	2.163	3.179	12.320	15.255	0.075
	GJR	1.810	2.992	9.988	14.599	0.058	2.220	3.235	12.911	16.159	0.076
	NAGARCH	**1.652**	**2.909**	**8.959**	**13.376**	**0.056**	2.246	3.163	13.371	16.262	0.071
2016	GARCH	1.917	2.689	11.565	16.122	0.051	2.397	3.242	14.637	18.648	0.081
	GJR	1.923	2.696	11.575	16.111	0.051	2.346	3.158	14.717	19.052	0.075
	NAGARCH	**1.794**	**2.619**	**10.570**	**15.042**	**0.047**	2.249	2.952	14.070	17.568	0.066
2017	GARCH	1.108	1.740	9.836	16.092	0.039	1.438	1.693	13.121	15.305	0.040
	GJR	1.103	1.736	9.732	15.934	0.039	1.487	1.766	13.588	16.055	0.043
	NAGARCH	**1.043**	**1.593**	**9.206**	**14.226**	**0.035**	1.661	1.886	15.333	17.570	0.047
Overall	GARCH	1.714	2.763	9.833	14.481	0.042	2.218	3.205	12.994	15.925	0.068
	GJR	1.721	2.770	9.860	14.506	0.042	2.206	3.106	13.227	16.208	0.064
	NAGARCH	**1.632**	**2.685**	**9.202**	**13.460**	**0.041**	2.217	3.113	13.286	16.038	0.062

This table presents the out-of-sample pricing errors. MAE is the average absolute error between the market price and the model price; RMSE is the square root of the average squared pricing error; MAE% and RMSE% are in relative terms expressed in percentages. QLIKE is given in equation (2.2.30). Numbers in bold are the lowest values among different models for a given year.

Table 2.3.4: Out-of-sample comparison of the VIX forecasts: DM test

Year	FHS-VI									Normal-VIX								
	GARCH			GJR			NAGARCH			GARCH			GJR			NAGARCH		
	MSE	DM1	DM2	MSE	DM1	DM2	MSE	DM1	DM2	MSE	DM1	DM2	MSE	DM1	DM2	MSE	DM1	DM2
Panel A. Out-of-sample VIX forecast year by year																		
2011	21.13	-4.81	1.08	21.17	-4.38	1.07	**19.48**	-3.75	-	28.94	-	4.99	24.79	-	4.74	28.26	-	2.75
2012	5.90	-2.47	1.47	5.82	-2.32	1.22	**5.12**	-2.15	-	8.10	-	2.08	7.45	-	2.01	6.58	-	1.95
2013	3.00	-2.15	0.56	3.05	-2.26	0.73	**2.77**	-2.49	-	4.20	-	3.85	4.04	-	1.80	3.67	-	1.49
2014	7.46	1.32	1.07	7.43	-0.15	1.16	**6.78**	-0.06	-	7.03	-	0.29	7.54	-	1.48	6.88	-	0.06
2015	9.62	-1.12	2.89	9.60	-1.79	2.68	**8.46**	-2.10	-	10.11	-	2.45	10.47	-	2.75	9.97	-	2.10
2016	8.13	-2.83	3.19	8.11	-2.62	3.07	**7.21**	-2.80	-	10.49	-	2.21	9.97	-	2.01	8.71	-	1.80
2017	3.30	0.31	1.21	3.29	-0.11	1.07	**2.69**	-1.96	-	2.87	-	0.18	3.12	-	1.38	3.56	-	1.76
Panel B. Out-of-sample VIX forecast over 2011-2017																		
h=1	2.63	-6.09	-0.93	**2.61**	-5.83	-1.04	2.81	-5.85	-	6.92	-	6.35	6.40	-	5.88	6.09	-	5.85
h=5	8.39	-2.36	3.35	8.37	-2.19	3.26	**7.52**	-2.90	-	10.27	-	3.60	9.65	-	3.47	9.69	-	2.90
h=20	19.29	-1.31	4.65	18.99	0.12	3.59	**17.53**	-2.72	-	22.46	-	2.17	18.75	-	2.65	18.94	-	2.72

This table presents the out-of-sample mean squared errors (MSE) of the VIX forecasts and the Diebold-Mariano (DM) test statistics based on one-week-ahead volatility forecasts. Panel A presents the MSE and DM statistics for a given year; while Panel B shows the MSE and DM statistics within the overall period for different forecast horizons. The DM test statistic is for the null hypothesis of equal accuracy and follows a N(0,1) distribution. DM1 is the DM statistic when comparing the same GARCH model estimated using different methods. DM2 is the DM statistic when considering the NAGARCH model using the FHS-VI method as the benchmark model. The values in bold are the lowest MSE values among different models.

different methods have a statistically significant difference, we also present the values of the DM test statistics in Panel A of Table 2.3.4 (denoted by DM1 in the table). In 5 out of 7 years, the GARCH model based on the FHS-VI method has negative DM statistics, which indicates that it generate smaller average MSE than the GARCH model based on the Normal-VIX method. Both the GJR and NAGARCH models that use FHS-VI produce lower average MSE than the corresponding models based on the Normal-VIX method for all the years. Surprisingly, for the year 2014, none of the models that use the FHS-VI method produces more accurate forecasts than those based on the Normal-VIX method. For the year 2017, only the NAGARCH model based on the FHS-VI outperforms its counterpart.

Interestingly, instead of the GJR model that proved superior in the in-sample period, the NAGARCH model has in general the smallest out-of-sample pricing errors. One possible reason is that the GJR model overfits the data in-sample. Panel A of Table 2.3.4 also considers the NAGARCH model based on the FHS-VI method as the benchmark model (denoted by DM2). All the DM2 statistics reported in Panel A are positive, indicating that the benchmark model has smaller average MSE values than the other models for all the years. Under the FHS-VI framework, the other two models, i.e., the GARCH and the GJR models, are not significantly different from the NAGARCH in their ability to produce VIX forecasts considering the yearly results. However, when comparing different estimation methods, the NAGARCH model that uses the FHS-VI method outperforms the models that use the Normal-VIX method.

In Table 2.3.5, we report the p-values of the SPA test with the null hypothesis that the benchmark model is not inferior to the other models. We consider each model as a benchmark model whilst the other five models are the competing models. The results in Panel A and Panel B of Table 2.3.5 show that for both MSE and QLIKE loss functions, the NAGARCH model has p-values equal to 1 for all the years. Therefore, we can not reject the null hypothesis that the NAGARCH model based on FHS-VI is superior to any of the alternatives. This is in line with our conclusions drawing from the DM test.

As shown in Figure 2.3.2, the models that use the FHS-VI method outperform the models based on the Normal-VIX method, especially when there is a big change in prices. Importantly, in terms of the VIX forecast performance, the NAGARCH model that uses the FHS-VI method is superior to all the other models.[18]

2.3.4 Computational time

The estimation is performed on a desktop with Intel i7 processor with a frequency of 3.2GHz and 16 GB of RAM. For the year 2017, which means estimation over 52

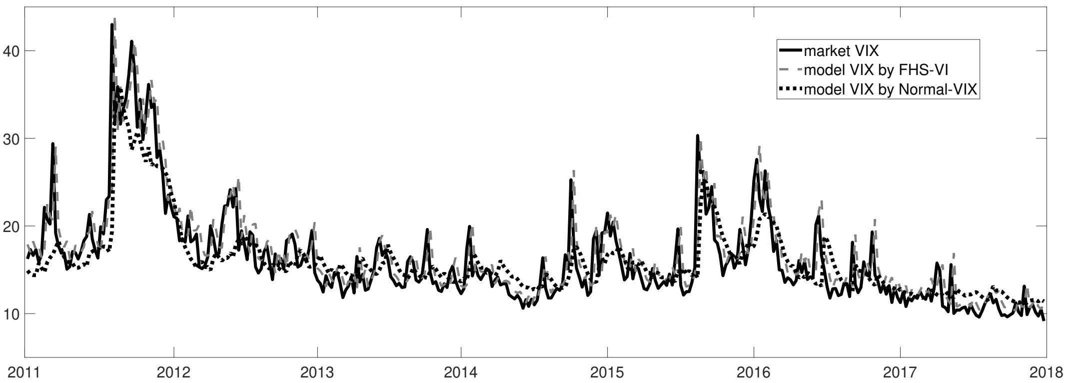
[18]Similarly, Kanniainen et al. (2014) also obtain that the NAGARCH model is better than the GJR model for option pricing when using joint information on the VIX index and the S&P 500 returns.

Table 2.3.5: Out-of-sample comparison of the VIX forecasts: SPA test

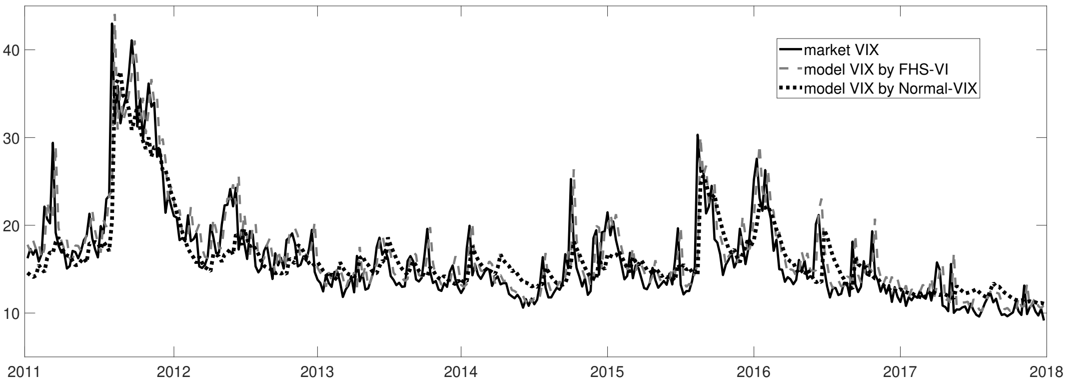
Year	FHS-VI			Normal-VIX		
	GARCH	GJR	NAGARCH	GARCH	GJR	NAGARCH
Panel A: Evaluation by MSE						
2011	0.081	0.066	**1.000**	0.002	0.003	0.008
2012	0.018	0.046	**1.000**	0.052	0.066	0.065
2013	0.235	0.249	**1.000**	0.000	0.019	0.032
2014	0.098	0.077	**1.000**	0.596	0.000	0.421
2015	0.002	0.005	**1.000**	0.012	0.000	0.000
2016	0.001	0.004	**1.000**	0.003	0.010	0.096
2017	0.026	0.064	**1.000**	0.329	0.076	0.000
Panel B: Evaluation by QLIKE						
2011	0.425	0.306	**1.000**	0.000	0.000	0.000
2012	0.230	0.238	**1.000**	0.024	0.016	0.011
2013	0.275	0.377	**1.000**	0.000	0.001	0.001
2014	0.055	0.098	**1.000**	0.012	0.000	0.161
2015	0.243	0.302	**1.000**	0.020	0.003	0.000
2016	0.002	0.009	**1.000**	0.001	0.010	0.057
2017	0.018	0.072	**1.000**	0.331	0.094	0.001
Panel C. Overall 2011-2017, evaluation by MSE for different horizons						
h=1	0.028	**1.000**	0.073	0.000	0.000	0.000
h=5	0.000	0.000	**1.000**	0.008	0.000	0.020
h=20	0.000	0.001	**1.000**	0.000	0.591	0.358
Panel D. Overall 2011-2017, evaluation by QLIKE for different horizons						
h=1	0.263	**1.000**	0.011	0.000	0.000	0.020
h=5	0.003	0.005	**1.000**	0.000	0.000	0.001
h=20	0.000	0.000	**1.000**	0.000	0.080	0.205

This table presents the SPA test results for out-of-sample VIX forecasts under two different loss functions. The SPA test statistic is used to test the null hypothesis that the benchmark model is not outperformed by the competing models. Each column is considered as a benchmark model whilst the other five models are the competitors. The values in bold are the highest SPA *p*-value for the given year. The number of bootstrap replications to calculate the *p*-values is 10,000.

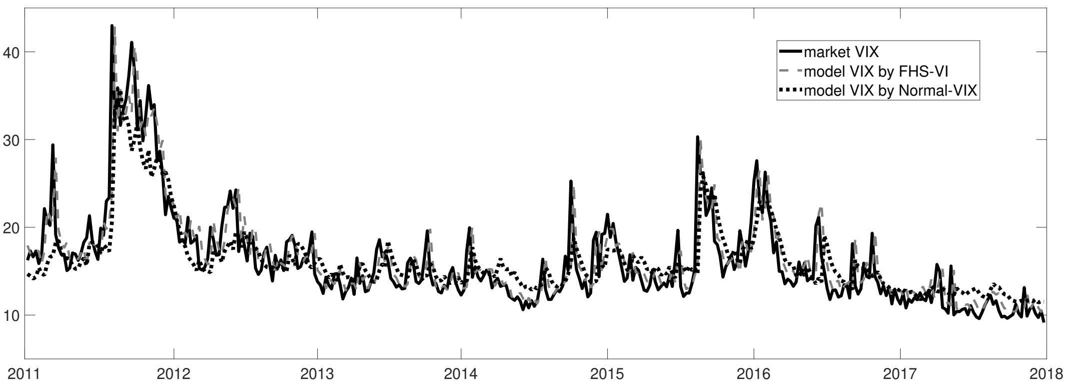
Figure 2.3.2: Out-of-sample comparison of the model VIX and the CBOE VIX



(a) GARCH



(b) GJR



(c) NAGARCH

weeks' estimation (with weekly re-estimations), the running time to calibrate once based on 373,377 option prices; this number of observation is roughly 50 times higher than the number of the volatility index prices, and this contributes to the large processing time needed for the options data. The estimation time is 149 min when using the FHS-options method. On the other hand, the running time for estimation over 52 weeks (still with weekly re-estimations) to calibrate once based on the GJR model is 20.8 min by using the information on VIX indices, i.e., the FHS-VI method. The total running time has little difference among GARCH, GJR and NAGARCH models when using the FHS-VI method, which is consistent with Kanniainen et al. (2014).

During the optimisation procedure, a grid search is performed for the initial values, which results in as many as 1000 iterations, and the estimation time depends on the grid size. Therefore, the estimation with the option-price-based FHS-options method, assuming 100 iterations, takes up to 4.8 h for a single week. The parameter calibration for the FHS-VI GJR model, based on the volatility indices, for one week and 100 iterations, is significantly faster at 40 min, which is a reduction of more than 86% in computational time compared to the FHS-options method.

2.4 Robustness checks

This section presents additional results, with respect to four different robustness checks we perform. First, we extend our analysis by using different forecasting horizons. Second, we consider alternative weights in the optimisation function given in equation (2.2.27), in order to adjust for the imbalance of the maturity weights caused by the equal weights given to the volatility indices. Third, we calculate pricing error statistics using different weighting approaches applied to the pricing errors of different volatility indices. Fourth, we present the robustness of our findings when computing the results using three indices only, which allows us to extend our sample period to include the 2008 financial crisis.

2.4.1 Alternative time horizons

Our previous findings show that the FHS-VI method significantly outperforms the Normal-VIX method for each model specification when forecasting VIX one-week-ahead ($h = 5$). In this section, we extend our analysis and report results for one-day-ahead ($h = 1$) and four-week-ahead ($h = 20$) VIX forecasts. To show the robustness of our results, we report both the DM test and SPA test implications for the three forecast horizons given above.

Panel B of Table 2.3.4 reports the DM test statistics using MSE for one-day-ahead, one-week-ahead and four-week-ahead VIX forecasting, respectively. Instead of the yearly analysis in Section 2.3, we only compare the model performance of the overall sample period, i.e., 2011-2017. The DM1 statistics denote the DM statistics comparing the GARCH models that use the FHS-VI method with their counterparts that use the Normal-VIX method. For one-day-ahead and one-week-ahead forecasts, the difference in forecasting performance is significantly different from zero when using the two methods. For the longer horizon forecasts, i.e., four-week-ahead forecasts, we can reject the null hypothesis of equal forecast accuracy of the two methods only for the NAGARCH model. The negative DM statistics indicate that all the models based on the FHS-VI approach, except for four-week-ahead forecasts of the GJR model, generate smaller average MSE than their counterparts based on the Normal-VIX method. Consistent with the test criteria in Section 2.3, the DM2 statistic in Panel B presents the out-of-sample forecast performance of the models when considering the NAGARCH based on the FHS-VI method as the benchmark model. The DM test statistics show that the NAGARCH model based on the FHS-VI method outperforms all the other models for weekly and monthly forecast horizons. For one-day-ahead forecasts, the NAGARCH model based on the FHS-VI method is found to have a superior predictive ability compared with the models that use the Normal-VIX method. On the other hand, the difference in average MSE loss favours the GJR model that uses the FHS-VI method for daily forecasts, though the difference is not statistically significant.

Panel C and Panel D of Table 2.3.5 present results on the SPA test based on forecasts for different horizons. For each model, the remaining five models are treated as competing models. As discussed above, p-values close to 1 indicate that we can not reject the null hypothesis of the benchmark model being superior to the other models. Both panels show evidence of a similar pattern of forecast ability: the NAGARCH model based on FHS-VI is found to be superior to all the other models for long-term volatility forecasts ($h = 5$ and $h = 20$), while the GJR model based on the FHS-VI method outperforms all the other models for short-run volatility forecasts[19].

2.4.2 Alternative weights used in the optimisation function

In Section 2.3, we assume each volatility index has the same weight in the optimisation function of equation (2.2.27). This weighting, however, places too much weight on the nearby risk-neutral volatilities. The volatilities of the first 7 days are included in all four indices, the volatilities of the first 22 days are included in three indices and so on. In this section, we consider weights in equation (2.2.27) that avoid this increased reliance on nearby maturities, and instead consider a set of index weights that would align the weights of the different volatility maturities. The adjusted RMSE is computed as in equation (2.2.27), but with modified

[19] The reason might be because the GJR model and the NAGARCH model use different forms of parameter to catch the leverage effect, i.e., the GJR model uses an indicator function while NAGARCH employs a quadratic one.

Table 2.4.1: In-sample pricing errors based on the optimisation function using time-weighting

Loss function		Time-weighting		Value-weighting		Equal-weighting				
Year	Model	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE%	RMSE%	QLIKE
2011	GARCH	0.228	0.518	0.316	0.643	0.322	0.650	1.231	1.924	0.001
	GJR	**0.204**	**0.499**	**0.285**	**0.617**	**0.290**	**0.625**	**1.112**	**1.810**	**0.001**
	NAGARCH	0.820	1.338	0.908	1.449	0.914	1.456	3.608	5.614	0.008
2012	GARCH	**0.176**	**0.313**	**0.242**	**0.385**	**0.257**	**0.401**	**1.356**	**2.120**	**0.001**
	GJR	0.182	0.316	0.244	0.386	0.262	0.404	1.407	2.174	0.001
	NAGARCH	0.585	1.024	0.662	1.076	0.674	1.085	3.393	5.356	0.008
2013	GARCH	**0.155**	**0.260**	**0.209**	**0.323**	**0.219**	**0.333**	**1.461**	**2.162**	**0.001**
	GJR	0.185	0.297	0.244	0.366	0.253	0.377	1.670	2.452	0.001
	NAGARCH	0.470	0.823	0.513	0.864	0.520	0.872	3.270	5.247	0.007
2014	GARCH	**0.146**	**0.302**	0.196	**0.366**	0.204	**0.378**	1.285	**2.019**	**0.001**
	GJR	0.149	0.307	**0.194**	0.369	**0.202**	0.382	**1.279**	2.065	0.001
	NAGARCH	0.366	0.839	0.415	0.907	0.424	0.922	2.578	4.976	0.007
2015	GARCH	0.219	0.382	0.294	0.465	0.305	0.478	1.684	2.432	0.001
	GJR	**0.209**	**0.375**	**0.284**	**0.457**	**0.295**	**0.469**	**1.617**	**2.345**	**0.001**
	NAGARCH	0.505	0.834	0.613	0.936	0.625	0.949	3.366	4.864	0.006
2016	GARCH	0.213	**0.431**	0.281	**0.511**	0.295	**0.535**	1.654	**2.600**	0.001
	GJR	**0.205**	0.437	**0.277**	0.519	**0.292**	0.544	**1.648**	2.650	0.001
	NAGARCH	0.373	0.759	0.459	0.873	0.477	0.903	2.507	4.026	0.004
2017	GARCH	0.141	0.300	0.188	0.351	0.200	0.372	1.605	2.758	0.001
	GJR	**0.136**	**0.289**	**0.178**	**0.334**	**0.190**	**0.355**	**1.527**	**2.586**	**0.001**
	NAGARCH	0.181	0.407	0.222	0.464	0.236	0.499	1.827	3.209	0.003
Overall	GARCH	0.183	0.368	0.247	0.447	0.257	0.462	1.468	**2.307**	**0.001**
	GJR	**0.181**	**0.368**	**0.244**	**0.445**	**0.255**	**0.461**	**1.465**	2.314	0.001
	NAGARCH	0.471	0.899	0.541	0.975	0.552	0.991	2.933	4.818	0.006

This table presents the in-sample pricing errors based on time weighting in the optimisation process. MAE is the average absolute error between the market price and the model price; RMSE is the square root of the averaged squared pricing error; these are calculated using time-weighting, value-weighting and equal weighting. MAE% and RMSE% are in relative terms expressed in percentages. QLIKE is defined in equation (2.2.30). Numbers in bold are the lowest values among different models for a given year.

Table 2.4.2: Out-of-sample pricing errors based on the optimisation function using time-weighting

Loss function		Time-weighting		Value-weighting		Equal-weighting				
Year	Model	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE%	RMSE%	QLIKE
2011	GARCH	**2.266**	3.771	2.589	4.330	2.619	4.375	9.763	14.115	0.057
	GJR	2.268	3.769	2.593	4.331	2.623	4.376	9.796	14.126	0.057
	NAGARCH	2.282	**3.660**	**2.529**	**4.174**	**2.552**	**4.215**	**9.440**	**13.119**	**0.051**
2012	GARCH	1.619	2.082	1.720	2.227	1.762	2.282	9.538	13.010	0.031
	GJR	1.616	2.074	1.736	2.224	1.778	2.277	9.626	13.007	0.031
	NAGARCH	**1.552**	**2.010**	**1.643**	**2.125**	**1.681**	**2.169**	**8.935**	**11.901**	**0.027**
2013	GARCH	**1.031**	**1.399**	**1.175**	**1.565**	**1.211**	**1.606**	**8.196**	10.998	**0.023**
	GJR	1.065	1.440	1.215	1.610	1.252	1.652	8.512	11.349	0.026
	NAGARCH	1.105	1.491	1.222	1.619	1.252	1.653	8.213	**10.673**	0.024
2014	GARCH	1.316	2.238	1.499	2.548	1.546	2.622	9.778	15.328	0.052
	GJR	1.315	**2.237**	1.509	2.546	1.558	2.621	9.902	15.385	0.052
	NAGARCH	**1.310**	2.248	**1.476**	**2.525**	**1.520**	**2.593**	**9.454**	**14.844**	**0.051**
2015	GARCH	1.515	2.534	1.710	2.880	1.758	2.957	9.707	14.381	0.058
	GJR	1.597	2.559	1.760	2.882	1.803	2.956	9.958	14.427	**0.055**
	NAGARCH	**1.449**	**2.497**	**1.623**	**2.848**	**1.665**	**2.925**	**9.030**	**13.395**	0.058
2016	GARCH	1.645	2.280	1.837	**2.553**	1.916	**2.654**	11.514	15.823	0.049
	GJR	1.618	**2.273**	1.845	2.567	1.927	2.669	11.587	15.921	0.049
	NAGARCH	**1.575**	2.280	**1.797**	2.576	**1.879**	2.681	**11.168**	**15.495**	**0.049**
2017	GARCH	0.924	1.450	1.018	1.062	1.093	1.714	9.681	15.774	0.038
	GJR	0.920	1.455	1.022	1.613	1.098	1.726	9.713	15.836	0.039
	NAGARCH	**0.860**	**1.333**	**0.951**	**1.474**	**1.023**	**1.579**	**8.983**	**13.995**	**0.034**
Overall	GARCH	1.324	2.294	1.330	2.503	1.702	2.743	9.744	14.306	0.044
	GJR	1.318	2.272	1.320	2.476	1.721	2.749	9.875	14.387	0.044
	NAGARCH	**1.289**	**2.231**	**1.277**	**2.420**	**1.654**	**2.680**	**9.322**	**13.445**	**0.042**

This table presents the out-of-sample pricing errors based on time weighting in the optimisation process. MAE is the average absolute error between the market price and the model price; RMSE is the square root of the average squared pricing error; these are calculated using time-weighting, value-weighting and equal weighting. MAE% and RMSE% are in relative terms expressed in percentages. QLIKE is defined in equation (2.2.30). Numbers in bold are the lowest values among different models for a given year.

