Forecasting VIX using filtered historical simulation

Article

Accepted Version

Jiang, Y. and Lazar, E. ORCID: https://orcid.org/0000-0002-8761-0754 (2022) Forecasting VIX using filtered historical simulation. Journal of Financial Econometrics, 20 (4). pp. 665-680. ISSN 1479-8417 doi: 10.1093/jjfinec/nbaa041 Available at https://centaur.reading.ac.uk/93184/

To link to this article DOI: http://dx.doi.org/10.1093/jjfinec/nbaa041

Publisher: Oxford University Press

Forecasting VIX using filtered historical simulation*

Yushuang Jiang[1] Emese Lazar[1†]

[1] ICMA Centre
Henley Business School
University of Reading

September 30, 2020

*We are grateful to the editor Andrew Patton and two anonymous referees for their insightful comments which helped improve the paper. We would also like to thank Michael Clements, Jin E. Zhang, and participants at MMF-EFiC Program 2018 for their helpful suggestions and discussions.

†Address correspondence to Emese Lazar, ICMA Centre, Henley Business School, University of Reading, Whiteknights, Reading RG6 6BA, UK, or email: e.lazar@icmacentre.ac.uk

Abstract

We propose a new VIX forecast method using GARCH models based on the filtered historical simulation put forward in Barone-Adesi et al. (2008). The flexible change of measure accommodates for non-normalities such as negative skewness and positive excess kurtosis. We present an application for four well-established volatility indices (VIX9D, VIX, VIX3M and VIX6M). Our results show that our proposed estimation method outperforms the Normal-VIX model of Hao and Zhang (2013) both in-sample and out-of-sample. Furthermore, the use of volatility indices reduces the computational burden significantly compared to the options based pricing method.

Key words: GARCH; historical filtered simulation; CBOE volatility index

JEL Classifications: C53, C58, G17

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-normalities 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 paper, we propose an alternative, faster approach to forecast volatility, which uses volatility indices information, extending the methodology of Hao and Zhang (2013). 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), and employs filtered historical returns. 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 likelihood estimation (MLE) using equity returns which produces estimates under the physical measure. Researchers then adjust the estimates to price options as in Bollerslev and Mikkelsen (1996). 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. 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 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. (2019) who estimate the extended leverage heterogeneous autoregressive gamma (LHARG) model of Majewski et al. (2015) using both the VIX term structure and VIX futures.

However, the current literature on GARCH option pricing using CBOE VIX considers only normally distributed returns. In the approach presented in this paper 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 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 poor empirical performance. Therefore, in this paper, we consider the volatility processes 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 the paper is organized as follows. Section 1 presents the new estimation method that uses the filtered historical simulation and the CBOE volatility indices. Section 2 provides the empirical results and analysis, and Section 3 concludes the study.

1 The models

In this section, we introduce the different GARCH model estimations we investigate in this study. 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.

1.1 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:

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

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:

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

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.

1.2 The Normal-VIX method

Hao and Zhang (2013) pioneer using 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[1]. The estimation is then carried out within a set of GARCH model specifications using both the returns and the VIX. The GJR model defined under the LRNVR is[2]

$$\begin{aligned}
\text{Physical measure:} \quad & \ln(S_t/S_{t-1}) = r_t + \lambda\sigma_t - \frac{1}{2}\sigma_t^2 + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t \\
& \sigma_t^2 = \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 + \gamma I_{t-1}\varepsilon_{t-1}^2 \\
\text{Risk-neutral measure:} \quad & \ln(S_t/S_{t-1}) = r_t - \frac{1}{2}\sigma_t^2 + \xi_t, \quad \xi_t = \sigma_t z_t \\
& \sigma_t^2 = \omega + \alpha(\xi_{t-1} - \lambda\sigma_{t-1})^2 + \beta\sigma_{t-1}^2 + \gamma I_{t-1}(\xi_{t-1} - \lambda\sigma_{t-1})^2
\end{aligned} \tag{3}$$

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:

$$Vix_t = A + B\sigma_{t+1}^2, \tag{4}$$

where

$$
\begin{aligned}
&Vix_t = (VIX_t/100)^2/252, \\
&A = \frac{\omega}{1-\eta}(1-B), \\
&B = \frac{1-\eta^n}{n(1-\eta)}, \\
&\eta = \alpha(1+\lambda^2) + \beta + \gamma S.
\end{aligned}
\tag{5}
$$

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.