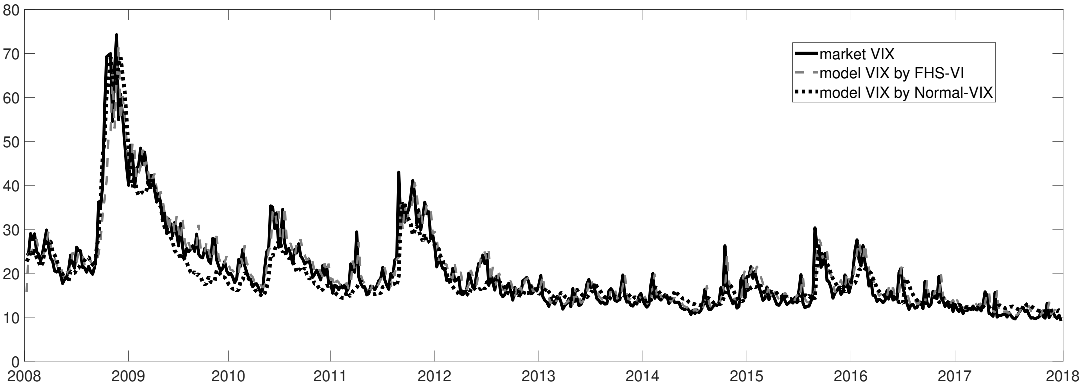
weights w_k calculated as follows: the four indices involve the risk-neutral volatilities of the next 126 days; we divide this into four periods according to the time horizons embedded in the volatility index. Period 1 includes the first 7 days, period 2 consists of day 8 to 22, period 3 is day 23 to 63, and period 4 is day 64 to day 126. If we use equation (2.2.27) with equal index weights w_k, the actual weights of the periods are 0.375, 0.25, 0.25 and 0.125, respectively. In this section we modify the weights of the volatility indices so that each period has the same weight; the modified weights of the volatility indices are then $w_1 = 0.125$, $w_2 = 0.25$, $w_3 = 0.125$ and $w_4 = 0.5$.

The right panels of Table 2.4.1 and Table 2.4.2 report the in-sample and out-of-sample pricing errors using modified weights in the optimisation function. The results are consistent with our earlier findings: the GJR model has the lowest pricing errors for most of the years in-sample, and, on the other hand, for the out-of-sample comparison, the NAGARCH model generates the smallest pricing errors in most cases. Notably, using the modified weights optimisation, both in-sample and out-of-sample pricing errors obtained with the FHS-VI method are lower than the pricing errors based on the Normal-VIX method, reported in Table 2.3.2 and Table 2.3.3.

2.4.3 Alternative weights used in the loss functions

In this section, we discuss the pricing error statistics based on modified weights for the volatility index in the loss functions - noting that our earlier results are based on equal weighting in equations (2.2.28) and (2.2.29). First, we modify the weights in the loss function to remove the increased reliance on the nearby volatilities, as in the previous section (we call this approach time-weighting). Second, we consider the loss functions in which the weights are proportional to the value of the volatility index (value-weighting). The loss functions are computed as in equation (2.2.28) and (2.2.29), but using non-equal weights. As such, we have two sets of alternative weights: w_k can be computed using the calculation detailed in Section 2.4.2, which equalises the effects of the different volatility maturities; or the weights can be considered to be proportional with the market values of the indices. The results based on the modified weights as above are reported in the left panel of Table 2.4.1 for in-sample comparison, and in Table 2.4.2 for out-of-sample comparison. Both sets of results are very similar to our findings based on the equally-weighted loss functions, i.e., the GJR model based on the FHS-VI method has the smallest pricing errors in-sample and the NAGARCH model that uses FHS-VI has the lowest pricing errors out-of-sample.

Figure 2.4.1: Out-of-sample comparison of VIX forecasts obtained using Normal-VIX, FHS-VI based on three indices, assuming GARCH, and CBOE VIX



2.4.4 Results based on three indices only

As mentioned in Section 2.3.1, our sample starts on 2 January 2011 due to the data availability of the VIX9D index. In this section, the estimation is carried out based on three indices only (VIX, VIX3m and VIX6m). This allows us to extend our sample with 3 additional years, starting on 7 January 2008, which is the starting date of VIX6M, with the added bonus that the financial crisis of 2008 is now included in the sample. Figure 2.4.1 presents the one-week-ahead VIX forecasts produced using three indices only, for the GARCH model. To be noted that the VIX reaches very high values during the financial crisis.

In Table 2.4.3 we compare the VIX forecasting performance of different models by calculating the p-values of the SPA test based on three indices. When forecasting one-day-ahead VIX, the p-values computed using the MSE and QLIKE loss functions for the GJR model based on the FHS-VI method are equal to 1,

Table 2.4.3: Out-of-sample SPA test results based on three indices

Horizon	FHS-VI			Normal-VIX		
	GARCH	GJR	NAGARCH	GARCH	GJR	NAGARCH
Panel A. Evaluation by MSE						
h=1	0.266	**1.000**	0.002	0.000	0.003	0.001
h=5	**1.000**	0.479	0.428	0.006	0.006	0.031
h=20	**1.000**	0.247	0.896	0.000	0.294	0.171
Panel B. Evaluation by QLIKE						
h=1	0.032	**1.000**	0.007	0.000	0.000	0.000
h=5	0.637	**1.000**	0.178	0.000	0.000	0.005
h=20	0.532	0.036	**1.000**	0.000	0.049	0.082

This table presents the SPA test statistics for VIX forecasts obtained using two loss functions for different horizons. The SPA test statistic is used to test the null hypothesis that the benchmark model is not outperformed by the competing models. The benchmark model is given at the top of the table. The number of bootstrap replications to calculate the p-values is 10,000. The values in bold are the highest SPA p-values for a given horizon.

indicating that we can not reject the null hypothesis that this model is superior to the other models for one-day-ahead forecasts. On the other hand, we find mixed evidence for longer-term forecasts. Using the MSE loss function, the GARCH models based on the FHS-VI method for weekly and monthly forecasts are found to be superior to the other models. However, when using the QLIKE loss function, the GJR model and the NAGARCH model based on the FHS-VI approach are found to be superior for weekly and monthly forecasts, respectively. It is also notable that the p-values based on the Normal-VIX method are much smaller than those based on the FHS-VI. Overall, for longer-term forecasts, the models based on FHS-VI outperform the models based on the Normal-VIX method, but it is difficult to differentiate among the FHS-VI based models in terms of superior predictive ability.

2.5 Conclusions

In this chapter, we propose to estimate several different GARCH models by using filtered historical simulations and a set of volatility indices. This approach produces estimates using the empirical innovation density that can accommodate for nonstandard features, such as negative skewness and positive excess kurtosis. To reduce the computational burden of using option prices, we employ four well-established volatility indices, i.e., the VIX9D, VIX, VIX3M and VIX6M, to do the calibration. We obtain that this approach dominates the alternative estimation method which only uses the VIX index and assumes a normal distribution, i.e., the Normal-VIX method. This outperformance holds both in-sample and out-of-sample for most of the years; we perform several robustness checks that confirm our results. Additionally, the parameter estimates are shown to be very stable compared to the FHS-options method and significantly reduce the computational time. An empirical analysis on the performance of our proposed estimation for option pricing would be a challenging exercise that we leave for future study.

Appendices

This supplemental appendix provides additional tables and figures.

Table A.1: Parameter estimates obtained using different methods for 2011

P measure	$\omega \times 10^5$		β		α		γ		θ		Ann. vol.	
Model	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev
Panel A. FHS-VI												
GARCH	0.153	0.009	0.907	0.003	0.085	0.003	-	-	-	-	0.207	0.099
GJR	0.174	0.013	0.917	0.004	0.000	0.000	0.143	0.006	-	-	0.213	0.108
NAGARCH	0.232	0.015	0.829	0.006	0.065	0.003	-	-	1.243	0.030	0.202	0.097
Q measure	$\omega^* \times 10^5$		β^*		α^*		γ^*		θ^*		Ann. vol.	
Model	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev
Panel B. FHS-VI												
GARCH	0.217	0.231	0.918	0.037	0.068	0.033	-	-	-	-	0.192	0.014
GJR	0.159	0.211	0.907	0.072	0.038	0.036	0.089	0.133	-	-	0.213	0.019
NAGARCH	0.173	0.216	0.878	0.067	0.062	0.029	-	-	1.037	0.306	0.209	0.024
Panel C. FHS-options												
GJR	1.027	2.252	0.866	0.117	0.017	0.040	0.164	0.088	-	-	0.237	0.025
Panel D. Normal-VIX												
GARCH	0.205	0.002	0.947	0.003	0.047	0.003	-	-	-	-	0.108	0.006
GJR	0.177	0.004	0.961	0.001	0.002	0.002	0.064	0.002	-	-	0.111	0.007
NAGARCH	0.211	0.003	0.936	0.002	0.039	0.001	-	-	0.733	0.067	0.110	0.007

This table presents the parameter estimates obtained using FHS-VI, FHS-options and Normal-VIX for the year 2011. Panel A reports the FHS-VI estimates under physical measure. Panel B, C and D present the parameter estimates obtained using the FHS-VI, FHS-options and Normal-VIX under the risk-neutral measure, respectively. Ann. vol. is the annualized volatility.

Table A.2: Parameter estimates obtained using different methods for 2012

P measure	$\omega \times 10^6$		β		α		γ		θ		Ann. vol.	
Model	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev
Panel A. FHS-VI												
GARCH	1.521	0.048	0.906	0.002	0.085	0.002	-	-	-	-	0.131	0.029
GJR	1.654	0.049	0.918	0.001	0.000	0.000	0.141	0.002	-	-	0.125	0.036
NAGARCH	2.094	0.070	0.830	0.005	0.061	0.002	-	-	1.272	0.061	0.130	0.036
Q measure	$\omega^* \times 10^6$		β^*		α^*		γ^*		θ^*		Ann. vol.	
Model	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev
Panel B. FHS-VI												
GARCH	1.321	1.776	0.914	0.032	0.075	0.026	-	-	-	-	0.200	0.016
GJR	0.832	1.358	0.860	0.066	0.050	0.030	0.144	0.080	-	-	0.173	0.015
NAGARCH	0.937	1.385	0.865	0.038	0.069	0.019	-	-	1.010	0.304	0.219	0.013
Panel C. FHS-options												
GJR	8.08	1.893	0.837	0.138	0.030	0.062	0.118	0.124	-	-	0.187	0.112
Panel D. Normal-VIX												
GARCH	1.982	0.032	0.949	0.000	0.047	0.001	-	-	-	-	0.164	0.020
GJR	1.711	0.071	0.960	0.002	0.002	0.003	0.064	0.004	-	-	0.168	0.022
NAGARCH	1.954	0.059	0.935	0.003	0.037	0.002	-	-	0.807	0.097	0.163	0.020

This table presents the parameter estimates obtained using FHS-VI, FHS-options and Normal-VIX for the year 2012. Panel A reports the FHS-VI estimates under physical measure. Panel B, C and D present the parameter estimates obtained using the FHS-VI, FHS-options and Normal-VIX under the risk-neutral measure, respectively. Ann. vol. is the annualized volatility.

Table A.3: Parameter estimates obtained using different methods for 2013

P measure	$\omega \times 10^6$		β		α		γ		θ		Ann. vol.	
Model	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev
Panel A. FHS-VI												
GARCH	1.569	0.058	0.904	0.002	0.086	0.002	-	-	-	-	0.127	0.026
GJR	1.764	0.068	0.913	0.002	0.000	0.000	0.147	0.003	-	-	0.124	0.028
NAGARCH	2.134	0.072	0.813	0.004	0.059	0.002	-	-	1.406	0.023	0.119	0.027
Q measure	$\omega^* \times 10^6$		β^*		α^*		γ^*		θ^*		Ann. vol.	
Model	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev
Panel B. FHS-VI												
GARCH	1.373	1.878	0.908	0.032	0.074	0.036	-	-	-	-	0.148	0.017
GJR	1.731	1.930	0.800	0.081	0.040	0.041	0.232	0.149	-	-	0.118	0.014
NAGARCH	1.334	1.657	0.888	0.065	0.050	0.027	-	-	0.953	0.473	0.170	0.020
Panel C. FHS-options												
GJR	1.761	4.282	0.789	0.172	0.032	0.061	0.167	0.168	-	-	0.174	0.048
Panel D. Normal-VIX												
GARCH	1.818	0.065	0.950	0.000	0.045	0.000	-	-	-	-	0.144	0.012
GJR	1.561	0.064	0.961	0.000	0.000	0.000	0.063	0.001	-	-	0.148	0.014
NAGARCH	1.742	0.057	0.929	0.002	0.033	0.001	-	-	1.021	0.043	0.138	0.012

This table presents the parameter estimates obtained using FHS-VI, FHS-options and Normal-VIX for the year 2013. Panel A reports the FHS-VI estimates under physical measure. Panel B, C and D present the parameter estimates obtained using the FHS-VI, FHS-options and Normal-VIX under the risk-neutral measure, respectively. Ann. vol. is the annualized volatility.

Table A.4: Parameter estimates obtained using different methods for 2014

P measure	$\omega \times 10^6$		β		α		γ		θ		Ann. vol.	
Model	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev
Panel A. FHS-VI												
GARCH	1.604	0.066	0.899	0.002	0.089	0.002	-	-	-	-	0.119	0.028
GJR	1.712	0.053	0.910	0.001	0.000	0.000	0.149	0.003	-	-	0.121	0.031
NAGARCH	2.051	0.057	0.807	0.002	0.060	0.002	-	-	1.438	0.027	0.116	0.031
Q measure	$\omega^* \times 10^6$		β^*		α^*		γ^*		θ^*		Ann. vol.	
Model	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev
Panel B. FHS-VI												
GARCH	2.168	2.202	0.914	0.041	0.064	0.031	-	-	-	-	0.174	0.018
GJR	1.803	2.028	0.819	0.093	0.024	0.033	0.219	0.131	-	-	0.140	0.015
NAGARCH	2.387	2.240	0.863	0.090	0.041	0.027	-	-	0.836	0.543	0.139	0.017
Panel C. FHS-options												
GJR	2.173	4.062	0.824	0.105	0.024	0.035	0.099	0.154	-	-	0.232	0.054
Panel D. Normal-VIX												
GARCH	1.681	0.045	0.950	0.000	0.045	0.000	-	-	-	-	0.139	0.018
GJR	1.582	0.051	0.955	0.001	0.000	0.000	0.054	0.001	-	-	0.147	0.017
NAGARCH	1.588	0.050	0.928	0.003	0.031	0.001	-	-	1.168	0.118	0.137	0.017

This table presents the parameter estimates obtained using FHS-VI, FHS-options and Normal-VIX for the year 2014. Panel A reports the FHS-VI estimates under physical measure. Panel B, C and D present the parameter estimates obtained using the FHS-VI, FHS-options and Normal-VIX under the risk-neutral measure, respectively. Ann. vol. is the annualized volatility.

Table A.5: Parameter estimates obtained using different methods for 2015

P measure	$\omega \times 10^6$		β		α		γ		θ		Ann. vol.	
Model	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev
Panel A. FHS-VI												
GARCH	1.716	0.069	0.894	0.003	0.092	0.002	-	-	-	-	0.151	0.047
GJR	1.817	0.078	0.903	0.003	0.000	0.000	0.159	0.003	-	-	0.150	0.053
NAGARCH	2.143	0.071	0.797	0.006	0.066	0.002	-	-	1.392	0.019	0.150	0.051
Q measure	$\omega^* \times 10^6$		β^*		α^*		γ^*		θ^*		Ann. vol.	
Model	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev
Panel B. FHS-VI												
GARCH	2.101	2.166	0.909	0.036	0.075	0.034	-	-	-	-	0.207	0.021
GJR	2.066	2.157	0.768	0.075	0.020	0.032	0.317	0.149	-	-	0.149	0.018
NAGARCH	1.889	1.938	0.850	0.062	0.058	0.020	-	-	1.135	0.400	0.185	0.025
Panel C. FHS-options												
GJR	3.542	4.201	0.809	0.171	0.071	0.085	0.159	0.131	-	-	0.148	0.058
Panel D. Normal-VIX												
GARCH	1.594	0.024	0.950	0.000	0.045	0.000	-	-	-	-	0.170	0.037
GJR	1.375	0.020	0.962	0.001	0.000	0.002	0.062	0.002	-	-	0.176	0.037
NAGARCH	1.619	0.217	0.930	0.005	0.034	0.087	-	-	0.956	0.070	0.169	0.034

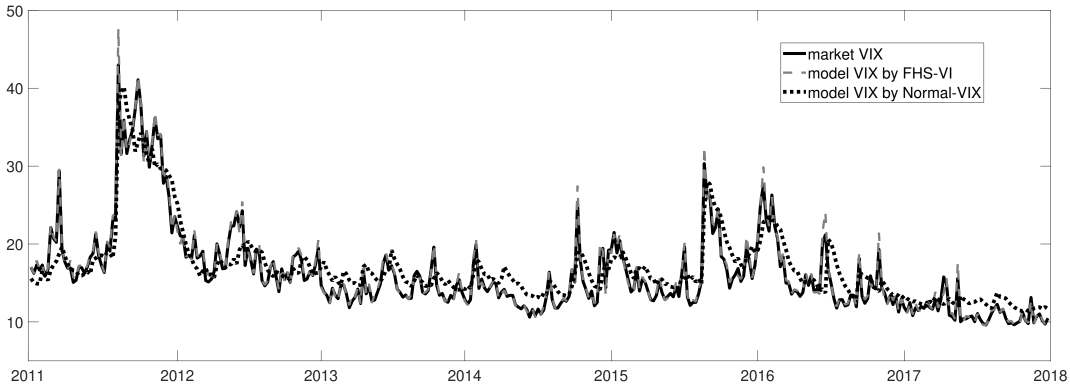
This table presents the parameter estimates obtained using FHS-VI, FHS-options and Normal-VIX for the year 2015. Panel A reports the FHS-VI estimates under physical measure. Panel B, C and D present the parameter estimates obtained using the FHS-VI, FHS-options and Normal-VIX under the risk-neutral measure, respectively. Ann. vol. is the annualized volatility.

Table A.6: Parameter estimates obtained using different methods for 2016

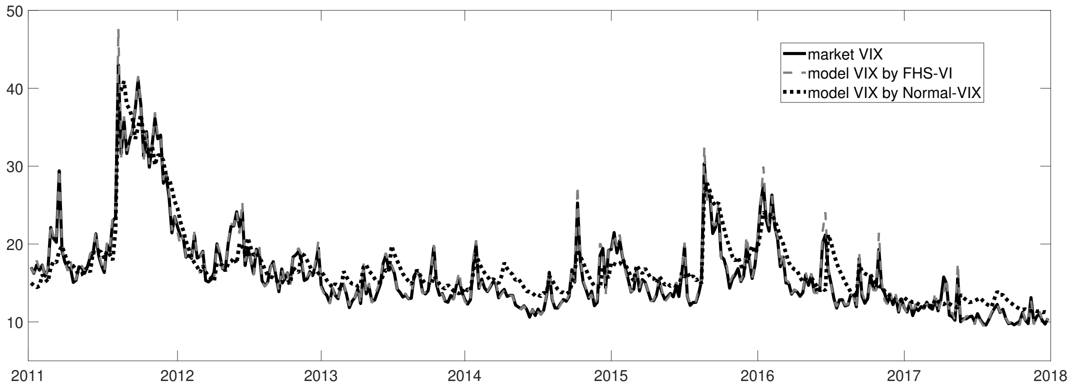
P measure	$\omega \times 10^6$		β		α		γ		θ		Ann. vol.	
Model	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev
Panel A. FHS-VI												
GARCH	2.038	0.214	0.884	0.006	0.098	0.003	-	-	-	-	0.136	0.048
GJR	2.118	0.154	0.894	0.003	0.000	0.000	0.169	0.003	-	-	0.135	0.055
NAGARCH	2.351	1.661	0.780	0.006	0.072	0.002	-	-	1.385	0.115	0.136	0.059
Q measure	$\omega^* \times 10^6$		β^*		α^*		γ^*		θ^*		Ann. vol.	
Model	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev
Panel B. FHS-VI												
GARCH	1.842	1.899	0.912	0.050	0.059	0.029	-	-	-	-	0.203	0.018
GJR	1.378	1.916	0.764	0.073	0.030	0.029	0.271	0.108	-	-	0.161	0.015
NAGARCH	2.651	2.026	0.858	0.100	0.042	0.030	-	-	0.683	0.610	0.171	0.023
Panel C. FHS-options												
GJR	6.073	4.142	0.714	0.067	0.041	0.014	0.121	0.108	-	-	0.123	0.015
Panel D. Normal-VIX												
GARCH	1.660	0.054	0.946	0.002	0.048	0.002	-	-	-	-	0.155	0.034
GJR	1.452	0.008	0.958	0.003	0.004	0.005	0.060	0.005	-	-	0.162	0.036
NAGARCH	1.689	0.006	0.930	0.002	0.038	0.003	-	-	0.845	0.084	0.154	0.035

This table presents the parameter estimates obtained using FHS-VI, FHS-options and Normal-VIX for the year 2016. Panel A reports the FHS-VI estimates under physical measure. Panel B, C and D present the parameter estimates obtained using the FHS-VI, FHS-options and Normal-VIX under the risk-neutral measure, respectively. Ann. vol. is the annualized volatility.