1.3 CBOE volatility indices

In this section, we briefly describe the CBOE volatility indices which measure the market expectation of volatility implied from 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 paper 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[3],

$$
\sigma^2 = \frac{2}{T}\sum_i \frac{\triangle K_i}{K_i^2} e^{RT} Q(K_i) - \frac{1}{T}[\frac{F}{K_0} - 1]^2, \tag{6}
$$

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:

$$VIX_t = 100 * \sqrt{\frac{\tau}{T} * \sum_{k=1}^{30} E^Q[\sigma_{t+k}^2 \mid \psi_t]}. \tag{7}$$

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

1.4 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 non-linear 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:

$$\ln(S_t/S_{t-1}) = \mu_t - \kappa_t + \varepsilon_t, \quad \varepsilon_t = \sigma_t z_t \tag{8}$$

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, where we estimate using the empirical distribution function. κ_t is the mean correction factor defined as:

$$\kappa_t = \ln(E_{t-1}[\exp\{\varepsilon_t\}]) \tag{9}$$

We have:

$$E_{t-1}[S_t/S_{t-1}] = E_{t-1}[\exp\{\mu_t - \kappa_t + \varepsilon_t\}] = \exp\{\mu_t\}. \tag{10}$$

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

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + g(\varepsilon_{t-1}) \tag{11}$$

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

$$\begin{aligned} \text{GARCH:} \quad & g(\varepsilon_{t-1}) = \alpha\varepsilon_{t-1}^2 \\ \text{NAGARCH:} \quad & g(\varepsilon_{t-1}) = \alpha(\varepsilon_{t-1} - \theta\sigma_{t-1})^2 \\ \text{GJR:} \quad & g(\varepsilon_{t-1}) = [\alpha + \gamma I(\varepsilon_{t-1} < 0)]\varepsilon_{t-1}^2 \end{aligned} \tag{12}$$

For the NAGARCH and GJR models, a positive θ and γ ensure an asymmetric response of the volatility to positive and negative returns.

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 (11):

$$\ln L_R = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^{n}\left\{\ln(\sigma_t^2) + \frac{(R_t - \mu_t + \kappa_t)^2}{\sigma_t^2}\right\} \tag{13}$$

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

Under the risk-neutral measure we have that:

$$\ln(S_t/S_{t-1}) = r_t - \kappa_t^* + \varepsilon_t^*, \quad \varepsilon_t^* = \sigma_t^* z_t^* \tag{14}$$

where r_t is the risk-free rate at time t which is same as in the LRNVR framework of equation (3), and κ_t^* is the mean correction factor under the risk-neutral measure:

$$\kappa_t^* = \ln(E_{t-1}[\exp\{\varepsilon_t^*\}]) \tag{15}$$

so that

$$E_{t-1}[S_t/S_{t-1}] = E_{t-1}[\exp\{r_t - \kappa_t^* + \varepsilon_t^*\}] = \exp\{r_t\}. \tag{16}$$

The conditional variance dynamics are as follows:

$$\sigma_t^{*2} = \omega^* + \beta^*\sigma_{t-1}^{*2} + g^*(\varepsilon_{t-1}^*) \tag{17}$$

where for the different models we have:

$$\begin{aligned} \text{GARCH:} \quad & g^*(\varepsilon_{t-1}^*) = \alpha^*\varepsilon_{t-1}^{*2} \\ \text{NAGARCH:} \quad & g^*(\varepsilon_{t-1}^*) = \alpha^*(\varepsilon_{t-1}^* - \theta^*\sigma_{t-1}^*)^2 \\ \text{GJR:} \quad & g^*(\varepsilon_{t-1}^*) = [\alpha^* + \gamma^* I(\varepsilon_{t-1}^* < 0)]\varepsilon_{t-1}^{*2}. \end{aligned} \tag{18}$$

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:

$$\frac{dQ}{dP} \mid \mathcal{F}_t = \exp\left(- \sum_{i=1}^{t} (\nu_i \varepsilon_i + \Psi_i(\nu_i)) \right) \tag{19}$$

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

$$E_{t-1}[\exp(-u\varepsilon_t)] \equiv \exp(\Psi_t(u)). \tag{20}$$

The mean correction factor κ_t in equation (8) thus can be viewed as $\Psi_t(-1)$. The authors then demonstrate the existence of an equivalent martingale measure and show that:

$$\sigma_t^{*2} \approx \sigma_t^2 - skew_t \sigma_t^3 \nu_t + \frac{kurt_t}{2} \sigma_t^4 \nu_t^2 \tag{21}$$

where ν_t is an approximation of the modified Sharpe ratio:

$$\nu_t \approx \frac{\mu_t - r_t}{\sigma_t^2} + \frac{1}{2} - \frac{\kappa_t}{\sigma_t^2} \tag{22}$$

Therefore, with a negative skewness and positive excess kurtosis, the risk-neutral conditional variance is greater than the conditional variance under the physical measure[5]. In this paper, 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.

Moreover, inspired by Barone-Adesi et al. (2008), we do not specify the change of probability measure from $\mathbb{P}$ to $\mathbb{Q}$. Instead, a new set of risk-neutral parameters are calibrated by using information on the CBOE volatility indices directly.[6] 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.

1.5 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 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 (13) with 3,500 historical returns.[7] Thus, the return innovations $\{\hat{z}_t\}$ are acquired.

2. Under the risk-neutral measure, a daily variance series is simulated for the next 6 months using the variance dynamics of equation (17). 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[8]. The conditional variance of the following 6 months $\{\sigma^{*2}_{t+1}, \sigma^{*2}_{t+2}..., \sigma^{*2}_{t+126}\}$[9] are 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: $E^Q_t[\sigma^2_{t+i}] = \frac{1}{N}\sum_{n=1}^{N} \sigma^{*(n)}_{t+i}$, where $\sigma^{*(n)}_{t+i}$ 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 paper, we use $N = 50,000$ paths[10].