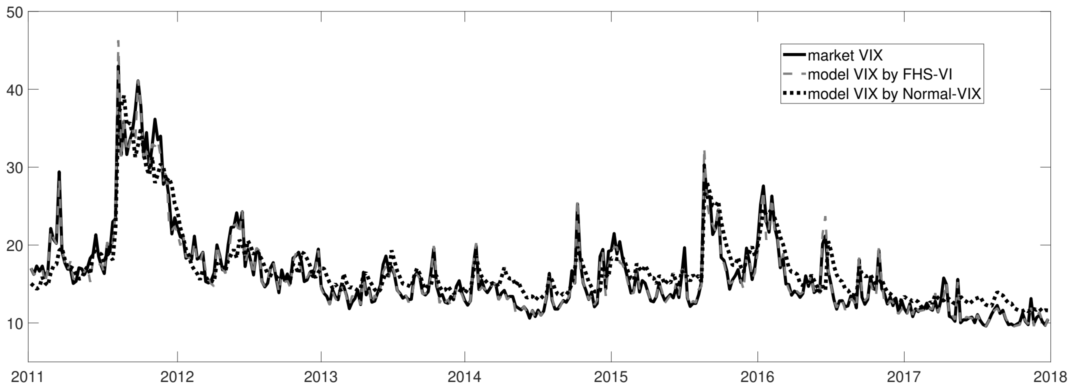
Figure A.1: In-sample comparison of the model VIX and the CBOE VIX



(a) **GARCH**

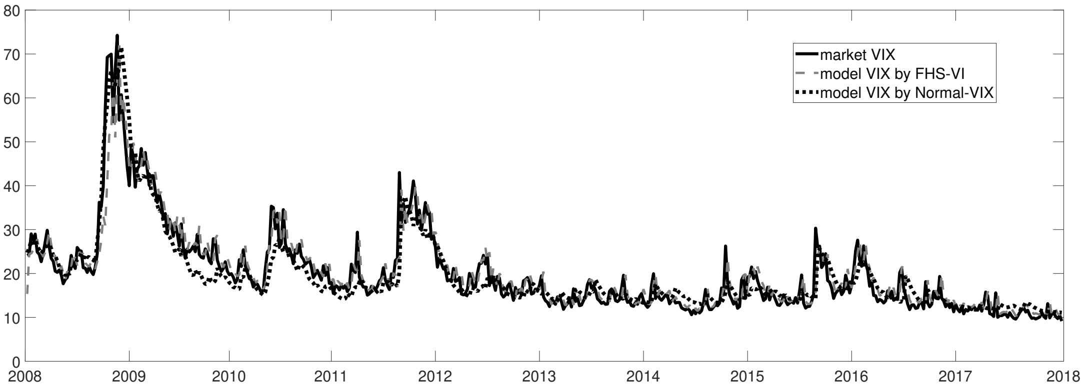


(b) **GJR**

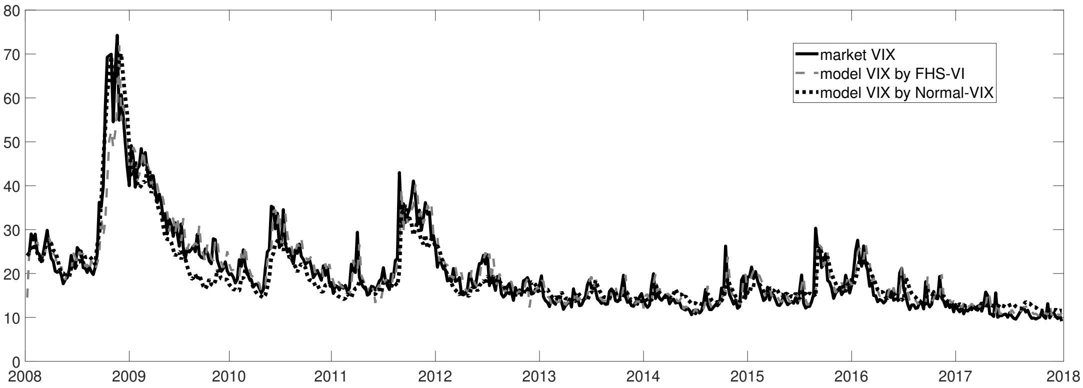


(c) **NAGARCH**

Figure A.2: Out-of-sample comparison of VIX forecasts obtained using Normal-VIX, FHS-VI based on three indices, and CBOE VIX



(a) GJR



(b) NAGARCH

Chapter 3

Does VIX term structure help to predict VIX futures prices: Evidence from COVID-19 Crisis

3.1 Introduction

The Chicago Board Options Exchange (CBOE) VIX measures the expected volatility associated with the S&P 500 index over the following 30 days, as implied by stock index option prices. The VIX index is often referred as the ’fear gauge’ by investors since it is established in 1993 (see Whaley, 2009). However, since the VIX index is not a tradable asset, the CBOE introduced VIX futures on March 26, 2004 to broaden hedging opportunities in volatility trading. Since its launch, the liquidity of the VIX futures market is steadily growing. For example, on September 8, 2020, the trading volume is 280,180 contracts or 8.9 billion USD in terms of market value[1]. Therefore, forecasting VIX futures prices are of great interest to academics and practitioners.

Apart from the VIX index which measures the implied volatility over the next 30 days, the CBOE also lists anther four volatility indices with different maturities to measure the implied volatility term structure of S&P 500 index: the CBOE S&P 500 9-day Volatility Index (VIX9D), the CBOE S&P 500 3-month Volatility Index (VIX3M), the CBOE S&P 500 6-month Volatility Index (VIX6M) and the CBOE S&P 500 1-year Volatility Index (VIX1Y). Moreover, the volatility term structure is found to be important in VIX futures pricing by Zhu and Zhang (2007). In this chapter, we explore whether combining VIX term structure and futures prices can improve the futures forecasting performance. In particular, we investigate how the VIX term structure affects the VIX futures forecasting under different conditions.

One intuitive determinant of the VIX futures prices is its underlying, i.e., the VIX index. However, given that the VIX itself is not easily traded, the no-arbitrage principle cannot be used to obtain a simple formula between the VIX futures prices and the spot VIX values as in the stock market[2]. Various stud-

[1] The data can be found from 'CBOE Futures Exchange Daily Market Statistics' via the following link: https://markets.cboe.com/us/futures/market_statistics/daily/.

[2] Theoretically, it is possible to trade VIX by replicating a portfolio of S&P 500 options, thus the VIX futures can be priced using the no-arbitrage principle, for example see Zhu and Zhang (2007)

ies focus on deriving the VIX futures prices through the instantaneous variance embedded by VIX via different volatility models. For example, Zhang and Zhu (2006) study VIX future prices with the Heston (1993) model. Lin (2007) and Zhu and Lian (2012) model the variance with simultaneous jumps in both the returns and the volatility process. In the discrete-time setting, Guo and Liu (2020) and Xie et al. (2020) propose new solutions under the model of Glosten et al. (1993) (GJR). A second strand of literature studies empirically the behaviours of the VIX futures and thus forecasts prices. For example, Dotsis et al. (2007) use the VIX futures data to evidence that the addition of jumps to a square root process improves the pricing performance. Zhang et al. (2010) also calibrate a mean-reverting variance model with jumps using VIX futures prices. Other related discussions on VIX futures include Konstantinidi and Skiadopoulos (2011a), Mencía and Sentana (2013), Taylor (2019), Ballestra et al. (2019) etc.

Most recently, several studies attempt to evaluate VIX futures prices with the model parameters estimated from both the underlying, i.e, VIX and the corresponding futures. Wang et al. (2017) propose a closed-form pricing formula based on the Heston and Nandi (2000) GARCH model using several data sets. They show that the joint estimation with VIX and the futures performs the best in terms of fitting the market VIX and the VIX futures prices simultaneously. Further, Huang et al. (2019b) calibrate an extended model of Majewski et al. (2015) with mixed information from the S&P 500 returns, the VIX term structure and VIX futures prices.

Despite both Wang et al. (2017) and Huang et al. (2019b) use the VIX/VIX term structure and the VIX futures, neither study shows statistical evidence whether the VIX term structure can help to predict the VIX futures prices. In this context, our paper also studies the forecasting accuracy of pricing methods but differs in terms of the data that it uses. We calibrate the GJR model with the filtered historical simulation (FHS) method proposed by Barone-Adesi et al. (2008) using different data. Then the risk-neutral expected value of daily variance can be updated accordingly, and we also obtain the expectation of the VIX squared for maturity T, which is the forward-starting variance swap. The study of Carr and Wu (2006) suggests that the VIX futures price is the difference between this forward variance swap and the risk-neutral variance of the VIX futures. Furthermore, Dupire (2006) and the CBOE white paper show that the fair price of VIX futures is equal to the price of forward variance minus a concavity adjustment, which can be expressed using the CBOE VIX volatility index, i.e., the VVIX[3]. Therefore, in this chapter, we evaluate the in-sample and out-of-sample pricing performance for the fair value of the VIX futures by using not only the VIX term structure or/and VIX futures data, but also the VVIX term structure.

The contributions of this chapter are threefold. First, this chapter is among the first to discuss how the VIX term structure affects the performance of VIX

[3]For the details of VVIX, please see Section 3.2.2.

futures pricing. We find that the addition of the VIX term structure to the VIX futures improves the VIX futures forecasting. The improvement is remarkable for the long-term VIX futures or for periods when the level of VIX is high. More interestingly, we also find that, during the 2020 COVID-19 pandemic, the joint estimation with VIX9D, VIX, VIX3M, VIX6M and the VIX futures provides the lowest pricing errors among all the methods. Additionally, we implement three statistical tests, i.e., the Diebold and Mariano (1995) test, the test of Giacomini and White (2006) and the Model Confidence Set (MCS) by Hansen et al. (2011), to evidence our results.

Second, different from the normal distribution assumption of returns used in the majority of the VIX futures pricing literature, the FHS method utilises the empirical innovation density extracted from historical returns. Also, Jiang and Lazar (2020) demonstrate that the GJR model with FHS provides better VIX forecasting performance than the traditional local risk-neutral valuation relationship (LRNVR) proposed by Duan (1995). In addition to the seven methods with different data sources applied to the GJR models with FHS, we also implement the model of Xie et al. (2020), which uses the GJR model under the LRNVR, as the benchmark model.

Our third contribution is the use of the VVIX term structure for VIX futures pricing. To our best knowledge, little (if any) literature considers the VVIX term structure in pricing VIX futures. To predict one-day-ahead VIX future prices, we also produce the VVIX forecasts through linear interpolation via the vector autoregression (VAR) model.

The remainder of this chapter is organized as follows. Section 3.2 describes the theoretical background of the model. Section 3.3 details the model estimation and forecasting procedure. Section 3.4 presents our results and analysis, and Section 3.5 concludes.

3.2 The model

3.2.1 GJR specifications

Following Barone-Adesi et al. (2008) and Jiang and Lazar (2020), we assume that the logarithm of the asset returns is governed by the GJR process of Glosten et al. (1993) with an empirical innovation density[4]. Under the physical measure $\mathbb{P}$:

$$\begin{aligned} &\ln(S_t/S_{t-1}) = \mu_t - \kappa_t + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t \\ &\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + [\alpha + \gamma I(\varepsilon_{t-1} < 0)]\varepsilon_{t-1}^2, \end{aligned} \tag{3.2.1}$$

where S_t is the S&P 500 index price at time t, μ_t is the expected rate of return, $z_t \mid \mathcal{F}_{t-1} \sim F(0,1)$, $\mathcal{F}_{t-1}$ is the information set up to time $t-1$. F

[4]Jiang and Lazar (2020) show that the GJR model is the best to forecast the short-term volatility. Also, there are existing literature using GJR model to price VIX futures, see Guo and Liu (2020) and Xie et al. (2020).

follows some unknown distribution with zero mean and unit variance which we estimate using the empirical distribution, and κ_t is the mean correction factor $\kappa_t = \ln(E_{t-1}[\exp\{\varepsilon_t\}])$. The dummy variable $I_t = 1$ when $z_t < 0$ and $I_t = 0$, otherwise. The leverage effect is captured by a positive γ.

Under the risk-neutral measure, we assume the following dynamics:

$$\begin{aligned} \ln(S_t/S_{t-1}) &= r_t - \kappa_t^* + \varepsilon_t^*, \quad \varepsilon_t^* = \sigma_t^* z_t^* \\ \sigma_t^{*2} &= \omega^* + \beta^* \sigma_{t-1}^{*2} + [\alpha^* + \gamma^* I(\varepsilon_{t-1}^* < 0)]\varepsilon_{t-1}^{*2}. \end{aligned} \tag{3.2.2}$$

where r_t is the risk-free rate at time t, z_t^* is assumed to follow the same distribution F as under the physical measure, κ_t^* is the mean correction factor under the risk-neutral measure, and $\theta^* = \{\omega^*, \alpha^*, \beta^*, \gamma^*\}$ are a set of risk-neutral parameters which are allowed to be different from those under the physical measure.

3.2.2 VIX and VVIX

The volatility index VIX measures the market participants' risk-neutral expectation of return volatility implied from options prices. The VIX index is computed from the out-of-money (OTM) S&P 500 option prices via:

$$VIX_t = 100 \times \sqrt{\frac{2}{\tau}\sum_i \frac{\triangle K_i}{K_i^2} e^{r\tau} Q(K_i) - \frac{1}{\tau}[\frac{F}{K_0} - 1]^2}, \tag{3.2.3}$$

where τ is 30 days, $Q(K_i)$ is the option price for strike K_i, $\triangle K_i$ is the interval between strike prices, F is the forward price of S&P 500, and K_0 is the first strike that is below the forward index level F.

In a discrete time setting, as described in Hao and Zhang (2013), the VIX can be obtained by taking the arithmetic average of the expected daily variance for the following month[5]:

$$VIX_t = 100 \times \sqrt{\frac{252}{22} \times E_t^Q \left[\sum_{\tau=1}^{22} \sigma_{t+\tau}^{*2} \right]}. \tag{3.2.4}$$

Similarly, the model price of the volatility index VIX9D, VIX3M, VIX6M and VIX1Y can be calculated as:

$$VIX9D_t = 100 \times \sqrt{\frac{252}{7} \times E_t^Q \left[\sum_{\tau=1}^{22} \sigma_{t+\tau}^{*2} \right]}, \tag{3.2.5}$$

$$VIX3M_t = 100 \times \sqrt{\frac{252}{63} \times E_t^Q \left[\sum_{\tau=1}^{22} \sigma_{t+\tau}^{*2} \right]}, \tag{3.2.6}$$

$$VIX6M_t = 100 \times \sqrt{\frac{252}{126} \times E_t^Q \left[\sum_{\tau=1}^{22} \sigma_{t+\tau}^{*2} \right]}, \tag{3.2.7}$$

$$VIX1Y_t = 100 \times \sqrt{E_t^Q \left[\sum_{\tau=1}^{22} \sigma_{t+\tau}^{*2} \right]}. \tag{3.2.8}$$

In February 2006, the CBOE launched VIX options which provides market participants more flexibility to trade volatility. Since its introduction, the VIX options market has been growing steadily. Thus, it is natural to study the implied

[5]We apply the trading day count convention in this chapter.

volatility of the VIX index, i.e., the VVIX and its term structure. Adapting the same calculation method for VIX in Equation (3.2.3), the VVIX term structure over different horizons is obtained by using VIX option prices. The VVIX measures the annualised expected volatility of the 30-day forward price of the VIX index[6]. Therefore, at time t, the squared VVIX term structure with an expiration date T can be written as:

$$VVIX_{t,T}^2 = 100 \times \frac{252}{T-t} \times E_t\left[\sum_{\tau=0}^{T-t} \sigma_\tau^2\right], \tag{3.2.9}$$

where σ_τ^2 is the variance of log-returns of the forward prices $F_\tau = VIX_\tau$. The CBOE white paper then illustrates that the squared VVIX term structure can also be approximated by[7]:

$$\begin{aligned} VVIX_{t,T}^2 \approx& 100 \times \frac{252}{T-t} \times VAR_t\left[\ln(\frac{F_T}{F_t})\right] \\ =& 100 \times \frac{252}{T-t} \times VAR_t\left[\ln(F_T)\right]. \end{aligned} \tag{3.2.10}$$

By considering Taylor's expansion for $\ln F_T$, Equation (3.2.10) becomes:

$$\begin{aligned} VVIX_{t,T}^2 \approx& 100 \times \frac{252}{T-t} \times VAR_t\left[\ln(F_t) + \frac{F_T - F_t}{F_t}\right] \\ =& 100 \times \frac{252}{T-t} \times \frac{VAR_t\left[F_T\right]}{F_t^2} \end{aligned} \tag{3.2.11}$$

[6]The VVIX represents the implied volatility of the 22-trading-day forward price of VIX.

[7]The white paper of the VVIX term structure can be found from CBOE website via this link: https://cdn.cboe.com/resources/indices/documents/vvix-termstructure.pdf.

3.2.3 Pricing VIX futures

The VIX futures are exchange-traded at the CBOE Futures Exchange (CFE). Their underlying asset is the VIX index. The contracts are cash settled on the Wednesday that is 30 days prior to the 3rd Friday of the calendar month following the expiring month. The primary purpose of VIX futures is to enable investors to trade and hedge volatility.

The CBOE white paper shows that the fair value of VIX futures at time t with an expiration date T is:

$$Fut_{t,T} = E_t^Q[VIX_T] = \sqrt{E_t^Q[VIX_T^2] - VAR_t[VIX_T]} \tag{3.2.12}$$

where VIX_T is the VIX level at time T. $VAR_t[VIX_T]$ is the variance of the futures price, and from Equation (3.2.11), its value for expiration $F_T = VIX_T$ is approximately equal to:

$$VAR_t[VIX_T] \approx F_t^2 \times \frac{T-t}{252} \times \left(\frac{VVIX_{t,T}}{100}\right)^2. \tag{3.2.13}$$

3.2.4 Benchmark model: XZR approach

For comparison, the model of Xie et al. (2020) (XZR model, hereafter) is used here as our benchmark model. Xie et al. (2020) propose an analytical approximation method to price VIX futures based on the information of model implied VIX.

The pricing formula is obtained assuming the GJR process. Under the physical measure:

$$\begin{aligned} \ln(S_t/S_{t-1}) &= r_t + \lambda\sigma_t - \frac{1}{2}\sigma_t^2 + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t \\ \sigma_t^2 &= \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma I_{t-1}\varepsilon_{t-1}^2, \end{aligned} \tag{3.2.14}$$

where r_t is the risk-free rate at time t, λ is the risk premium, $z_t \mid \mathcal{F}_{t-1} \sim N(0,1)$, and $\mathcal{F}_t$ is the information set up to t. I_t is one when z_t is negative and zero otherwise. Following the local risk-neutral valuation relationship (LRNVR) proposed by Duan (1995), the GJR process under the risk-neutral measure is:

$$\begin{aligned} \ln(S_t/S_{t-1}) &= r_t - \frac{1}{2}\sigma_t^2 + \xi_t, \quad \xi_t = \sigma_t z_t \\ \sigma_t^2 &= \omega + \alpha(\xi_{t-1} - \lambda\sigma_{t-1})^2 + \beta\sigma_{t-1}^2 + \gamma I_{t-1}(\xi_{t-1} - \lambda\sigma_{t-1})^2, \end{aligned} \tag{3.2.15}$$

The GJR parameters $\{\omega, \alpha, \beta, \gamma, \lambda\}$ are then calibrated using maximum likelihood estimation (MLE) based on the following equations:

$$\begin{aligned} VIX_t^{mkt} &= VIX_t^{imp}\eta_t, \quad \eta_t \sim LN(-\sigma_\eta^2/2, \sigma_\eta^2) \\ VIX_t^{imp} &= 100(\sqrt{\Phi + \Psi\sigma_{t+1}^2}) \\ \Phi &= 252 \times \frac{\omega}{1-\rho}(1 - \frac{\Psi}{252}), \quad \Psi = 252\frac{1-\rho^n}{n(1-\rho)} \\ \rho &= \alpha(1+\lambda^2) + \beta + \gamma[\frac{\lambda}{\sqrt{2\pi}}e^{-\frac{\lambda^2}{2}} + (1+\lambda^2)N(\lambda)] \end{aligned} \tag{3.2.16}$$

where VIX_t^{mkt} is the market price of VIX and VIX_t^{imp} is the model implied VIX.

η_t is assumed to be log-normally distributed to make sure VIX_t^{imp} is an unbiased estimator of VIX_t^{mkt}.

The pricing formula of VIX futures is then obtained using the GJR model estimates based on Taylor's expansion[8]:

$$F_{t,T} = E_t^Q[VIX_T] = VIX_t^{imp} E_t^Q[1 + f^{(1)}(h_{T+1} - h_{t+1}) + \frac{f^{(2)}}{2}(h_{T+1} - h_{t+1})^2]. \tag{3.2.17}$$

To simplify the notation, we use h_t to denote σ_t^2. $f^{(1)}$ and $f^{(2)}$ are the first two derivatives of $f(h_{T+1}|h_{t+1})$:

$$f(h_{T+1}|h_{t+1}) = \sqrt{1 + \frac{\Psi(h_{T+1} - h_{t+1})}{\Phi + \Psi h_{t+1}}}. \tag{3.2.18}$$

3.3 Estimation and forecasting

3.3.1 Model estimation

Under the physical measure, we adopt the same estimation procedure as in Barone-Adesi et al. (2008) and Jiang and Lazar (2020). In other words, the GJR model parameters are estimated on each Wednesday using quasi-maximum likeli-

[8]For more details, please see Xie et al. (2020).

hood function (QML) with 3,500 historical returns[9]. Under the risk-neutral measure, the calibration is based on minimising the root mean square error (RMSE) between the prices given by the model and the market prices. To study whether the VIX term structure can help predict the VIX futures prices, we propose seven new methods to calibrate model parameters. Each method differs in term of the data source it uses in the estimation process:

1. Fut+5VIs (e): VIX futures data and all five volatility indices, i.e, VIX9D, VIX, VIX3M, VIX6M and VIX1Y. When considering the volatility indices, we assume equal weighting.

2. Fut+5VIs (t): VIX futures data and all five volatility indices. Following Jiang and Lazar (2020), we also adjust the index weights by taking into account the increased reliance on nearby maturities when using equal weighting. The weights of the volatility indices are then modified based on equal weights for each time period.

3. Fut+4VIs: VIX futures data and four volatility indices, i.e, VIX9D, VIX, VIX3M and VIX6M. Since most of VIX futures data in the sample has a maturity less than six months, we consider excluding the VIX1Y index because of its long maturity.

4. Fut+VIX1Y: VIX futures data and VIX1Y index. When imposing equal

[9]See Jiang and Lazar (2020) for more details about the estimation procedure.

weights on the volatility for each day in a year, we end up with VIX1Y only.

5. Fut: Futures only. It is natural to forecast futures prices using historical information of itself only.

6. 5VIs: All five indices. Apart from using futures data only and futures and VIX term structure, we consider using information on the VIX term structure only.

7. 4VIs: Four volatility indices, i.e, VIX9D, VIX, VIX3M and VIX6M. For the maturities of VIX futures in the sample, we also consider using the four indices only; these are the ones which measure the volatility over a period less than or equal to six months.

In this chapter, we aim to minimise the following expression on each Wednesday with respect to θ^* in Equation (3.2.2)[10]:

$$\sqrt{w_F \times MSE_{Fut} + w_V \times MSE_{VI}} \tag{3.3.1}$$

with

$$MSE_{Fut} = \sum_{i=1}^{m} \left[w_i \times \left(Fut^{(i)market} - Fut^{(i)model} \right)^2 \right] \tag{3.3.2}$$

[10]Other optimisation function could also be considered, for example, the ad-hoc linear model in Kanniainen et al. (2014).

and

$$MSE_{VI} = \sum_{k=1}^{n} \left[w_k \times \left(VI^{(k)market} - VI^{(k)model} \right)^2 \right], \tag{3.3.3}$$

where w_F and w_V are the weights of MSE_{Fut} and MSE_{VI}, respectively[11]. m is the total number of VIX futures that observed in the market on Wednesdays. $w_i = \frac{1}{m}$ is the weight for the i^{th} VIX future on that day. $Fut^{(i)market}$ represents the market price of the i^{th} VIX futures and $Fut^{(i)model}$ is the future price calculated matching the same maturity of $Fut^{(i)market}$ using the models described in Section 3.2. n is the number of volatility indices used in the estimation. w_k is the weight of the volatility index, assuming equal weighting or non-equal weighting to adjust for the different maturities.[12]

3.3.2 VVIX forecasting

According to the VIX futures pricing formula of Equation (3.2.12) and (3.2.13), we also have to model the VVIX value in order to predict VIX futures prices. In this chapter, we apply the VAR(1) model, which assumes that the VVIX time series with different maturities affect each other over time, to forecast the VVIX

[11] The ratio between the number of VIX futures contact and the number of VIX term structure is approximately 2 : 1. To better illustrate the usefulness of VIX term structure, we use $w_F = \frac{1}{2}$ and $w_V = \frac{1}{2}$ when both information of VIX futures and VIX term structure are included in the estimation; the alternative weights depending on the number of futures and VIX could also be considered. When only futures data is included, $w_F = 1$ and $w_V = 0$ and $w_F = 0$ and $w_V = 1$ if only the VIX term structure is used.

[12] When assuming equal weighting, $w_k = \frac{1}{n}$; while using five volatility indices and non-equal weighting, ie., for method Fut+5VIs (t), $w_1 = 0.1$, $w_2 = 0.2$, $w_3 = 0.1$, $w_4 = 0.2$ and $w_5 = 0.4$.

term structure.[13] The forecasted VVIX value for each given maturity is obtained by linear interpolation[14]:

1. A new coordinate is constructed based on 13 points between 10 and 130 days, i.e., [10, 20,...130], by linearly interpolating the existing VVIX term structure[15].

2. Once we compute the forecasted VVIX values based on the coordinate points, the second linear interpolation is performed to match the target forecasting maturity in the sample.

The VAR(1) model is defined as:

$$\triangle VVIX_t = C + A\triangle VVIX_{t-1} + e_t. \tag{3.3.4}$$

where $\triangle VVIX_t$ is a $\mathit{13} \times 1$ vector of daily changes between $t-1$ and t, C is a $\mathit{13} \times 1$ vector of constants, A is a $\mathit{13} \times \mathit{13}$ matrix of coefficients, and e_t is a $\mathit{13} \times 1$ vector of error terms.

[13]We also fit the VVIX series by using the ARMA model which presents the similar results (results are available on request).

[14]We implement two-step linear interpolations. Since the maturities of VVIX do not have any patterns, we use the first interpolation to construct a new coordinate in order to forecast the future values. The forecasted values are then based on new coordinate; therefore, we have to do the second interpolation to match the forecasts to our target maturities.

[15]The VAR(1) model is estimated using 1,500 historical observations or the maximum data available, if this is less than 1,500 observations.

3.3.3 Model evaluation

The forecasts are generated based on a rolling window of 3,500 observations. After each model estimation, the one-day-ahead out-of-sample forecasts are obtained for each Thursday in the sample, i.e., from January 2011 to October 2020. To measure the pricing VIX futures performance of the different models, we compare the following loss functions:

$$MAE = \frac{1}{N}\sum_{j=1}^{N}\left|Fut_j^{market} - Fut_j^{model}\right| \tag{3.3.5}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left[\left(Fut_j^{market} - Fut_j^{model}\right)^2\right]} \tag{3.3.6}$$

$$MAE\% = \frac{1}{N}\sum_{j=1}^{N}\left|\frac{Fut_j^{market}}{Fut_j^{model}} - 1\right| \tag{3.3.7}$$

$$RMSE\% = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left[\left(\frac{Fut_j^{market}}{Fut_j^{model}} - 1\right)^2\right]} \tag{3.3.8}$$

where N is the total number of observations in the sample, and Fut_j^{market} and Fut_j^{model} denote the market price and the model price of VIX futures, respectively.

In addition, to test the significant differences of forecasting accuracy among the models, we employ the DM test proposed by Diebold and Mariano (1995), the conditional predictive ability test proposed by Giacomini and White (2006)

(GW test) and the model confidence set (MCS) proposed by Hansen et al. (2011).

The DM test is a pairwise test with the null hypothesis of equal forecast accuracy of two different models. The two sets of estimated forecast errors are defined as $\hat{e}_{1,t}$ and $\hat{e}_{2,t}$, respectively. The difference in loss between two forecasting methods is denoted as: $d_t = L(\hat{e}_{1,t}) - L(\hat{e}_{2,t})$. The DM statistic below:

$$DM = \frac{\bar{d}}{\sqrt{2\pi\hat{f}_d(0)/T}} \tag{3.3.9}$$

has an asymptotic standard normal distribution under the null, where $\bar{d}$ is the sample mean of the loss differential, and $2\pi\hat{f}_d(0)$ is a consistent estimator of the asymptotic variance[16].

Different from the DM test, the GW test considers conditional predictive ability taking into account estimation uncertainty. The null hypothesis of equal conditional predictive ability can be written as $E(d_t \mid \mathcal{F}_{t-1}) = 0$, where $\mathcal{F}_t$ is the information set up to time t. Giacomini and White (2006) show that the null of the GW test can be tested using a Wald statistic:

$$GW_t = T\left(T^{-1}\sum_{t=1}^{T} d_t\right)' \hat{\Omega}_T^{-1}\left(T^{-1}\sum_{t=1}^{T} d_t\right) \sim \chi_1^2 \tag{3.3.10}$$

where $\hat{\Omega}_T^{-1}$ is a heteroskedasticity and autocorrelation consistent (HAC) estimator of the asymptotic variance.

[16]In this chapter, the DM test is calculated based on the MSE of the different pricing methods.

To compare the forecasting performance among a set of models, we employ the methodology of MCS which can provide a subset of models that include the best model at a given confidence level. The semi-quadratic statistic is defined as[17]:

$$T_{SQ} = \sum_{i,j \in \mathcal{M}} \frac{(\bar{d}_{i,j})^2}{\widehat{var}(\bar{d}_{i,j})} \tag{3.3.11}$$

where $\bar{d}_{i,j}$ is the average loss differentials between model i and model j in set $\mathcal{M}$ and its variance $\widehat{var}(\bar{d}_{i,j})$ is obtained by using block bootstrap with 12 blocks and 10,000 replications in this chapter[18]. We consider the 75% confidence level in line with Hansen et al. (2011)[19].

3.4 Empirical results

3.4.1 Data

The VIX term structure, i.e., VIX, VIX9D, VIX3M, VIX6M and VIX1Y, and the VVIX term structure are downloaded from the CBOE website. The VIX9D data starts in January 2011; therefore, our sample period covers January 2011 to October 2020. The VIX futures data are also collected from the CBOE website. Following Zhu and Lian (2012) and Huang et al. (2019b), to avoid liquidity-related

[17]The range statistics are obtained as well with similar results which are available on request.
[18]More details about the MCS procudure can be found in Hansen et al. (2011).
[19]The 95% confidence level yields similar results which are available on request.

bias, the VIX futures prices with time to maturity less than five days, or open interest less than 200 contracts are discarded. Also, we exclude the VIX futures data which don't have a matched VVIX term structure. The VIX futures prices thus yield 14,990 observations in total.

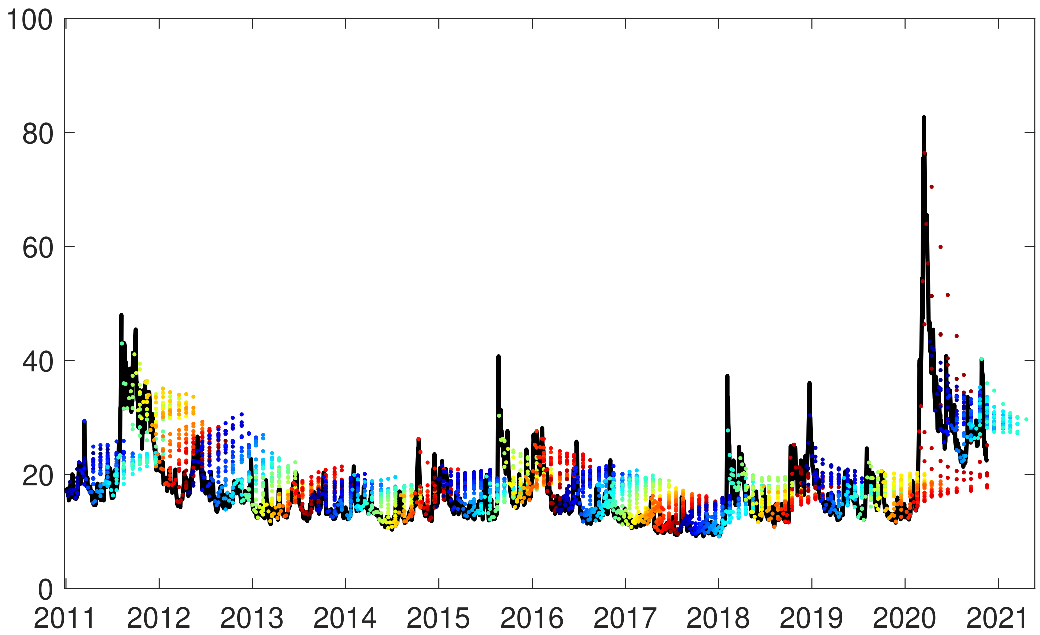
Table 3.4.1: Summary statistics for VIX futures prices

	Obs	Mean	Std.Dev	Skew.	Kurt.	Min	Max
All	14,990	19.348	5.342	1.658	4.369	10.025	70.475
Maturity							
≤ 30	3,528	17.645	5.977	2.386	8.945	10.025	70.475
$(30, 60]$	3,385	19.171	5.452	1.892	4.726	11.975	59.925
$(60, 90]$	3,411	19.770	4.953	1.579	2.696	12.975	51.500
$(90, 120]$	3,421	20.261	4.574	1.265	0.810	13.725	37.475
> 120	1,246	20.992	4.814	1.088	0.010	14.525	35.150
VIX level							
VIX ≤ 15	7,268	16.003	2.104	0.090	0.923	10.025	26.600
VIX > 15	7,722	22.497	5.549	1.434	3.517	13.875	70.475

This table presents the summary statistics for VIX futures prices from January 2011 to October 2020. The data are summarised by different maturities and VIX levels.

Summary statistics for the sample period is presented in Table 3.4.1. The price of VIX futures is on average \$19.348 with a standard deviation of \$5.234, a minimum of \$10.025 and a maximum of \$70.475. The table also shows the VIX futures term structure. When the time to maturity increases, the VIX futures prices tend to become more expensive and less volatile. It is also notable that the extreme value is highly likely in the VIX futures with short maturities. In addition, the VIX futures are cheaper when the VIX index is lower, i.e., the average price of VIX futures is 16.003 when the VIX value is smaller than or equal to 15, and the average price is 22.497 when the VIX level is higher than 15. Figure 3.4.1 shows this relationship intuitively by plotting the spot VIX value and VIX futures curves from 2001 to 2020. Furthermore, we observe that the VIX futures curve is upward sloping for most of the time, i.e., longer-term VIX futures are more expensive than near term VIX futures. However, the VIX futures curve demonstrates a downward sloping pattern when the spot VIX index spikes, for example, during the 2020 COVID-19 Crisis.

Figure 3.4.1: VIX and VIX futures curves



Notes: The black line stands for the spot VIX index value and the colorful dots stand for the VIX futures prices with different expiration dates.

Table 3.4.2: In-sample pricing errors of VIX futures models

	MAE	RMSE	MAE%	RMSE%	Std.Dev	Corr.coef.
Panel A: Jan 2011 - Dec 2019						
Fut+5VIs(e)	0.276	0.366	1.563	2.086	0.351	0.997
Fut+5VIs(t)	0.294	0.392	1.668	2.244	0.378	0.996
Fut+4VIs	0.223	0.312	1.259	1.764	0.305	0.998
Fut+VIX1Y	0.249	0.377	1.410	2.194	0.376	0.996
Fut	**0.093**	**0.143**	**0.508**	**0.754**	**0.143**	**0.999**
5VIs	0.861	1.339	4.774	7.177	1.277	0.955
4VIs	1.078	1.790	5.928	9.653	1.768	0.917
XZR	2.098	2.712	11.479	14.25	2.708	0.771
Panel B: Jan 2020 - Oct 2020						
Fut+5VIs(e)	0.965	1.658	3.127	4.578	1.581	0.982
Fut+5VIs(t)	1.166	1.976	3.717	5.500	1.831	0.975
Fut+4VIs	0.822	1.387	2.735	3.943	1.374	0.985
Fut+VIX1Y	1.580	3.283	4.681	7.753	3.012	0.929
Fut	**0.509**	**0.751**	**1.824**	**2.629**	**0.752**	**0.995**
5VIs	3.021	5.459	9.437	15.361	4.915	0.773
4VIs	2.664	5.159	8.533	14.456	4.783	0.792
XZR	7.235	9.013	24.366	29.040	8.487	0.707

This table presents the in-sample pricing errors of VIX futures models. MAE is the average absolute error between the market price and the model price; RMSE is the square root of the average squared pricing error (market price - model price); MAE% and RMSE% are expressed in relative terms (percentages); Std.Dev is the standard deviation of pricing errors; Corr.coef. is the correlation coefficient between the model price and the market price. Numbers in bold are the loss which are the lowest across different models.

3.4.2 In-sample pricing performance

To compare which method has better performance to evaluate VIX futures prices in-sample, Table 3.4.2 presents several pricing errors described in section 3.3.3[20]. We also report the standard deviation (Std.Dev) of pricing errors, i.e., the difference between market prices and model prices, and the correlation coefficients between these two series.

We divide the whole sample period into two subperiods to investigate whether VIX term structure data can help price VIX futures. Panel A of Table 3.4.2 shows the in-sample pricing errors for the period from January 2011 to December 2019. Not surprisingly, the 'Fut' method that only uses futures data has the lowest values of loss functions in fitting VIX futures series. On the contrary, the benchmark model 'XZR' approach, which only includes the VIX information and assumes a normal distribution for the return innovations, displays the highest pricing errors among all the methods. When we add more volatility indices and considering filtered historical returns, both the '4VIs' and '5VIs' provide a significant improvement in pricing errors compared with the 'XZR' method. Most importantly, after we incorporate the VIX term structure along with the futures data, i.e., for the 'Fut+5VIs (e)', 'Fut+5VIs (t)', 'Fut+4VIs' and 'Fut+VIX1Y' models, the pricing errors are slightly higher than those estimated for the 'Fut' model. However, these four models still provide a good pricing performance in

[20]The estimated parameters of different models are available upon request.

terms of correlation coefficients between market prices and model prices, with correlation values more than 0.995.

Panel B of Table 3.4.2 reports the in-sample pricing performance for the period January 2020 to October 2020. The pricing errors produced by all eight different models during 2020 are much higher than those given by the corresponding models for 2011-2019. Again, the 'Fut' model delivers the lowest loss values for VIX futures pricing and the 'XZR' model gives the highest pricing errors. Among the other approaches, 'Fut+5VIs (e)' and 'Fut+4VIs' report higher pricing errors than the 'Fut' method, but still give high correlation coefficients between market prices and model prices with values of 0.982 and 0.985, respectively.

3.4.3 Out-of-sample pricing performance

To test whether the pricing models overfit the VIX futures prices in-sample, we also perform one-day ahead out-of-sample forecasting. Table 3.4.3 presents the out-of-sample pricing errors of different models. Panel A covers the period from January 2011 to December 2019. Similar to the in-sample comparison for the same period, the 'Fut' model shows the best pricing performance out-of-sample. However, it's noted that the errors of all the pricing methods increase compared with the in-sample results, especially for the 'Fut' and 'Fut+VIX1Y' model. Thus the difference between the other models and the 'Fut' model decrease. For example, the difference of RMSE provided by 'Fut+4VIs' and 'Fut' is 0.034 out-of-sample

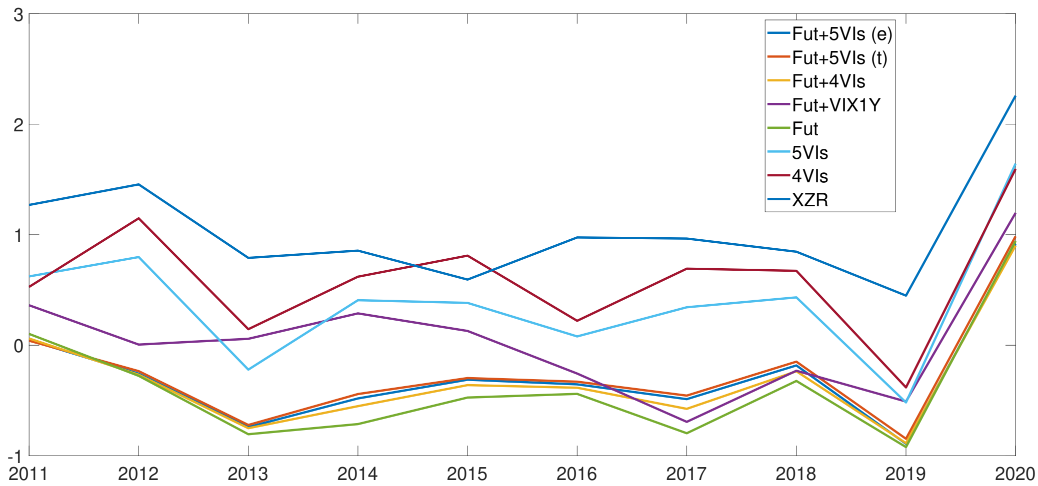
Table 3.4.3: Out-of-sample pricing errors of VIX futures models

	MAE	RMSE	MAE%	RMSE%	Std.Dev	Corr.coef.
Panel A: Jan 2011 - Dec 2019						
Fut+5VIs(e)	0.486	0.711	2.659	3.722	0.696	0.986
Fut+5VIs(t)	0.500	0.726	2.735	3.818	0.712	0.986
Fut+4VIs	0.456	0.691	2.478	3.579	0.683	0.987
Fut+VIX1Y	0.603	0.994	3.278	5.311	0.991	0.973
Fut	**0.415**	**0.657**	**2.196**	**3.262**	**0.656**	**0.988**
5VIs	0.975	1.460	5.387	7.779	1.389	0.946
4VIs	1.190	1.895	6.531	10.163	1.866	0.907
XZR	2.119	2.715	11.602	14.299	2.692	0.777
Panel B: Jan 2020 - Oct 2020						
Fut+5VIs(e)	**1.356**	2.498	**4.272**	**6.697**	**2.390**	**0.951**
Fut+5VIs(t)	1.469	2.680	4.589	7.188	2.491	0.947
Fut+4VIs	1.376	**2.464**	4.402	6.799	2.427	0.948
Fut+VIX1Y	1.803	3.311	5.668	9.543	3.289	0.907
Fut	1.443	2.567	4.548	7.128	2.553	0.942
5VIs	3.089	5.173	9.808	15.185	4.542	0.804
4VIs	2.733	4.927	8.865	14.495	4.496	0.813
XZR	7.696	9.560	25.691	30.633	8.783	0.643

This table presents the out-of-sample pricing errors of VIX futures forecasts. MAE is the average absolute error between the market price and the model price; RMSE is the square root of the average squared pricing error (market price - model price); MAE% and RMSE% are expressed in relative terms (percentages); Std.Dev is the standard deviation of pricing errors; Corr.coef. is the correlation coefficient between the model price and the market price. Numbers in bold are the loss which are the lowest across different models.

which is only 5% of the pricing errors given by the 'Fut' model.

Figure 3.4.2: Out-of-sample RMSE comparison of VIX futures models



Notes: The logarithm of the RMSE estimates of various models.

Panel B of Table 3.4.3 reports the out-of-sample pricing errors for 2020. We find that the loss values are almost tripled for all the pricing models compared with those from 2011-2019. Notably, instead of the 'Fut' model that has the smallest pricing errors for both in-sample and out-of-sample from 2011 to 2019, the 'Fut+5VIs' model offers the lowest errors when considering the MAE, MAE% and RMSE% loss functions and 'Fut+4VIs' gives the lowest RMSE. Figure 3.4.2 demonstrates the change of RMSE for different pricing methods and for by years[21]. Apparently, the pricing errors from the volatility indices related models, i.e., 'XZR', '4VIs' and '5VIs', are higher than those from the other models for almost all the years. On the other hand, when the VIX term structure is combined with

[21] For better illustration effect, we consider the logarithm of the RMSE for each method.

futures data, 'Fut+5VIs (e)', 'Fut+5VIs (t)', 'Fut+4VIs' develop similar patterns and values with the 'Fut' model.

Diebold and Mariano test

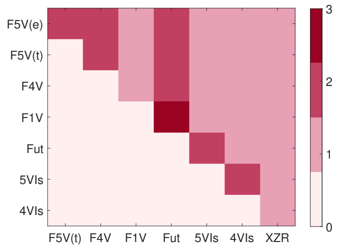
To further answer whether the VIX term structure can help to predict the VIX futures prices, we also want to know whether the addition of VIX term structure to the futures data has a statistically significant difference as compared to using the futures data only. Table 3.4.4 provides the DM test statistics by years based on 'Fut' as the benchmark model. For the year 2011, 'Fut+5VIs (e)', 'Fut+5VIs (t)' and 'Fut+4VIs' report negative DM statistics, which indicates that they generate lower MSE on average than 'Fut'; especially 'Fut+5VIs (t)' proved to outperform the 'Fut' method with a significant DM value. For the year 2020, 'Fut+5VIs (e)' and 'Fut+4VIs' display smaller MSE than 'Fut', although not significantly different. From 2012 to 2019, 'Fut' gives the lowest MSE. However, the 'Fut' model is found not significantly better than at least one of the other models for the years 2012-2013, 2016 and 2018-2019. Interestingly, 'Fut+4VIs' offers the lowest MSE for the overall period, i.e., 2011-2020. This finding suggests that adding the VIX term structure to futures data can provide lower pricing errors for VIX futures pricing.

Table 3.4.4: Out-of-sample comparison of VIX futures pricing: DM test results

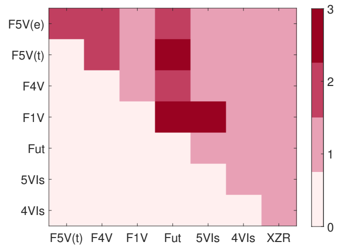
	Fut+5VIs (e)		Fut+5VIs (t)		Fut+4VIs		Fut+VIX1Y		Fut		5VIs		4VIs		XZR	
Year	MSE	DM	MSE	DM	MSE	DM	MSE	DM	MSE	DM	MSE	DM	MSE	DM	MSE	DM
2011	1.100	-1.901	**1.087**	-2.054	1.130	-1.577	2.062	3.356	1.228	-	3.473	2.263	2.875	3.976	12.665	9.938
2012	0.597	0.864	0.626	1.773	0.586	0.505	1.013	2.439	**0.575**	-	4.934	5.236	9.942	6.333	18.361	11.715
2013	0.227	2.212	0.236	2.661	0.223	1.733	1.124	4.841	**0.200**	-	0.644	4.876	1.337	4.820	4.858	17.806
2014	0.383	4.654	0.413	5.931	0.332	2.723	1.778	5.790	**0.240**	-	2.260	3.327	3.458	6.629	5.555	19.317
2015	0.535	5.499	0.553	5.976	0.486	3.938	1.295	5.076	**0.388**	-	2.151	8.858	5.060	5.907	3.301	10.950
2016	0.493	1.734	0.516	2.101	0.464	1.120	0.598	3.432	**0.415**	-	1.173	9.042	1.556	6.135	7.052	8.838
2017	0.377	8.415	0.402	8.918	0.316	5.491	0.250	5.520	**0.204**	-	1.989	10.402	3.997	7.983	6.907	14.628
2018	0.694	3.032	0.744	3.393	0.624	2.444	0.631	1.277	**0.524**	-	2.378	5.056	3.845	2.790	5.406	13.183
2019	0.168	0.678	0.184	1.835	0.170	0.844	0.362	3.295	**0.158**	-	0.355	5.439	0.466	6.486	2.462	9.618
2020	6.240	-0.374	7.182	0.781	**6.071**	-0.593	10.964	2.622	6.590	-	26.758	5.110	24.277	4.535	91.330	11.435
Overall	1.045	0.392	1.153	1.970	**1.004**	-0.081	1.926	5.550	1.010		4.447	8.393	5.537	10.438	15.258	17.015

This table presents the out-of-sample mean squared errors (MSE) of the VIX futures forecasts and the Diebold-Mariano (DM) test statistics. The DM test statistic is for the null hypothesis of equal accuracy and follows a N(0,1) distribution. The values in bold are the lowest MSE values for a given year and the DM values are based on 'Fut' as the benchmark model.

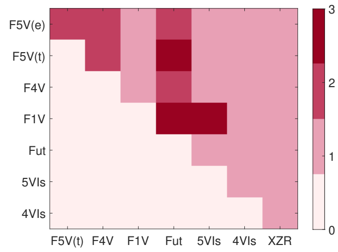
Figure 3.4.3: Out-of-sample GW test results for 2011 - 2020



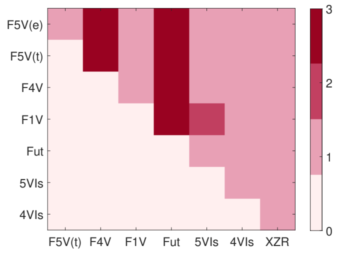
(a) Year 2011



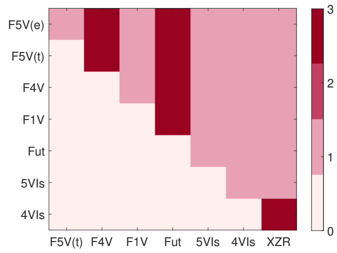
(b) Year 2012



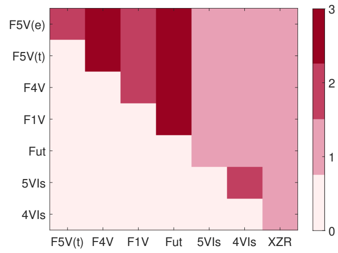
(c) Year 2013



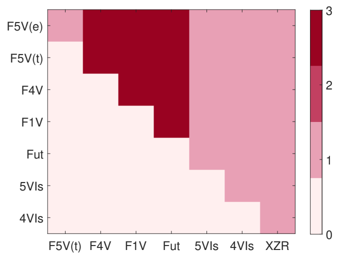
(d) Year 2014



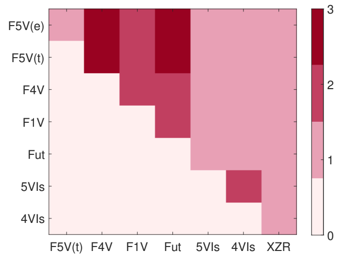
(e) Year 2015



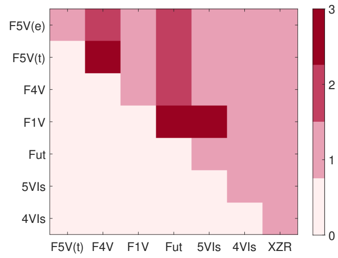
(f) Year 2016



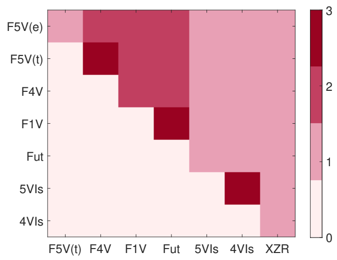
(g) Year 2017



(h) Year 2018



(i) Year 2019



(j) Year 2020

Notes: Color map based on the GW test comparing the MSE loss values. The null hypothesis is that the row model and column model have equal conditional predictive ability. Color 0 means no comparison between two models; color 1 blocks signify that the row model has higher MSE than the column model at 5% level; color 2 means that there is no difference in the conditional forecasting ability between the row and column models; color 3 signifies that the row model has lower MSE than the column model at 5% level. The models are denoted by short abbreviations in the following order: F5V(e) denotes Fut+5VIs (e), F5V(t) denotes Fut+5VIs (t), F4V denotes Fut+4VIs, F1V stands for Fut+VIX1Y and the other four models use the same notations as defined in Section 3.3.1.

Giacomini and White test

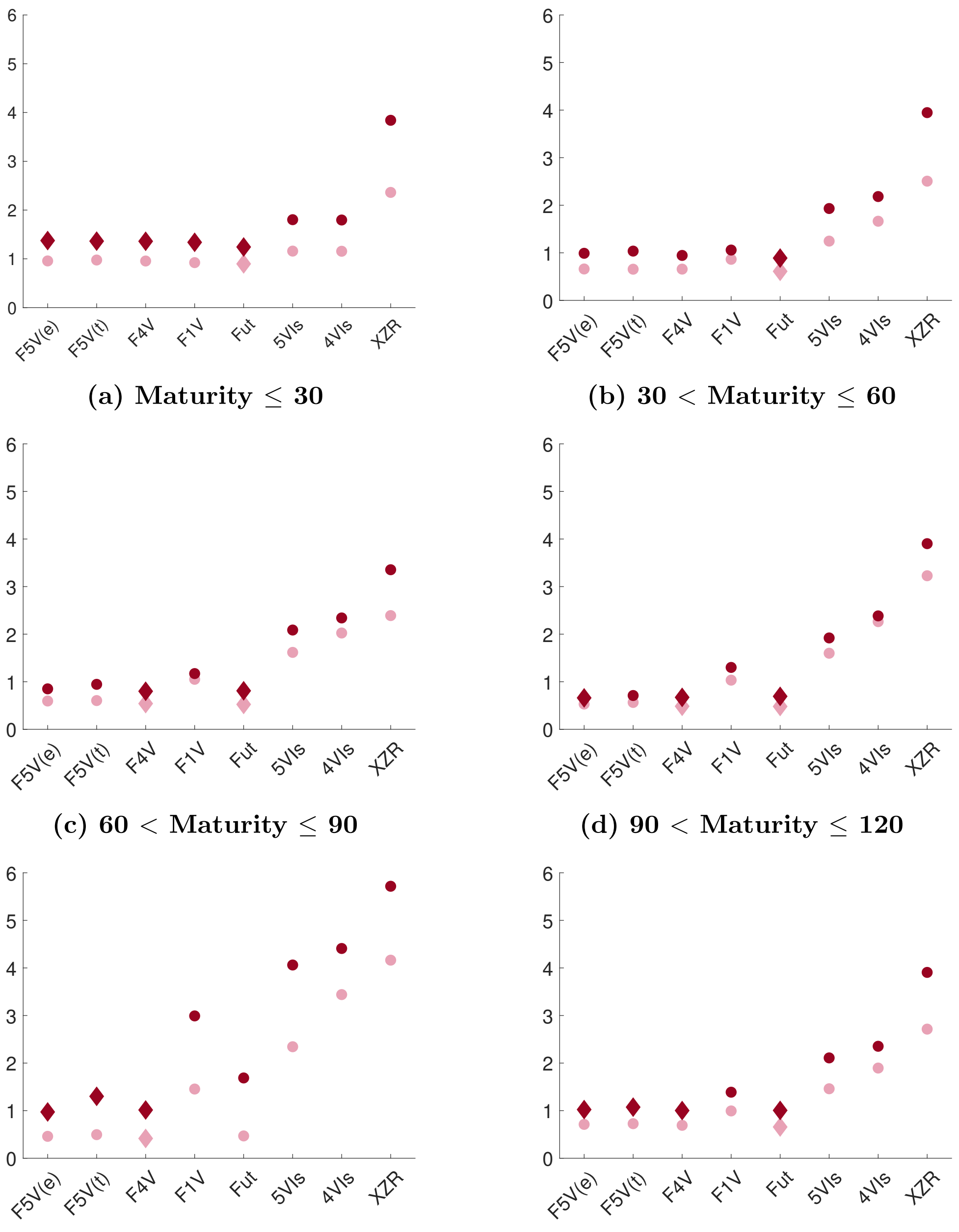
Figure 3.4.3 plots the colour map based on the GW test by years. The dark-red block (Colour 3) means that the row model has a lower MSE than the column model at 5% significance level; the pink block (Colour 1) means that the column model has a lower MSE than the row model at 5% level. For the years 2014-2017, the 'Fut' model has dark-red blocks for all the vertical comparisons vertically and pink blocks horizontally, which means that it outperforms all the other approaches during this period. However, for the rest of the sample period, 'Fut' is not significantly different from at least one method which contains both futures data and the VIX term structure information. It is also notable that for 6 out of 10 years, 'Fut' is found not to be superior to 'Fut+4VIs' in the predictive ability of VIX futures prices.

Model confidence set

The MCS test is carried out for a detailed comparison of the pricing errors based on different time to maturity and VIX levels, presented in Figure 3.4.4 and Figure 3.4.5[22]. Figure 3.4.4 reports the MCS test according to maturity for both 2011-2019 and 2011-2020. Consistent with our earlier findings, the pricing errors tend to be larger when we include the 2020 data for all different maturities. Only for the VIX futures with a maturity between 30 days and 60 days, 'Fut' is a

[22]Since we do not have the enough observations for the MCS test for 2020 year only, we then compare the pricing performance between the period 2011-2019 and 2011-2020.

Figure 3.4.4: Out-of-sample RMSE of VIX futures models for different maturities



(a) Maturity $\leq$ 30

(b) 30 $<$ Maturity $\leq$ 60

(c) 60 $<$ Maturity $\leq$ 90

(d) 90 $<$ Maturity $\leq$ 120

(e) Maturity $\geq$ 120

(f) All Maturities

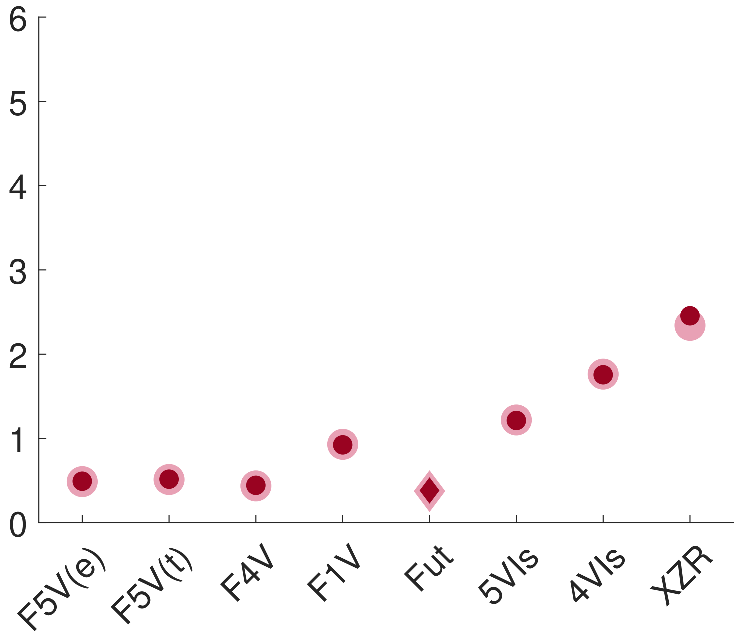
Notes: The RMSE loss values for different levels of time to maturity (days). Different colors denote a different sample time period, i.e., pink dots/diamonds indicate outputs obtained over the sample period Jan 2011- Dec 2019; dark red dots/diamonds are outputs estimated over Jan 2011- Oct 2020. The models represented by diamonds belong to the MCS at 75% confidence level. The models are denoted by short abbreviations in the following order: F5V(e) denotes Fut+5VIs (e), F5V(t) denotes Fut+5VIs (t), F4V denotes Fut+4VIs, F1V stands for Fut+VIX1Y and the other four models use the same notations as defined in Section 3.3.1.

single best-performing method regardless whether the year 2020 is added to the estimation or not. When the time to maturity ranges from 60 days to 120 days, the out-of-sample pricing performance of 'Fut+4VIs' is not significantly different from the 'Fut' model. Importantly, for the long-term VIX futures, i.e., when the maturity is longer than 120 days, 'Fut+4VIs' is proved to produce more accurate forecasts than the 'Fut' model for both periods. When considering all futures prices from 2011 to 2020, the MCS ends up with 'Fut+5VIs (e)', 'Fut+5VIs (t)', 'Fut+4VIs' and 'Fut'. This finding is consistent with our earlier results from the DM test and the GW test.

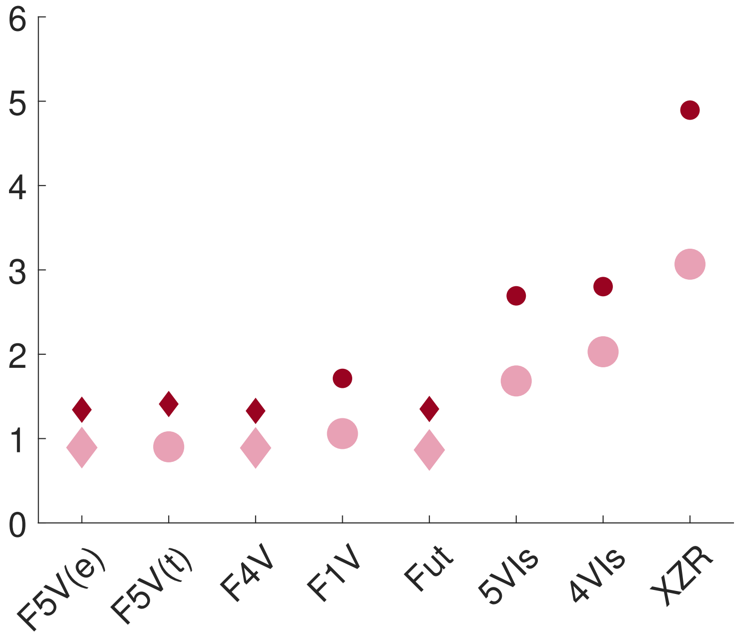
Figure 3.4.5 summarises the RMSE values by different VIX levels. All the different methods tend to have a higher RMSE when the VIX level is higher. For low VIX levels, i.e., VIX ≤ 15, only 'Fut' is included in the MCS for cases, when the COVID period is included or excluded from the dataset. But when the VIX level is higher than 15, the MCS contains 'Fut+5VIs (e)', 'Fut+4VIs' and 'Fut' for 2011-2019 and 'Fut+5VIs (e)', 'Fut+5VIs (t)', 'Fut+4VIs' and 'Fut' for 2011-2020.

All the results above provide evidence that the use of VIX term structure improves the VIX futures forecasting, especially when the VIX level is high. Moreover, Table 3.4.5 suggests that combining the volatility indices with futures data can give close VIX forecasts compared to using the VIX term structure only.

Figure 3.4.5: Out-of-sample RMSE of VIX futures models for different VIX levels



(a) VIX $\leq$ 15



(b) VIX $>$ 15

Notes: The RMSE loss values estimated over different VIX levels. Different colors denote a different sample time period, i.e., pink dots/diamonds indicate outputs for the sample period Jan 2011- Dec 2019; dark red dots/diamonds are outputs estimated over Jan 2011- Oct 2020. The models represented by diamonds belong to the MCS at 75% confidence level. The models are denoted by short abbreviations in the following order: F5V(e) denotes Fut+5VIs (e), F5V(t) denotes Fut+5VIs (t), F4V denotes Fut+4VIs, F1V stands for Fut+VIX1Y and the other four models use the same notations as defined in Section 3.3.1.

Table 3.4.5: Out-of-sample pricing errors of VIX models

	MAE	RMSE	MAE%	RMSE%	Std.Dev	Corr.coef.
Fut+5VIs(e)	1.194	2.354	6.097	9.040	2.352	0.952
Fut+5VIs(t)	1.221	2.378	6.267	9.211	2.374	0.951
Fut+4VIs	1.189	2.375	6.066	9.047	2.373	0.951
Fut+VIX1Y	2.568	4.791	14.938	29.089	4.683	0.801
Fut	1.502	2.790	7.660	10.531	2.790	0.934
5VIs	1.081	2.009	5.659	7.778	2.002	0.965
4VIs	1.043	1.983	5.425	7.622	1.978	0.966
XZR	2.542	4.536	13.139	16.831	4.536	0.817

This table presents the out-of-sample pricing errors of CBOE VIX forecasts from January 2011 to October 2020. MAE is the average absolute error between the market price and the model price; RMSE is the square root of the average squared pricing error (market price - model price); MAE% and RMSE% are expressed in relative terms (percentages); Std.Dev is the standard deviation of pricing errors; Corr.coef. is the correlation coefficient between the model price and the market price.

3.5 Conclusions

In this chapter, we examine the efficiency of including the VIX term structure in the VIX futures pricing model. The GJR model parameters are calibrated from the data using filtered historical simulation. We also include the VVIX term structure in the VIX futures pricing model. We find that the out-of-sample performance of the models that use the VIX term structure and the VIX futures is not significantly different from the model that use the futures data only for most of the years, but yields significant outperformance compared to the models based on the VIX/VIX term structure. Most importantly, the evidence of the 2020 COVID-19 crisis suggests that the addition of the VIX term structure improves model performance in terms of achieving lower pricing errors. Meanwhile, the MCS test shows that the use of the VIX term structure can also deliver better forecasts for the VIX futures with a maturity longer than 120 days or when the VIX level is higher than 15.

This chapter has at least two implications. First, the model that uses futures data only underperforms when it comes to forecasting futures prices, especially for the extreme observations, i.e., during the 2020 COVID-19 crisis. Second, the GJR model with filtered historical simulation is a better choice for describing market volatility compared to the tradition local risk-neutral valuation relationship.

Appendices

This supplemental appendix provides additional tables and figures.

Table A.1: Parameter estimates obtained using different methods for 2017

	$\omega \times 10^5$		β		$\alpha \times 10^2$		γ		Ann. vol.	
Model	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev	Mean	Std.dev
Panel A. P measure										
GJR	2.267	0.110	0.886	0.002	0.000	0.000	0.176	0.002	0.089	0.015
Panel B. Q measure										
Fut+5VIs(e)	3.027	9.109	0.855	0.120	4.841	6.142	0.128	0.109	0.114	0.104
Fut+5VIs(t)	2.883	9.104	0.839	0.117	6.265	6.905	0.133	0.110	0.116	0.096
Fut+4VIs	3.513	9.409	0.926	0.132	2.077	3.745	0.036	0.082	0.105	0.100
Fut+VIX1Y	0.496	0.262	0.889	0.049	7.083	5.904	0.064	0.086	0.117	0.011
Fut	1.133	0.499	0.948	0.050	2.430	4.159	0.035	0.062	0.107	0.014
5VIs	3.103	9.463	0.801	0.151	1.072	1.090	0.113	0.168	0.107	0.079
4VIs	3.286	8.340	0.760	0.213	1.560	1.597	0.083	0.226	0.107	0.100

This table presents the parameter estimates obtained using different models for the year 2017. Panel A reports the parameters under the physical measure. Panel B presents the parameter estimates obtained under the risk-neutral measure. Ann. vol. is the annualized volatility.

Chapter 4

Optimal forecasting of VVIX: forecast combinations *vs.* LASSO

4.1 Introduction

Volatility plays a significant role in financial markets and risk management practices. The Chicago Board Options Exchange (CBOE) volatility index (VIX) has been the most popular barometer of market sentiment since it was launched in 1993. The high VIX level implies ascendancy of fear while a low VIX level signals dominance of greed. To help investors have an even deeper insight into volatility, the CBOE introduced VVIX to measures the uncertainty in market sentiment in 2012. The goal of the VVIX index is to capture the expected volatility of the 30-day forward VIX index. Hence it is often called “the volatility-of-volatility” or "the vol-of-vol". Just as the VIX is calculated from the S&P 500 options, the VVIX is calculated using the same methodology to a cross-section of the VIX options. The reliable forecasts of the VVIX index are key to capture the future changes of the VIX index, thus is of interest to academic researchers and market participants.

The literature on the VVIX index has primarily focused on the characteristic of the volatility-of-volatility, as measured by VVIX. For example, Park (2015)employs the VVIX as a proxy to document a positive correlation between the volatility-of-volatility and the current prices of tail risk hedging options. Hollstein and Prokopczuk (2018) suggest that the volatility-of-volatility is significantly priced in the market returns and implies a negative risk premium. Huang et al. (2019a) demonstrate that time-varying volatility-of-volatility is a significant risk factor that affects VIX option returns. Bu et al. (2019) document that stocks with higher sensitivities to daily changes in volatility-of-volatility have higher returns than those with lower sensitivities. Jeon et al. (2020) show that incorporating VVIX into models significantly increases the predictive power compared to traditional volatility models. Other related studies on VVIX include Zang et al. (2017), Krause (2019) etc.

However, only few studies (if any) discuss forecasting the VVIX index. To fill this gap, we compare several popular volatility forecasting models out-of-sample and attempt to answer the following simple question: is there an optimal forecasting method for the VVIX index? In the context of volatility forecasting, a number of studies suggest that the combination of individual forecasts has often been found to outperform individual forecasts, see, for example, Becker and Clements (2008), Patton and Sheppard (2009), Wang et al. (2016) etc. There are three potential explanations: 1) the combination of individual forecasting models covers the information from each model; 2) they are likely to provide insurance against structural breaks while the individual model may be very differently affected; 3) there is a possible variance reduction since individual forecasting models may be differentially mis-specified[1]. On the other hand, the least absolute shrinkage and selection operator (LASSO), introduced by Tibshirani (1996), is another popular predictive tool in financial forecasting, see, for example, Audrino and Knaus (2016), Zhang et al. (2019a), Zhang et al. (2019b), etc. The advantages of using the LASSO include: 1) it can select the most important predictors by producing zero estimated coefficients; 2) the LASSO can pick one predictor among several highly correlated ones.

In this chapter, we generate daily VVIX forecasts for the year 2016-2020 using thirteen different models. First, we consider three individual models: a simple linear regression using a set of lagged variables; the autoregressive moving average (ARMA) model; and the heterogeneous autoregressive (HAR) model of Corsi (2009). Second, we implement eight popular combining methods in the

[1] Please see Clemen (1989), Clements and Hendry (2004) and Timmermann (2006) for reviews of forecast combinations.

literature on forecast combination approaches) to produce the VVIX forecasts (see Rapach et al., 2010; Hsiao and Wan, 2014). Third, we perform two LASSO type regressions, i.e., the original LASSO proposed by Tibshirani (1996) and the elastic net of Zou and Hastie (2005), using all the predictors from the individual models.

To the best of our knowledge, this study is among the first to attempt VVIX forecasting in the literature. We find that a median combining method outperforms all the other models by providing the lowest squared errors of the forecasts for the period covering 2016-2020. Furthermore, the model confidence set (MCS) test shows that both the simple average combining method and the median method have significantly superior forecasting performance than all the other models. In addition, our results on LASSO-type models suggest that the daily changes in average monthly VVIX play an important role for VVIX forecasting.

The remainder of the study is organized as follows. Section 4.2 introduces the daily behavior of the VVIX index. Section 4.3 describes the different forecasting models and the evaluation criteria. Section 4.4 presents our results and analysis, and Section 4.5 concludes.

4.2 The CBOE VVIX

4.2.1 The background of the VVIX index

The CBOE introduced the futures and options on the VIX in March 2004 and February 2006, respectively, to trade and hedge against changes in volatility. Nowadays, VIX options and futures are among the most actively traded contracts in the financial market. Trading on VIX derivatives enables practitioners to invest in market volatility regardless of the actual direction of the S&P 500 index, and further provides more opportunities to diversify their portfolios. Huang et al. (2019a) report that apart from the VIX index, the VVIX is also a significant risk factor that affects VIX option returns. Different from the VIX which measures the implied volatility of the S&P 500 market, the VVIX, on the other hand, shows how rapidly market volatility changes rather than measures the volatility itself. Hence, investors should consider the levels of both the VIX and the VVIX index when trading VIX options and futures. For example, if both the VIX and the VVIX are observed to have a high level, then they are expected to decrease to their long-run mean in the future. And thus investors may profit from a bear call spread which consists of a long call option with a higher strike price and a short call with a lower strike price. Moreover, Park (2015) demonstrates that the VVIX index has predictive power for the returns of tail risk hedging options, such as the S&P 500 puts and VIX calls.

The model-free formula that CBOE employs to calculate VVIX in a similar manner to the calculation of the VIX:

$$VVIX_t = 100 \times \sqrt{\frac{2}{\tau}\sum_i \frac{\triangle K_i}{K_i^2} e^{r\tau} Q(K_i) - \frac{1}{\tau}[\frac{F}{K_0} - 1]^2}, \tag{4.2.1}$$

where τ is time to expiration, $Q(K_i)$ is the VIX option price with strike K_i, $\triangle K_i$ is the interval between strike prices, F is the forward price derived from the VIX options, and K_0 is the first strike that is below the forward index level F.

4.2.2 Daily behavior of VVIX

In this chapter, we investigate the daily VVIX index for the period covering January 2007 to December 2020, which gives a total sample of 3,522 observations. Table 4.2.1 shows the summary statistics for the whole sample period. The VVIX exhibits a long term mean of 91.3 and ranges from 59.74 to 207.59. Not surprisingly, the VVIX index is right-skewed and leptokurtic. The p-value of the Jarque-Bera test in the column also demonstrates the nonnormality of the VVIX index. All these descriptive statistics show variations over the sample period. For example, the means and the ranges of the VVIX index increase as the years go by. Also, the VVIX is more volatile during 2020 than 2007-2015 and 2016-2019 which is evidenced by a high standard deviation of 19.39. Since the VVIX measures the expected volatility of the VIX index, Table 4.2.1 also reports the VVIX statistics summarised by different levels of VIX. The VVIX index has a higher value on average and is more volatile when the corresponding VIX level is high.

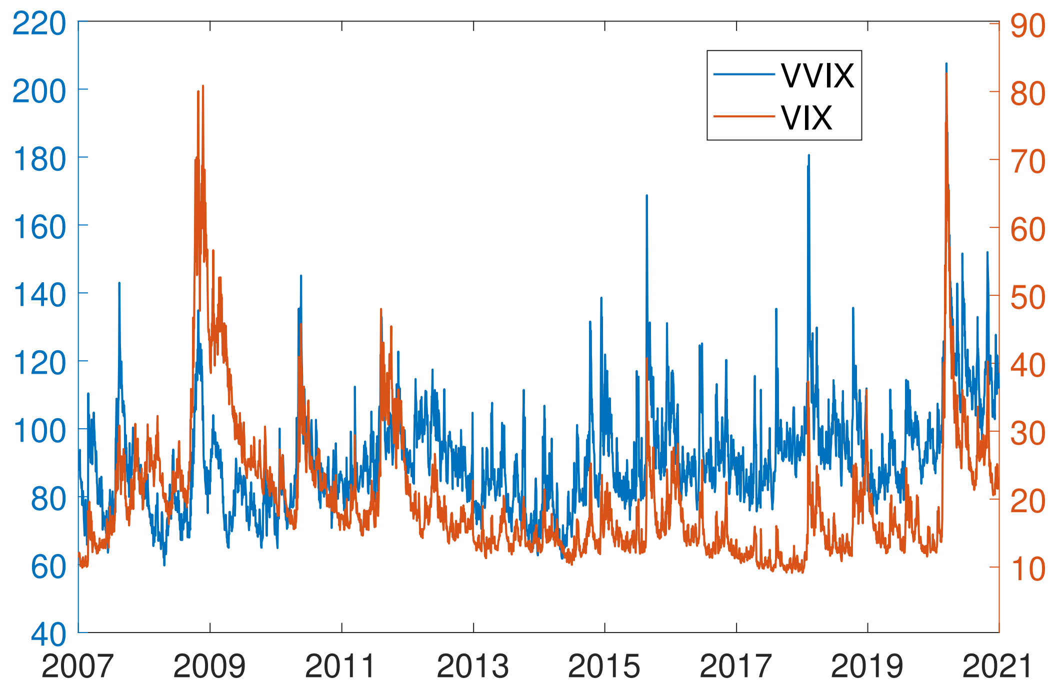
Table 4.2.1: Summary statistics on VVIX

	Obs	Mean	Std.Dev	Skew.	Kurt.	Min	Max	JB test
All	3,522	91.300	15.752	1.430	4.571	59.74	207.59	0.000
Periods								
2007-2015	2,265	87.079	13.485	0.955	1.600	59.74	168.75	0.000
2016-2019	1,005	94.026	11.607	2.020	9.105	74.98	180.61	0.000
2020	252	118.362	19.390	1.534	3.511	86.87	207.59	0.000
VIX level								
VIX$\leq$ 15	1,284	85.000	9.249	-0.047	-0.285	61.76	114.39	0.081
15<VIX$\leq$ 20	958	91.262	10.834	0.324	0.145	63.06	135.32	0.001
20<VIX$\leq$ 35	1,037	95.502	17.970	0.463	0.040	59.74	180.61	0.000
VIX>35	243	106.807	27.968	0.943	0.692	64.95	207.59	0.000

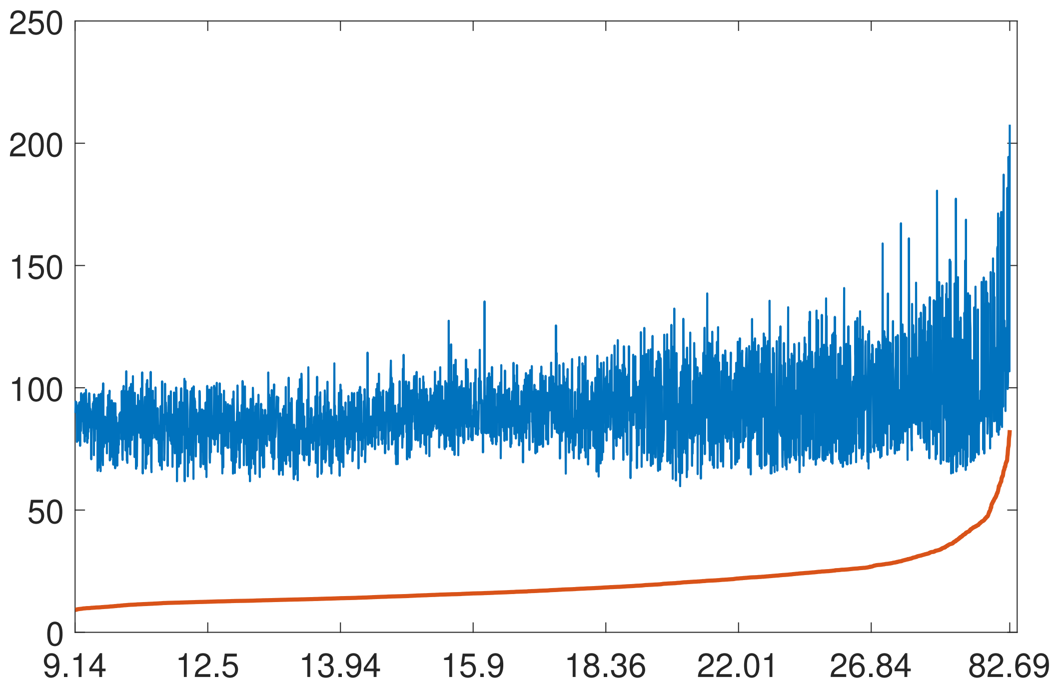
This table presents the summary statistics for the VVIX index from January 2007 to December 2020. The data are summarised by different periods and VIX levels. The p-value of Jarque–Bera (JB) test for normality is reported.

Panel A of Figure 4.2.1 plots the time evolution of both the VVIX and the VIX index over the whole sample period. The level of the volatility of volatility, i.e., the VVIX, is much higher than that of the volatility index itself, i.e., the VIX. But both the VIX and the VVIX demonstrate a mean-reverting property on a long-run basis. Interestingly, the spikes in the VVIX often tend to be accompanied by spikes in the VIX. However, it is notable that the VVIX index is more sensitive (higher spikes) to the economic uncertainty than the VIX, for example, during August 2007 (the beginning of the subprime crisis), August 2015 (the Chinese

Figure 4.2.1: The VIX and the VVIX



(a) The VIX and the VVIX

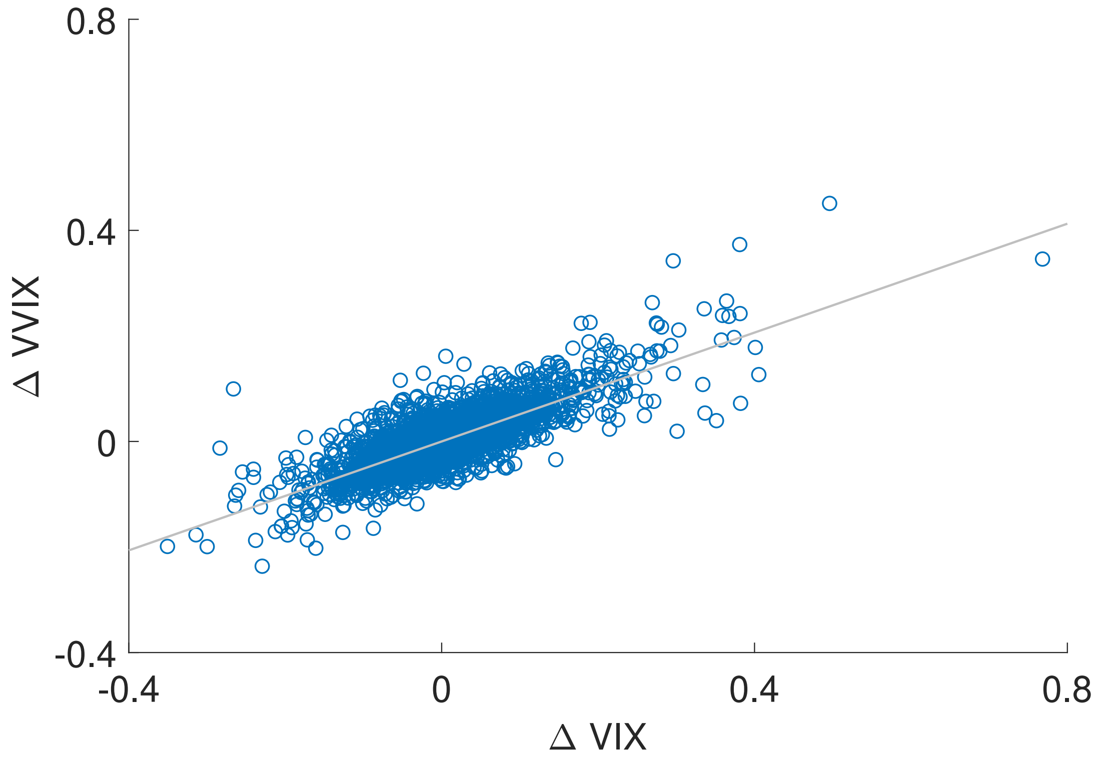


(b) VVIX and VIX sorted in ascending order of VIX

Notes: Panel (a) shows the evolution of the daily VIX and VVIX index series from January 3, 2007 to December 31, 2020. Panel (b) demonstrates the variation of the VVIX index when the VIX is sorted in ascending order. The y-axis ticker of Panel (b) is the value of daily VVIX and the x-axis shows values of the VIX index.

stock market crash), February 2019 (Donald Trump's trade war with China), etc.

Figure 4.2.2: Scatter plot of Δ VIX and Δ VVIX



Notes: Scatter plot of the changes in the VVIX and the changes in the VIX from January 3, 2007 to December 31, 2020. For a better illustration, we take the logarithm of VVIX and VIX, respectively. The grey line is the fitted values from the estimated linear regression.

Panel B of Figure 4.2.1 illustrates the variation of the VVIX when the VIX is sorted from smallest to largest. It shows little correlation pattern between the VVIX and the VIX. However, both the magnitude and the variation of the VVIX are much larger for higher values of the VIX compared to those for a lower level of the VIX. It is noteworthy that, when we consider the logarithm of the VVIX and the VIX values, their first differential exhibits a strong correlation as shown in Figure 4.2.2.

4.3 The forecasting models

4.3.1 Linear regression model

We perform a linear regression of the daily changes in the VVIX index on a set of lagged variables. Intuitively, the lagged daily changes in VVIX are included along with its underlying, i.e., the VIX index and the S&P 500 index returns in the regression. Additionally, we examine whether the information on the VIX term structure explains future values of the VVIX index. To assess whether there is an asymmetric relationship between the daily changes in VVIX and the explanatory variables, we also consider the absolute values of these variables[2]. A mixed selection procedure, i.e., a combination of forward selection and backward selection, is then employed to decide the number of the predictors. Based on both the Akaike information criterion (AIC) and the Bayesian information criterion (BIC), the following regression is estimated:

$$\begin{aligned} \Delta VVIX_t = \beta_0 + \beta_1 \Delta VVIX_{t-1} + \beta_2 |\Delta VVIX_{t-1}| + \beta_3 \Delta VIX_{t-1} \\ + \beta_4 |\Delta VIX_{t-1}| + \beta_5 |VX^{diff}_{t-1}| + \beta_6 ret_{t-1} + \varepsilon_t, \end{aligned} \tag{4.3.1}$$

[2]Konstantinidi and Skiadopoulos (2011b) suggest that the slope of yield curve has predicative power for the VIX futures market, hence we also take this variable into consideration. We examine the yield curve slope within different maturities; however, the estimated regression shows that the information on the yield curve does not explain the daily changes in the VVIX index.

where β_0 is a constant, $\Delta VVIX_t$ and ΔVIX_t denote the daily changes in the VVIX index and the VIX index from time $t-1$ to t, respectively. $|\Delta VVIX_{t-1}|$ and $|\Delta VIX_{t-1}|$ are the absolute values of $\Delta VVIX_t$ and ΔVIX_t, respectively. $|VX_t^{diff}|$ is the absolute value of the difference between the CBOE S&P 500 3 Month Volatility Index (VIX3M) and the VIX index. ret_t is the log return of S&P 500 at time t.

4.3.2 ARMA model

The Augmented Dickey-Fuller (ADF) test proposed by Dickey and Fuller (1981) suggests that the VVIX index is non-stationary in the levels but stationary in first differences. Therefore, by considering both AIC and BIC, we estimate the following ARMA(1,1) model:

$$\Delta VVIX_t = c + \varphi_1 \Delta VVIX_{t-1} + \theta_2 \varepsilon_{t-1} + \varepsilon_t. \tag{4.3.2}$$

4.3.3 HAR model

The third model we amploy is based on the HAR model proposed by Corsi (2009). Since its origination, the HAR model is attracting increased attention in volatility modelling because it can approximate long memory processes in a simple way, is so-called 'parsimonious', and is easy to estimate using ordinary least squares (OLS). There are various extensions of the HAR model, for example, Andersen et al. (2007) and Corsi et al. (2010) introduce a jump component, Patton and Sheppard (2015) incorporate a leverage effect, etc. In this chapter, we focus on the basic HAR model to investigates the extent to which the VVIX information of the previous day, week, and month can contribute to explaining the current VVIX value. To align with the two models described in the above sections, we employ the changes in VVIX instead of the VVIX value itself[3]. The model is expressed as:

$$
\begin{aligned}
\Delta VVIX_t &= \beta_0 + \beta_1 \Delta VVIX_{t-1} + \beta_2 \Delta VVIX^w_{t-1} + \beta_3 \Delta VVIX^m_{t-1} + \varepsilon_t \\
VVIX^w_{t-1} &= \frac{1}{5}\sum_{i=1}^{5} \Delta VVIX_{t-i} \\
VVIX^m_{t-1} &= \frac{1}{22}\sum_{i=1}^{22} \Delta VVIX_{t-i},
\end{aligned}
\tag{4.3.3}
$$

where $\Delta VVIX_t$ is the daily change in the VVIX index at time t, and $\Delta VVIX^w_t$ and $\Delta VVIX^m_t$ denote the changes in average weekly and monthly VVIX levels at t, respectively [4].

[3]We compare the two regressions, i.e, the regression using VVIX and the regression using the changes. Interestingly, all the coefficients in the regression using changes are significantly different from zero at 5%, while only the coefficient of lagged daily VVIX is significant in the regression using the VVIX index.

[4]In this study, we use the trading day count convention. Hence the weekly and monthly VVIX levels are calculated as the average values over the past 5 and 22 days, respectively.

4.3.4 Forecast combinations

So far, we have three models to forecast the VVIX index. However, it has been well-known that combinations of individual forecasts often outperform the individual forecasts, see, for example, Becker and Clements (2008), Patton and Sheppard (2009), Wang et al. (2016), etc. The popular forecast combination methods include the weighted average approach and the regression approach.

Weighted average approach

The combination forecasts of VVIX using the weighted average method can be written as:

$$\widehat{VVIX}_{t+1} = \sum_{i=1}^{n} \omega_{i,t} \widehat{VVIX}_{i,t+1}, \tag{4.3.4}$$

where $\widehat{VVIX}_t$ represents the combination forecast of VVIX at time t, n is the number of individual models and in our case $n = 3$, $\omega_{i,t}$ is the combining weight for the i-th individual forecast estimated at time t, and $\widehat{VVIX}_{i,t}$ is the forecast of VVIX value which is produced by the i-th model.

Following Rapach et al. (2010), Zhu and Zhu (2013), and Hsiao and Wan (2014), among others, we consider the following four popular weighting methods:

- Mean combination. The mean combination refers to the simple average of all forecasting models, i.e., $\omega_{i,t} = \frac{1}{3}$.

- Median combination. The median combination employs the median of all individual forecasts.

- Discount mean square prediction error (DMSPE) combining method. The DMSPE method allocates greater weights to the forecasting models which have better forecasting performance. The combining weight of the i-th individual predictive model is defined as:

$$\omega_{i,t} = \frac{\phi_{i,t}^{-1}}{\sum_{j=1}^{n} \phi_{j,t}^{-1}}, \tag{4.3.5}$$

where

$$\phi_{i,t} = \sum_{s=m+1}^{t} \theta^{t-s}(VVIX_s - \widehat{VVIX}_{i,s})^2, \tag{4.3.6}$$

m is the number of observations in-sample, $VVIX_s$ is the market price of VVIX at time s, $\widehat{VVIX}_{i,s}$ is the VVIX forecast from the i-th model at time s, θ is a discount factor for which we consider two values, 1 and 0.9.

- Bayesian averaging method. In this method, the combining weight is based on the BIC value of the in-sample period:

$$\omega_{i,t} = \frac{\exp\left(-\frac{1}{2}\Delta BIC_i\right)}{\sum_{j=1}^{3} \exp\left(-\frac{1}{2}\Delta BIC_j\right)}, \tag{4.3.7}$$

where BIC_i is the BIC value for the i-th model and $\Delta BIC_i = BIC_i - min_j(BIC_j)$.

Regression approach

Granger and Ramanathan (1984) propose three regression approaches to combine forecasts:

$$
\begin{aligned}
VVIX_t &= \sum_{i=1}^{n} w_i VF_{i,t} + u_t \quad s.t. \sum_{j=1}^{3} w_j = 1 \\
VVIX_t &= \sum_{i=1}^{n} w_i VF_{i,t} + u_t \\
VVIX_t &= w_0 + \sum_{i=1}^{n} w_i VF_{i,t} + u_t,
\end{aligned}
\tag{4.3.8}
$$

where w_i is the unknown parameter for the i-th model, $VF_{i,t}$ is the VVIX forecasts for time t using the i-th individual model. In this chapter, the three regression models are referred as REG1, REG2 and REG3, respectively. REG1 and REG2 can be considered as constrained regression models of REG3.

4.3.5 LASSO regressions

The least absolute shrinkage and selection operator (LASSO) regression has been introduced by Tibshirani (1996). It is a regression analysis that penalizes the coefficients of the independent variables to shrink some of them to zero. The goal of LASSO regression is to identify the most important variables associated with the response variable. The LASSO forecast is defined as:

$$
\widehat{VVIX}_{t+1} = \hat{\beta}_0 + \sum_{i=1}^{P} \hat{\beta}_i x_{i,t}, \tag{4.3.9}
$$

and the LASSO estimate is given by:

$$\begin{aligned} \hat{\beta} = & \arg\min_{\beta} \frac{1}{2} \sum_{j=1}^{t-1} \left(VVIX_{j+1} - \beta_0 - \sum_{i=1}^{P} \beta_i x_{i,j} \right)^2, \\ s.t. \quad & \sum_{i=1}^{P} |\beta_i| \leq \psi, \end{aligned} \tag{4.3.10}$$

where $x_{i,t}$ is the i-th predictor available at time t, P is the total number of the predictors which are all the predictors included in the individual models. $\sum_{i=1}^{P} |\beta_i|$ denotes the L_1 LASSO penalty which makes the solutions nonlinear in the VVIX forecasts. ψ is a pre-specified parameter that determines the degree of shrinkage. When ψ is sufficiently small, some of the coefficients may become zero which leads to the selection of a subset of the variables.

We can also write the LASSO in the so-called Lagrangian form:

$$\hat{\beta} = \arg\min_{\beta} \left\{ \sum_{j=1}^{t-1} \left(\widehat{VVIX}_{j+1} - \beta_0 - \sum_{i=1}^{P} \beta_i x_{i,j} \right)^2 + \lambda \sum_{i=1}^{P} |\beta_i| \right\}, \tag{4.3.11}$$

where λ controls the amount of L_1 regularization and serves a similar role as ψ.

Apart from the original LASSO method, we also consider another popular LASSO approach which is the elastic net proposed by Zou and Hastie (2005). The elastic net forecast is the same as Equation (4.3.9), whereas the corresponding estimate is:

$$\hat{\beta} = \arg\min_{\beta} \left\{ \sum_{j=1}^{t-1} \left(\widehat{VVIX}_{j+1} - \beta_0 - \sum_{i=1}^{P} \beta_i x_{i,j} \right)^2 + \lambda \sum_{i=1}^{P} \left(\alpha \beta_i^2 + (1-\alpha) |\beta_i| \right) \right\}, \tag{4.3.12}$$

where $\alpha \in [0, 1]$, and when $\alpha = 0$ the elastic net method turns into the original LASSO.

In order to forecast the VVIX index for time $t+1$ using the LASSO and elastic net regressions, we need to decide the optimal value of λ and α *ex ante* with all the information up to t. In this chapter, we employ the split cross-validation to identify the optimal λ and α. The detailed procedure is as follows:

1. To avoid unfair penalty on the predictors with a small range, we standardise the features before fitting the model, i.e., we subtract the mean of the feature and then divide it by the standard deviation of the feature[5].

2. Determine a minimum number of observations for fitting the model, denoted as m. In this chapter, we use the number of observations during 2007-2014 as the minimum number. At the first iteration, we train the model on the data $VVIX_1, VVIX_2, ...VVIX_m$ and forecast the price for the next day, $\widehat{VVIX}_{m+1}$. Then the forecast error is $e_{m+1} = VVIX_{m+1} - \widehat{VVIX}_{m+1}$.

3. For the second iteration, fit the model to the data $VVIX_1, VVIX_2, ...VVIX_{m+1}$

[5]We also perform the max-min normalisation to scale the features; the results are similar and available on request.

and calculate the forecast error e_{m+2}, and so on, up to time t.

4. Compute the MSE as $\frac{1}{t-m}\sum_{i=m+1}^{t} e_i^2$. Identify the optimal λ and/or α which yield the lowest MSE.

5. Fit the model to $VVIX_1, VVIX_2, ...VVIX_t$ with the optimal estimates of λ and/or α from the last step. Then the model parameters $\hat{\beta}$ in the equation (4.3.11) and (4.3.12) can be obtained.

6. Once the model parameters $\hat{\beta}$ and shrinkage factors λ and/or α are known, we can forecast $VVIX_{t+1}$ with all the predictors up to time t.

4.3.6 Model evaluation

To quantitatively evaluate the forecasting accuracy of different models, we follow the literature and use the three popular loss functions[6]:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|VVIX_i - \widehat{VVIX}_i\right|, \tag{4.3.13}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(VVIX_i - \widehat{VVIX}_i\right)^2}, \tag{4.3.14}$$

$$QLIKE = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{VVIX_i}{\widehat{VVIX}_i} - \log\frac{VVIX_i}{\widehat{VVIX}_i} - 1\right), \tag{4.3.15}$$

[6]See Patton (2011a) for a range of loss functions which are employed in the literature of volatility forecast evaluation.

where N is the number of total observations out-of-sample, $VVIX_i$ is the market value of the VVIX index at time i, and $\widehat{VVIX}_i$ denotes the VVIX forecast of a given model. Since the difference between the highest VVIX value and the lowest one is more than 100, we also report the MAE and RMSE in relative terms, i.e., MAE% and RMSE%, respectively.

$$MAE\% = \frac{1}{N}\sum_{j=1}^{N}\left|\frac{\widehat{VVIX}_i}{VVIX_i} - 1\right|, \tag{4.3.16}$$

$$RMSE\% = \sqrt{\frac{1}{N}\sum_{j=1}^{N}\left[\left(\frac{\widehat{VVIX}_i}{VVIX_i} - 1\right)^2\right]}. \tag{4.3.17}$$

Also, to evaluate the statistical forecast accuracy of the models, we consider both the Diebold and Mariano (1995) (DM) test and the model confidence set (MCS) test of Hansen et al. (2011). The DM test is employed to examine the significance of the differences between two series of forecasts. Specifically, the difference in errors is defined as: $d_t = L(\hat{e}_{1,t}) - L(\hat{e}_{2,t})$, then the DM statistic is given by:

$$DM = \frac{\bar{d}}{\sqrt{2\pi\hat{f}_d(0)/T}}, \tag{4.3.18}$$

where $\bar{d}$ is the mean of d_t, and $2\pi\hat{f}_d(0)$ is a consistent estimator of the asymptotic variance.

The MCS is the subset of models which contains the best models at a given confidence level. In this chapter, we consider the 75% confidence level and two methods, which calculate the test statistics using the sums of absolute values (R method) and the sums of squared loss differentials (SQ method), respectively.

4.4 Empirical results

4.4.1 Data

The VVIX data is downloaded from the CBOE website. With respect to the predictors for the individual models of linear regression, LASSO and Elastic net models, the VIX, the VIX3M and the S&P 500 prices are from www.finance.yahoo.com; the data on the yield curve is from U.S. Department of the Treasury website.

During the process of model estimation, some of the forecasting methods need a holdout period to estimate the parameters, for example, the weights in the weighted average combing method (except the Mean and Median combinations), the parameters in the model combining regression, λ and/or α in LASSO and Elastic net models. We then consider the 2007-2014 period as the in-sample period and the year 2015 as the first holdout period. Therefore, for the out-of-sample evaluation, we examine the pricing performance over 2016-2020 for all the models on a rolling-window basis. In addition, we divide the whole out-of-sample period into two sub-periods: 1) a relatively peaceful period covering 2016-2019 which has an overall standard deviation of 11.607; 2) year 2020 which is more volatile (standard deviation of 19.390) due to the uncertainty related to the COVID-19 crisis.

4.4.2 Out-of-sample forecast evaluation

Table 4.4.2: Out-of-sample forecasting errors (1)

Model	MAE	RMSE	QLIKE	MAE%	RMSE%	MCS
Random walk	3.579	5.482	1.263	3.510	5.159	-
Regression	3.551	5.488	1.258	3.477	5.160	-
ARMA	3.486	5.423	1.240	3.441	5.141	-
HAR	3.531	5.453	1.250	3.472	5.144	-
Mean	3.494	5.416	1.235	3.438	5.118	✓
Median	3.456	5.409	1.232	3.413	5.130	✓
DMSPE1	3.494	5.417	1.235	3.438	5.119	-
DMSPE0.9	3.495	5.418	1.235	3.438	5.119	-
Bayesian	3.531	5.453	1.250	3.472	5.143	-
REG1	3.483	5.459	1.242	3.429	5.147	-
REG2	3.565	5.531	1.271	3.504	5.197	-
REG3	3.570	5.571	1.277	3.503	5.207	-
LASSO	3.518	5.432	1.239	3.457	5.120	-
Elastic net	3.516	5.433	1.241	3.455	5.124	-

This table presents the out-of-sample pricing errors for the full sample period covering 2016-2020. MAE is the average absolute error between the market price and the model price; RMSE is the square root of the average squared pricing error; QLIKE is given in equation (4.3.15). MAE% and RMSE% are in relative terms expressed in percentages with respect to the VVIX. The last column presents an indicator whether the model is in the MCS at 75% confidence level based on MSE.

In Table 4.4.2 we present the out-of-sample forecast errors for the whole period, i.e, 2016-2020. Among all the models, the 'Median' combination model exhibits the lowest errors when considering the MAE, RMSE and MAE%, and the 'Mean' estimator has the lowest value under RMSE%. The last column of Table 4.4.2

reports the results of the MCS procedure. Based on MSE, the full sample MCS contains the 'Mean' and 'Median' models at 75% confidence level.

Table 4.4.3 lists the out-of-sample forecasting errors of different models. In addition to the models described in Section 4.3, we report the errors obtained from the random walk. Also, the DM statistics are obtained based on the random walk model as the benchmark. Panel A covers the period from January 2011 to December 2019. Similar to the results for the whole sample period, the 'Median' model delivers the lowest values under all the loss functions except for RMSE%, and the 'Mean' estimator has the lowest RMSE%. However, when considering the RMSE and QLIKE only, 'Regression', 'Mean', 'Median', 'DMSP1', 'DMSP0.9', 'LASSO' and 'Elastic net' give similar errors[7]. It is notable that the DM statistics of all the models are negative, which indicates that all the models have smaller average MSE than the random walk in this period. More importantly, for the seven models we mentioned above, we can reject the null hypothesis of equal forecast accuracy at 5% significance level. In other words, these seven models outperform the random walk model for the period 2011-2019.

Panel B of Table 4.4.3 presents the out-of-sample performance for the year 2020. The RMSE of all the models increases while RMSE% decreases, which might be induced by the high average level of VVIX during the pandemic. Notably, the random walk model has the lowest RMSE among all the forecasting

[7]Patton (2011a) shows that, among all the loss functions, only MSE and QLIKE are robust to the noise in the volatility proxy.

Table 4.4.3: Out-of-sample forecasting errors (2)

Model	MAE	RMSE	QLIKE	MAE%	RMSE%	DM stat
Panel A. Year 2016-2019						
Random Walk	3.453	5.331	1.313	3.551	5.263	-
Regression	3.369	5.243	1.278	3.478	5.206	-2.308
ARMA	3.343	5.218	1.274	3.466	5.213	-1.797
HAR	3.379	5.260	1.287	3.491	5.220	-1.738
Mean	3.343	5.219	1.269	3.457	5.192	-2.604
Median	3.315	5.205	1.266	3.438	5.202	-2.017
DMSPE1	3.344	5.219	1.269	3.457	5.192	-2.600
DMSPE0.9	3.344	5.221	1.269	3.458	5.193	-2.596
Bayesian	3.379	5.260	1.286	3.490	5.219	-1.750
REG1	3.324	5.228	1.269	3.444	5.207	-1.918
REG2	3.402	5.292	1.299	3.520	5.260	-0.739
REG3	3.384	5.303	1.301	3.507	5.271	-0.288
LASSO	3.375	5.240	1.276	3.482	5.197	-2.512
Elastic net	3.372	5.241	1.277	3.479	5.201	-2.507
Panel B. Year 2020						
Random Walk	4.085	6.048	1.066	3.349	4.722	-
Regression	4.277	6.371	1.181	3.469	4.973	1.574
ARMA	4.055	6.173	1.106	3.342	4.842	0.597
HAR	4.137	6.164	1.104	3.399	4.827	0.745
Mean	4.096	6.137	1.099	3.361	4.813	0.730
Median	4.018	6.156	1.098	3.313	4.830	0.511
DMSPE1	4.096	6.142	1.100	3.361	4.816	0.760
DMSPE0.9	4.096	6.143	1.099	3.361	4.814	0.772
Bayesian	4.137	6.164	1.104	3.399	4.827	0.745
REG1	4.116	6.300	1.133	3.371	4.899	1.304
REG2	4.213	6.395	1.160	3.438	4.940	1.682
REG3	4.315	6.528	1.179	3.483	4.942	2.376
LASSO	4.090	6.140	1.095	3.358	4.805	0.510
Elastic net	4.091	6.141	1.095	3.359	4.805	0.502

This table presents the out-of-sample pricing errors. MAE is the average absolute error between the market price and the model price; RMSE is the square root of the average squared pricing error; QLIKE is given in equation (4.3.15). MAE% and RMSE% are in relative terms expressed in percentages with respect to the VVIX. DM denotes the Diebold-Mariano test statistic based on the MSE with the null hypothesis of equal accuracy and follows a N(0,1) distribution. The benchmark model is the random walk.

models, although it is not significantly different from all the other models (except 'REG3') in its ability to generate VVIX forecasts as shown by the DM statistics.

Table 4.4.4: Out-of-sample forecasting errors by VIX levels

	VIX≤15		15<VIX≤ 20		20<VIX≤ 35		VIX>35	
	obs.=667		*obs.*=257		*obs.*=282		*obs.*=51	
Model	RMSE	DM	RMSE	DM	RMSE	DM	RMSE	DM
Random walk	3.727	-	6.106	-	6.273	-	12.394	-
Regression	3.677	-2.781	5.952	-3.421	6.133	-1.301	13.379	2.074
ARMA	3.576	-3.029	5.974	-1.799	6.078	-1.357	13.173	1.454
HAR	3.636	-3.338	6.041	-0.997	6.141	-1.479	12.950	1.422
Mean	3.613	-4.130	5.965	-2.821	6.081	-1.953	12.977	1.789
Median	3.561	-3.488	5.953	-2.224	6.047	-1.565	13.215	1.534
DMSPE1	3.613	-4.131	5.965	-2.815	6.081	-1.950	12.989	1.806
DMSPE0.9	3.614	-4.188	5.965	-2.879	6.080	-1.974	13.002	1.853
Bayesian	3.635	-3.374	6.040	-1.007	6.141	-1.477	12.951	1.424
REG1	3.545	-3.724	5.977	-1.735	6.147	-1.281	13.466	2.143
REG2	3.596	-3.323	6.092	-0.162	6.236	-0.371	13.518	2.158
REG3	3.615	-2.343	5.945	-1.551	6.269	-0.034	14.090	2.237
LASSO	3.651	-3.319	5.970	-3.573	6.093	-2.165	12.965	1.600
Elastic net	3.657	-3.260	5.972	-3.736	6.085	-2.231	12.966	1.613

This table presents the out-of-sample RMSE for different levels of the VIX. DM denotes the Diebold-Mariano test statistic based on the MSE with the null hypothesis of equal accuracy and follows a N(0,1) distribution. The benchmark model is the random walk. *obs.* is the number of observations.

To investigate whether the market volatility affects the VVIX forecasting performance, Table 4.4.4 reports the out-of-sample errors by VIX levels. When the VIX level is less than or equal to 15, all the models outperform the random walk with, surprisingly, the 'REG1' presenting the lowest MSE. When the value of the VIX is in the range of (15,20], 'Regression', 'Mean', 'Median', 'DMSP1', 'DMSP0.9', 'LASSO' and 'Elastic net' models have more accurate forecasts compared with the random walk. We find that these seven models are identical with the models which outperform the random walk for the period 2016-2019. Furthermore, when the VIX falls into (20,35], only 'LASSO' and 'Elastic net' exhibit a significant outperformance over the random walk. However, if the VIX increases above 35, which accounts for 4% of the total observations, based on the DM statistics we can conclude that most of the models are not significantly different from the random walk in terms of VVIX forecasting. Also noteworthy is the fact as the level of the VIX index rises, the RMSE increases and the number of models which are superior to the random walk decreases.

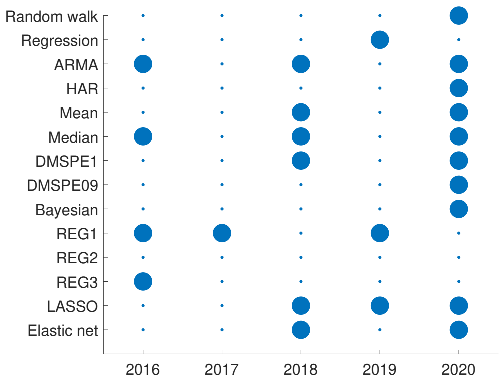
Table 4.4.5: Out-of-sample forecasting comparison: MCS test

Model	MSE based		QLIKE based	
	R method	SQ method	R method	SQ method
Random walk	1	1	1	1
Regression	1	1	2	2
ARMA	3	3	2	2
HAR	1	1	1	1
Mean	2	3	3	5
Median	3	4	5	5
DMSPE1	2	3	2	4
DMSPE0.9	1	2	1	2
Bayesian	1	1	1	1
REG1	3	4	3	4
REG2	0	0	0	1
REG3	1	1	2	2
LASSO	3	3	3	3
Elastic net	2	3	3	3

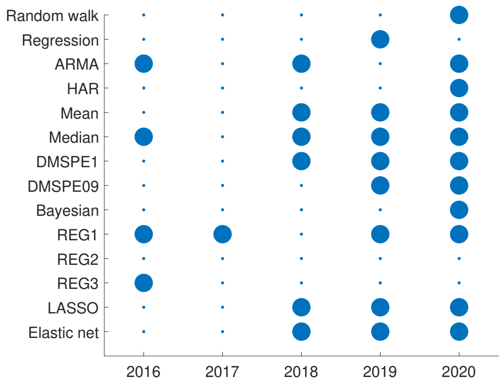
This table presents the number of out-of-sample years for which each model is within the MCS at 75% confidence level; the data sample is from 2016 to 2020. The MCS test employs two methods: the range method (R Method) and semi-quadratic method (SQ Method). The loss functions considered are MSE and QLIKE. The MCS test is based on 10,000 bootstraps.

Apart from the discussion on the forecasting errors and DM test, we also conduct two additional analysis. The first one is the MCS test which identifies the best subset out of the entire model family. In addition to Table 4.4.2 which shows the MCS results for the whole period, Figure 4.4.3 illustrates the MCS test by years. Panel A to Panel D displays the results of MCS using the R method and SQ method under MSE and QLIKE, respectively. The models selected by MCS are quite similar across different methods for each year, except that 'REG1' dominates in 2017 when considering the MSE-based procedure. Overall, the 'Median' model is most likely to survive in the MCS when considering all the methods and loss functions used in the test. Also, Table 4.4.5 confirms this finding by presenting the number of years in which each model is in the MCS.

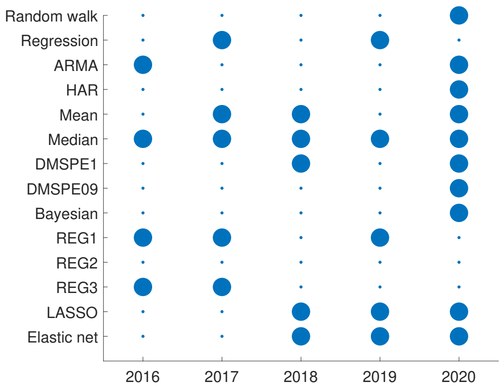
Figure 4.4.3: Out-of-sample comparison: MCS test results



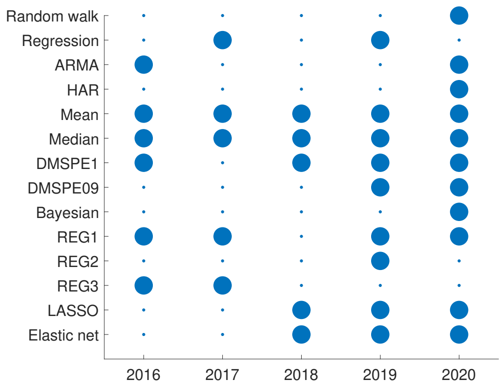
(a) R Method based on MSE



(b) SQ Method based on MSE



(c) R Method based on QLIKE



(d) SQ Method based on QLIKE

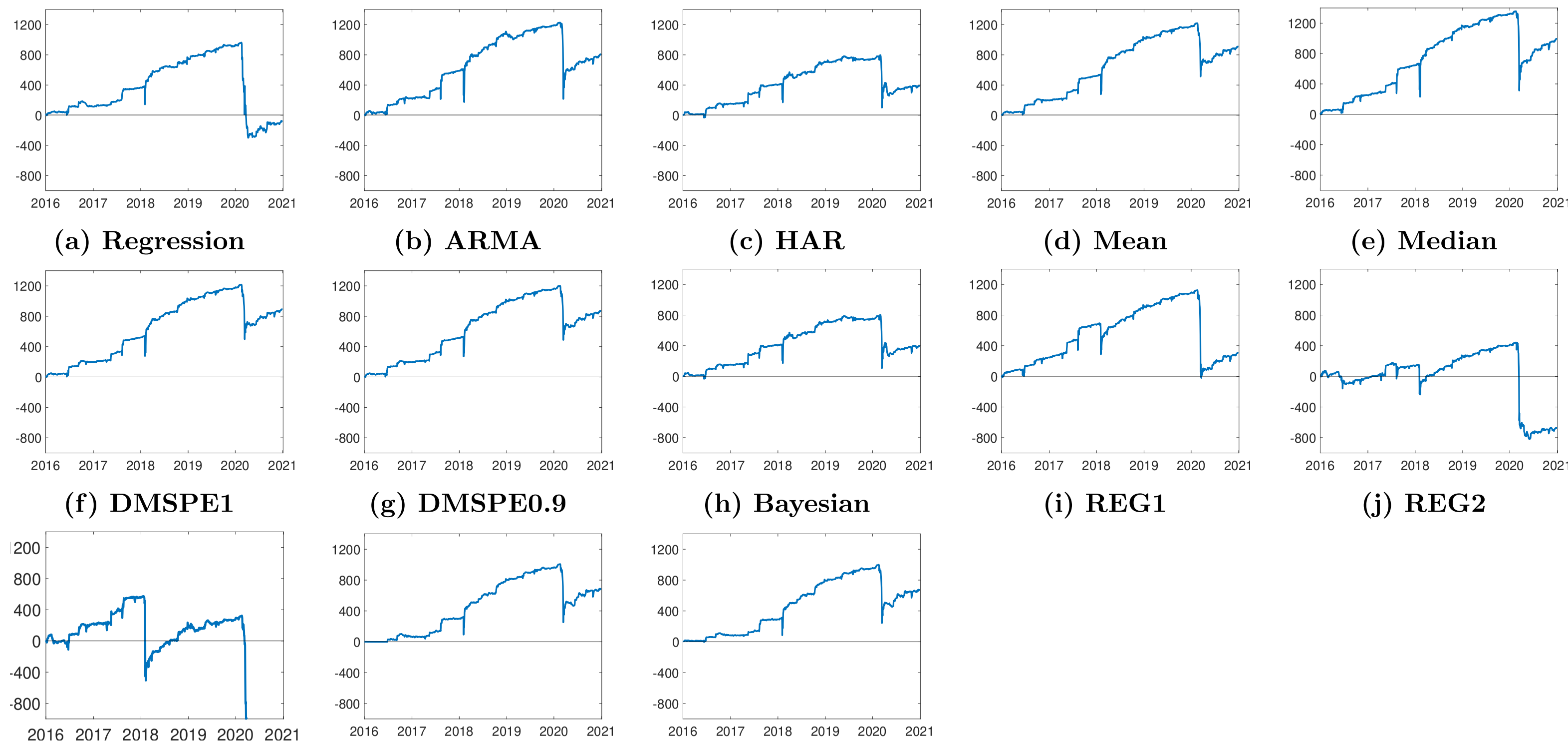
Notes: This chart illustrates the MCS test results by years. The bubble signifies that the corresponding model is within the MCS at 75% confidence level for a given year. The data ranges from 2016 to 2020. The MCS test employs two methods: the range method (R Method) and semi-quadratic method (SQ Method). The loss functions considered are MSE and QLIKE. The MCS test is based on 10,000 bootstraps.

Following Rapach et al. (2010), we plot the time series of the difference between the cumulative squared error (CSE) of the random walk and the cumulative squared error of each model over the whole period, as shown in Figure 4.4.4. This gives a simple visual impression of how each forecasting model differs from the random walk during the out-of-sample period. All the models (except 'REG2' and 'REG3') display an upward trend before the year 2020, which indicates that those models consistently outperform the random walk during 2016-2019. However, the line in each panel drops dramatically in March 2020, which means that all forecasting models fail to capture the spikes in the VVIX index. Furthermore, we compare the height of the curve at the end of the period: the high endpoint in Panel (e) demonstrates that the 'Median' model has a lower MSE than the other models. This is consistent with our earlier findings. Also, 'ARMA', 'Mean', 'DMSPE1', 'DMSPE0.9', 'LASSO' and 'Elastic net' are shown to outperform the random walk over the whole sample.

4.4.3 Variable selection

Although the 'Median' model combination provides the best out-of-sample performance considering the overall period, 'LASSO' and 'Elastic net' also deliver more accurate forecasts than the random walk. Considering that these models, i.e., 'LASSO' and 'Elastic net', are able to offer model selection among a number of feasible variables, it is of interest to examine the importance of the various predictors for VVIX forecasting. We consider the following 17 variables from the individual models, which also include all the candidates for 'Regression': the

Figure 4.4.4: Cumulative square prediction errors



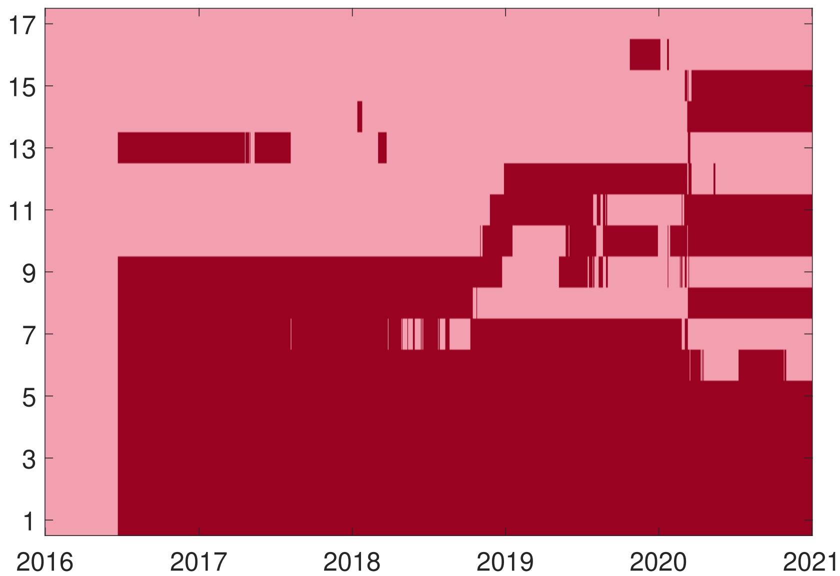
(a) Regression (b) ARMA (c) HAR (d) Mean (e) Median

(f) DMSPE1 (g) DMSPE0.9 (h) Bayesian (i) REG1 (j) REG2

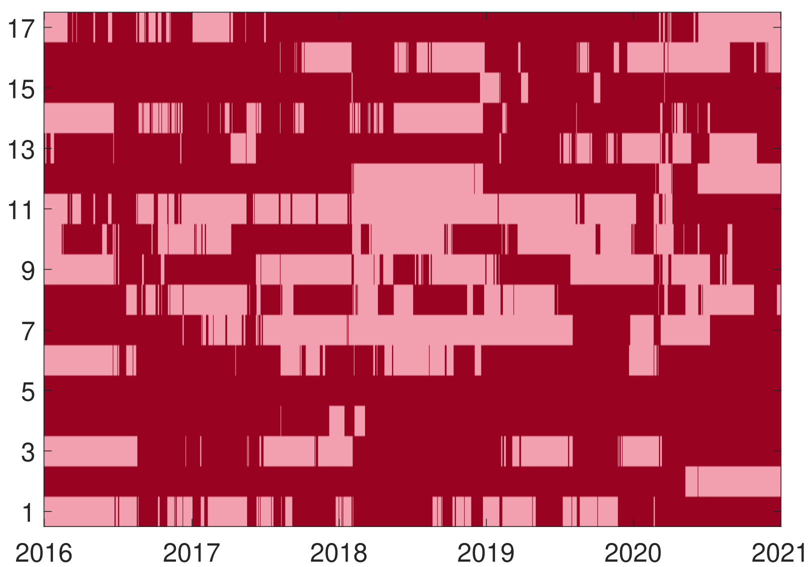
(k) REG3 (l) LASSO (m) Elastic net

Notes: Cumulative square prediction errors for the random walk model minus the cumulative square prediction errors for each forecasting model, January 2016 to December 2020.

Figure 4.4.5: Variable selection for 2016-2020



(a) LASSO



(b) Elastic net

Notes: This chart shows the variables selected for forecasting the VVIX index over the period January 2016 to December 2020. The y-axis is the 17 variables which are potentially related to the daily changes of VVIX. The dark-red blocks indicate that the variable is selected by the predictive regression on a given day.

lagged daily changes in VVIX ($\Delta VVIX$), the daily changes in the VIX (ΔVIX), daily changes in the average weekly VVIX ($\Delta VVIX^w$), the daily changes in the average monthly VVIX ($\Delta VVIX^m$), the difference between the VIX3M and the VIX index (VX^{diff}), the log-returns of the S&P 500 index (ret), the squared returns (ret^2), the difference between the 3-month treasury bill rate and the 1-month treasury bill rate (TB_{6m}^{diff}), the difference between the 6-month treasury bill rate and the 1-month treasury bill rate (TB_{3m}^{diff}), and their absolute values (except the squared returns).

Figure 4.4.5 shows the variable selection results of the 'LASSO' and 'Elastic net' methods, respectively. An impressive finding is that, compared with 'Elastic net', the 'LASSO' method tends to select less variables and thus have more 'blocks' predictors over time. In 'LASSO', the top five variables selected are ΔVIX, $\Delta VVIX^m$, $|\Delta VVIX^w|$, $|\Delta VVIX^m|$ and $|VX^{diff}|$, respectively; while the top five selected by 'Elastic net' are $\Delta VVIX^m$, $|\Delta VVIX^m|$, ret, $\Delta VVIX$ and TB_{3m}^{diff}. It is notable that the changes in the average monthly VVIX and its absolute value appear more often in the selected predictors of VVIX for both methods.

4.5 Conclusions

In the financial market, VVIX is an important indicator of how rapidly market volatility changes rather than of the volatility itself. Motivated by the success of forecast combinations and LASSO-type shrinkage methods, this chapter seeks to answer the question: is there an optimal forecasting method for the VVIX index.

In this chapter, we examine the forecasting performance of three individual models, eight combination methods and two LASSO-type models out-of-sample over the period 2016-2020. We find that the simple 'Median' method yields the lowest MSE across years. Moreover, the results of the MCS procedure shows that both the 'Mean' and 'Median' methods outperform the other models for the overall period. Furthermore, the model selection results of both 'LASSO' and 'Elastic net' methods suggest that, instead of daily changes in the VVIX index, the changes in the monthly VVIX are of vital importance in predicting the VVIX.

Chapter 5

Conclusions and Further Research

5.1 Summary of the Findings and Contributions of the Thesis

This thesis makes original contributions to the volatility forecasting literature, specifically regarding the uses of volatility indices. Thus, it benefits both academicians and financial practitioners because it provides valuable lessons regarding the information contained in the volatility indices.

In Chapter 2, we propose a new VIX forecasting method employing the filtered historical simulations put forward in Barone-Adesi et al. (2008) and the information on the VIX term structure. This approach provides estimates using the empirical innovation density that can capture the non-normal features of returns, such as negative skewness and positive excess kurtosis. Different from the traditional methods that use a cross-section of options data, we calibrate the model by applying four well-established volatility indices, i.e., VIX9D, VIX, VIX3M and VIX6M. We find that this method outperforms, both in-sample and out-of-sample, the benchmark model which only uses the VIX index and assumes a normal distribution. Additionally, the NAGARCH model based on the new method is superior to all the other competing models for long-term volatility forecasts, while the GJR model under the proposed estimation approach outperforms all the other models for short-run volatility forecasts. Also, we perform statistical tests and several robustness checks that confirm our results. More importantly, we provide evidence that our proposed estimation method significantly reduces the computational time.

In Chapter 3, we explore the usefulness of adding the VIX term structure to VIX futures pricing models. Similarly to Chapter 2, the estimation assumes the empirical innovation density to accommodate for the non-normality of returns. The parameters of the GJR model are then calibrated from the historical futures data, or the information on the VIX term structure, or their combinations. Our analysis of the out-of-sample forecasting performance suggests that, for most of the years, the performance of the models that use both the VIX term structure and the futures prices is not significantly different from the performance of the models that only include data on the futures. Also, we examine the forecasting performance of different models during the recent 2020 COVID-19 crisis. Our empirical results show that the use of the VIX term structure leads to the lower pricing errors. Moreover, we investigate the model performance when using different maturities and VIX levels. Our findings are that, compared to the model that uses only futures data, models that incorporate information on the VIX term structure into VIX futures pricing models can provide better forecasts when 1) the future's maturity is longer than 120 days; or 2) the VIX level is higher than 15.

In Chapter 4, we analyse VVIX forecasting methods. Motivated by the success of forecast combinations and the LASSO-type shrinkage methods, we attempt to answer the following question: is there an optimal VVIX forecasting method? If yes, then is this based on forecast combinations or LASSO? We find that forecast combinations perform best. We compare the forecasting performance of three individual models, eight combining methods and two LASSO-type models out-of-sample. The results show that the simple median combining method delivers the lowest forecasting errors across the years. However, we find that both the mean and median combining methods end up in the model confidence set at 75% confidence level for the full sample period. In addition, we discuss the model selection results of two shrinkage methods, i.e., LASSO and elastic net. Interestingly, instead of daily changes in the VVIX, the changes in monthly VVIX are key to predict the VVIX.

5.2 Suggestions for Future Research

Although this thesis has important implications for volatility forecasting, there are still many gaps. In this section, we address future research directions based on the findings of this thesis.

Forecasting VIX Chapter 2 estimate a volatility model that assumes an empirical innovation density which captures the non-normality of returns. First, we briefly discuss the variance risk premium in Section 2.3.2 with a focus on forecasting the VIX index. As suggested by Bollerslev et al. (2009), an estimate of the variance risk premium predicts stock returns. Therefore, given the estimated variance under the physical measure and the risk-neutral estimated variance, our analysis may be extended to include more details on the variance risk premium. For example, one can compare the value of variance risk premium captured by the new methods proposed compared to models introduced in the recent literature.

Second, this model may be extended to include a jump component in order to capture the spikes of the volatility dynamics. In the table that reports results on the out-of-sample comparison (Table 2.3.3), the GARCH-type models that use the proposed method exhibit the highest pricing errors in 2011, the year in which the VIX is quite volatile, compared to the results for the other years. Therefore, it is worth to investigate whether a jump component improves the forecasting performance during turbulent markets.

Third, the proposed model is calibrated from the VIX term structure, which are calculated using out-of-the-money options. It would be interesting to examine the option pricing performance of the new method.

VIX futures Chapter 3 explores the effects of the VIX term structure on the one-day-ahead VIX futures forecasts. First, a promising extension would be considering longer forecasting horizons, i.e., one-week-ahead and one-month-ahead forecasts. Also, it is not rare in the VIX futures pricing literature to evaluate the model performance by basis, which is defined as the difference between the VIX level and the VIX futures prices.

Second, inspired by the VVIX forecasting results in Chapter 4, another feasible extension would enhance the VVIX term structure forecasts using combination methods in order to improve on the VIX futures pricing model.

VVIX forecasts In Chapter 4, we employ a traditional split cross validation method for time series to estimate α and λ *ex ante* in a LASSO-type regression. However, recent literature, for example, Zhang et al. (2019a), propose a new algorithm to identify α and λ, which delivers a better forecasting performance. It may be worth combining LASSO with the new algorithm.

Finally, with respect to the combining methods, one may examine the newly proposed 'partially egalitarian LASSO' (peLASSO) method of Diebold and Shin (2019), which discards some forecasts and shrinks the survivors toward equality. The peLASSO, which is shown to outperform simple average and median forecasts, would be an interesting extension to attempt.

References

Andersen, T. G., Bollerslev, T., Diebold, F. X., 2007. Roughing It Up: Including Jump Components in the Measurement, Modeling, and Forecasting of Return Volatility. The Review of Economics and Statistics 89, 701–720.

Audrino, F., Knaus, S. D., 2016. Lassoing the HAR Model: A Model Selection Perspective on Realized Volatility Dynamics. Econometric Reviews 35, 1485–1521.

Ballestra, L. V., Guizzardi, A., Palladini, F., 2019. Forecasting and trading on the VIX futures market: A neural network approach based on open to close returns and coincident indicators. International Journal of Forecasting 35, 1250–1262.

Bandi, F. M., Perron, B., 2006. Long Memory and the Relation Between Implied and Realized Volatility. Journal of Financial Econometrics 4, 636–670.

Barone-Adesi, G., Engle, R. F., Mancini, L., 2008. A GARCH option pricing model with filtered historical simulation. Review of Financial Studies 21, 1223–1258.

Becker, R., Clements, A. E., 2008. Are combination forecasts of S&P 500 volatility statistically superior? International Journal of Forecasting 24, 122–133.

Bekaert, G., Hoerova, M., 2014. The VIX, the variance premium and stock market volatility. Journal of Econometrics 183, 181–192.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

Bollerslev, T., 1987. A conditionally heteroskedastic time series model for speculative prices and rates of return. Review of Economics and Statistics 69, 542–547.

Bollerslev, T., Gibson, M., Zhou, H., 2011. Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. Journal of Econometrics 160, 235 – 245, realized Volatility.

Bollerslev, T., Tauchen, G., Zhou, H., 2009. Expected Stock Returns and Variance Risk Premia. The Review of Financial Studies 22, 4463–4492.

Bollerslev, T., Wooldridge, J. M., 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. Econometric Reviews 11, 143–172.

Bu, R., Fu, X., Jawadi, F., 2019. Does the volatility of volatility risk forecast future stock returns? Journal of International Financial Markets, Institutions and Money 61, 16–36.

Byun, S. J., Min, B., 2013. Conditional volatility and the GARCH option pricing model with non-normal innovations. Journal of Futures Markets 33, 1–28.

Carr, P., Lee, R., 2009. Volatility derivatives. Annual Review of Financial Economics 1, 319–339.

Carr, P., Madan, D., 1998. Towards a theory of volatility trading. In Robert Jarrow (Ed.). Volatility Estimation Techniques for Pricing Derivatives pp. 417–427.

Carr, P., Wu, L., 2006. A Tale of Two Indices. The Journal of Derivatives 13, 13 – 29.

Carr, P., Wu, L., 2009. Variance risk premiums. Review of Financial Studies 22, 1311–1341.

Christoffersen, P., Elkamhi, R., Feunou, B., Jacobs, K., 2009. Option valuation with conditional heteroskedasticity and nonnormality. Review of Financial Studies 23, 2139–2183.

Christoffersen, P., Heston, S., Jacobs, K., 2006. Option valuation with conditional skewness. Journal of Econometrics 131, 253–284.

Christoffersen, P., Jacobs, K., 2004. Which GARCH model for option valuation? Management Science 50, 1204–1221.

Christoffersen, P., Jacobs, K., Ornthanalai, C., 2013. GARCH option valuation: Theory and evidence. Journal of Derivatives 21, 8–41.

Clemen, R. T., 1989. Combining forecasts: A review and annotated bibliography. International Journal of Forecasting 5, 559–583.

Clements, M. P., Hendry, D. F., 2004. Pooling of forecasts.

Corrado, C. J., Miller, Jr., T. W., 2005. The forecast quality of cboe implied volatility indexes. Journal of Futures Markets 25, 339–373.

Corsi, F., 2009. A Simple Approximate Long-Memory Model of Realized Volatility. Journal of Financial Econometrics 7, 174–196.

Corsi, F., Pirino, D., Renò, R., 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. Journal of Econometrics 159, 276–288.

Demeterfi, K., Derman, E., Kamal, M., Zou, J., 1999. A guide to volatility and variance swaps. Journal of Derivatives 6, 9–32.

Dickey, D. A., Fuller, W. A., 1981. Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. Econometrica 49, 1057–1072.

Diebold, F. X., Mariano, R. S., 1995. Comparing Predictive Accuracy. Journal of Business & Economic Statistics 20, 134–144.

Diebold, F. X., Shin, M., 2019. Machine learning for regularized survey forecast combination: Partially-egalitarian lasso and its derivatives. International Journal of Forecasting 35, 1679–1691.

Dotsis, G., Psychoyios, D., Skiadopoulos, G., 2007. An empirical comparison of continuous-time models of implied volatility indices. Journal of Banking & Finance 31, 3584–3603.

Duan, J.-C., 1995. The GARCH option pricing model. Mathematical Finance 5, 13–32.

Duan, J.-C., Yeh, C.-Y., 2010. Jump and volatility risk premiums implied by VIX. Journal of Economic Dynamics and Control 34, 2232–2244.

Dumas, B., Fleming, J., Whaley, R. E., 1998. Implied volatility functions: Empirical tests. Journal of Finance 53, 2059–2106.

Dupire, B., 2006. Model free results on volatility derivatives (working paper). SAMSI, Research Triangle Park .

Engle, R. F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987–1007.

Engle, R. F., Ng, V. K., 1993. Measuring and testing the impact of news on volatility. Journal of Finance 48, 1749–1778.

Fernandes, M., Medeiros, M. C., Scharth, M., 2014. Modeling and predicting the cboe market volatility index. Journal of Banking and Finance 40, 1–10.

Giacomini, R., White, H., 2006. Tests of Conditional Predictive Ability. Econometrica 74, 1545–1578.

Glosten, L. R., Jagannathan, R., Runkle, D. E., 1993. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. The Journal of Finance 48, 1779–1801.

Granger, C. W. J., Ramanathan, R., 1984. Improved methods of combining forecasts. Journal of Forecasting 3, 197–204.

Guo, S., Liu, Q., 2020. Efficient Out-of-Sample Pricing of VIX Futures. Journal of Derivatives 27, 126 LP – 139.

Hansen, P. R., 2005. A test for superior predictive ability. Journal of Business & Economic Statistics 23, 365–380.

Hansen, P. R., Lunde, A., Nason, J. M., 2011. The Model Confidence Set. Econometrica 79, 453–497.

Hao, J., Zhang, J. E., 2013. GARCH option pricing models, the CBOE VIX, and variance risk premium. Journal of Financial Econometrics 11, 556–580.

Heston, S. L., 1993. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. The Review of Financial Studies 6, 327–343.

Heston, S. L., Nandi, S., 2000. A closed-form GARCH option valuation model. Review of Financial Studies 13, 585–625.

Hollstein, F., Prokopczuk, M., 2018. How Aggregate Volatility-of-Volatility Affects Stock Returns. The Review of Asset Pricing Studies 8, 253–292.

Hsiao, C., Wan, S. K., 2014. Is there an optimal forecast combination? Journal of Econometrics 178, 294–309.

Huang, D., Schlag, C., Shaliastovich, I., Thimme, J., 2019a. Volatility-of-Volatility Risk. Journal of Financial and Quantitative Analysis 54, 2423–2452.

Huang, Z., Tong, C., Wang, T., 2019b. VIX term structure and vix futures pricing with realized volatility. Journal of Futures Markets 39, 72–93.

Jeon, B., Seo, S. W., Kim, J. S., 2020. Uncertainty and the volatility forecasting power of option-implied volatility. Journal of Futures Markets 40, 1109–1126.

Jiang, Y., Lazar, E., 2020. Forecasting VIX using filtered historical simulation. Journal of Financial Econometrics Nbaa041.

Kambouroudis, D. S., McMillan, D. G., 2016. Does VIX or volume improve GARCH volatility forecasts? Applied Economics 48, 1210–1228.

Kanniainen, J., Lin, B., Yang, H., 2014. Estimating and using GARCH models with VIX data for option valuation. Journal of Banking & Finance 43, 200–211.

Konstantinidi, E., Skiadopoulos, G., 2011a. Are VIX futures prices predictable? An empirical investigation. International Journal of Forecasting 27, 543–560.

Konstantinidi, E., Skiadopoulos, G., 2011b. Are VIX futures prices predictable? An empirical investigation. International Journal of Forecasting 27, 543–560.

Konstantinidi, E., Skiadopoulos, G., Tzagkaraki, E., 2008. Can the evolution of implied volatility be forecasted? evidence from european and us implied volatility indices. Journal of Banking and Finance 32, 2401–2411.

Krause, T. A., 2019. Hedge fund returns and uncertainty. The North American Journal of Economics and Finance 47, 597–601.

Lin, Y.-N., 2007. Pricing VIX futures: Evidence from integrated physical and risk-neutral probability measures. Journal of Futures Markets 27, 1175–1217.

Liu, Q., Guo, S., Qiao, G., 2015. VIX forecasting and variance risk premium: A new GARCH approach. North American Journal of Economics and Finance 34, 314–322.

Majewski, A. A., Bormetti, G., Corsi, F., 2015. Smile from the past: A general option pricing framework with multiple volatility and leverage components. Journal of Econometrics 187, 521–531.

Mandelbrot, B., 1963. The variation of certain speculative prices. Journal of Business 36, 394–419.

Mencía, J., Sentana, E., 2013. Valuation of VIX derivatives. Journal of Financial Economics 108, 367–391.

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

Park, Y.-H., 2015. Volatility-of-volatility and tail risk hedging returns. Journal of Financial Markets 26, 38–63.

Patton, A. J., 2011a. Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics 160, 246–256.

Patton, A. J., 2011b. Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics 160, 246 – 256.

Patton, A. J., Sheppard, K., 2009. Optimal combinations of realised volatility estimators. International Journal of Forecasting 25, 218–238.

Patton, A. J., Sheppard, K., 2015. Good volatility, bad volatility: Signed jumps and the persistence of volatility. Review of Economics and Statistics 97, 683–697.

Rapach, D. E., Strauss, J. K., Zhou, G., 2010. Out-of-Sample Equity Premium Prediction: Combination Forecasts and Links to the Real Economy. The Review of Financial Studies 23, 821–862.

Stentoft, L., 2008. American option pricing using GARCH models and the normal inverse gaussian distribution. Journal of Financial Econometrics 6, 540–582.

Szado, E., 2009. VIX futures and options: A case study of portfolio diversification during the 2008 financial crisis. The Journal of Alternative Investments 12, 68–85.

Taylor, N., 2019. Forecasting returns in the VIX futures market. International Journal of Forecasting 35, 1193–1210.

Tibshirani, R., 1996. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological) 58, 267–288.

Timmermann, A., 2006. Forecast combinations. In: G. Elliott, C. W. J. Granger, &. A. G. T. (ed.), *Handbook of economic forecasting*, North Holland Press., Amsterdam.

Wang, T., Shen, Y., Jiang, Y., Huang, Z., 2017. Pricing the CBOE VIX futures with the Heston–Nandi garch model. Journal of Futures Markets 37, 641–659.

Wang, Y., Ma, F., Wei, Y., Wu, C., 2016. Forecasting realized volatility in a changing world: A dynamic model averaging approach. Journal of Banking & Finance 64, 136–149.

Whaley, R. E., 1993. Derivatives on market volatility: Hedging tools long overdue. Journal of Derivatives 1, 71–84.

Whaley, R. E., 2009. Understanding the VIX. The Journal of Portfolio Management 35, 98 – 105.

Xie, H., Zhou, M., Ruan, T., 2020. Pricing VIX futures under the GJR-GARCH process: An analytical approximation method. Journal of Derivatives 27, 77–88.

Zang, X., Ni, J., Huang, J.-Z., Wu, L., 2017. Double-jump diffusion model for VIX: evidence from VVIX. Quantitative Finance 17, 227–240.

Zhang, J. E., Shu, J., Brenner, M., 2010. The new market for volatility trading. Journal of Futures Markets 30, 809–833.

Zhang, J. E., Zhu, Y., 2006. VIX futures. Journal of Futures Markets 26, 521–531.

Zhang, Y., Ma, F., Wang, Y., 2019a. Forecasting crude oil prices with a large set of predictors: Can LASSO select powerful predictors? Journal of Empirical Finance 54, 97–117.

Zhang, Y., Wei, Y., Zhang, Y., Jin, D., 2019b. Forecasting oil price volatility: Forecast combination versus shrinkage method. Energy Economics 80, 423–433.

Zhu, S.-P., Lian, G.-H., 2012. An analytical formula for VIX futures and its applications. Journal of Futures Markets 32, 166–190.

Zhu, X., Zhu, J., 2013. Predicting stock returns: A regime-switching combination approach and economic links. Journal of Banking & Finance 37, 4120–4133.

Zhu, Y., Zhang, J. E., 2007. Variance term structure and VIX futures pricing. International Journal of Theoretical and Applied Finance 10, 111–127.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67, 301–320.