4. According to the definition of VIX and equation (7), the GARCH model implied VIX (model VIX, hereafter) under the trading day count convention can be calculated as:

$$VIX^{model}_t = 100 * \sqrt{\frac{252}{22} * \sum_{i=1}^{22} E^Q_t[\sigma^{*2}_{t+i}]} \tag{23}$$

Similarly:

$$VIX9D^{model}_t = 100 * \sqrt{\frac{252}{7} * \sum_{i=1}^{7} E^Q_t[\sigma^{*2}_{t+i}]} \tag{24}$$

$$VIX3M^{model}_t = 100 * \sqrt{\frac{252}{63} * \sum_{i=1}^{63} E^Q_t[\sigma^{*2}_{t+i}]} \tag{25}$$

$$VIX6M^{model}_t = 100 * \sqrt{\frac{252}{126} * \sum_{i=1}^{126} E^Q_t[\sigma^{*2}_{t+i}]} \tag{26}$$

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

$$\sqrt{\sum_{k=1}^{4} \left[w_k * \left(VI^{(k)market} - VI^{(k)model} \right)^2 \right]} \tag{27}$$

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.

1.6 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:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{4}\left[w_k * |VI_i^{(k)market} - VI_i^{(k)model}|\right] \tag{28}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{4}\left[w_k * \left(VI_i^{(k)market} - VI_i^{(k)model}\right)^2\right]} \tag{29}$$

where $w_k = 0.25$ is the weight of each index assuming equal weighting, N 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 (2011) 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 $w_k = 0.25$):

$$QLIKE = \frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{4}\left[w_k * \left(\frac{{VI_i^{(k)market}}^2}{{VI_i^{(k)model}}^2} - \log(\frac{{VI_i^{(k)market}}^2}{{VI_i^{(k)model}}^2}) - 1\right)\right]. \tag{30}$$

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 2.2 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:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(VIX_i^{model} - VIX_i^{market}\right)^2 \tag{31}$$

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:

$$DM = \frac{\bar{d}}{\sqrt{2\pi \hat{f}_d(0)/T}} \tag{32}$$

has an asymptotic standard normal distribution under the null hypothesis, where $\bar{d}$ 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 of $\sqrt{T}\hat{d}$. In this paper, 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 $\max_{k=1,...,m} E[d_{k,t}] \leq 0$. Then the statistic for the SPA test is calculated as:

$$T^{SPA} = \max_{k=1,...,m} \frac{n^{1/2}\bar{d}_k}{\hat{\omega}_k} \tag{33}$$

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 $\omega_k^2 = var(n^{1/2}\bar{d}_k)$. 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 Empirical analysis

2.1 Data

The CBOE volatility indices used in this paper 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.[11] 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.[12]

Figure 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.

2.2 In-sample model comparison

In this section, we carry out the estimation of the different GARCH models using different methods described in Section 1. 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 methods. Table 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, 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 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 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, which shows the out-of-sample VIX forecasts for different models.

2.3 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 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 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 different methods have a statistically significant difference, we also present the values of the DM test statistics in Panel A of Table

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. Panel A of Table 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 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 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, 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.[13]

2.4 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 weeks' estimation (with weekly re-estimations), the running time to calibrate once based on 373,377 option prices 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.

3 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 (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.

3.1 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 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, 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, 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 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.

3.2 Alternative weights used in the optimisation function

In Section 2, we assume each volatility index has the same weight in the optimisation function of equation (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 (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 (27), but with modified 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 (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 6 and Table 7 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 and Table 3.

3.3 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 (28) and (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 (28) and (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 3.2, 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 6 for in-sample comparison, and in Table 7 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.

3.4 Results based on three indices only

As mentioned in Section 2.1, 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 3 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 8 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, 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.

4 Conclusions

In this study, 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.

References

Barone-Adesi, G., R. F. Engle, and L. Mancini, 2008. A GARCH Option Pricing Model with Filtered Historical Simulation. *Review of Financial Studies* 21(3): 1223–1258.

Bekaert, G. and M. Hoerova, 2014. The VIX, the variance premium and stock market volatility. *Journal of Econometrics* 183(2): 181–192.

Bollerslev, T., 1986. Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31(3): 307–327.

Bollerslev, T., 1987. A Conditionally Heteroskedastic Time Series Model for Speculative Prices and Rates of Return. *Review of Economics and Statistics* 69(3): 542–547.

Bollerslev, T., M. Gibson, and H. Zhou, 2011. Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *Journal of Econometrics* 160(1): 235 – 245. Realized Volatility.

Bollerslev, T. and H. O. Mikkelsen, 1996. Modeling and Pricing Long Memory in Stock Market Volatility. *Journal of Econometrics* 73(1): 151–184.

Bollerslev, T. and J. M. Wooldridge, 1992. Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances. *Econometric Reviews* 11(2): 143–172.

Byun, S. J. and B. Min, 2013. Conditional Volatility and the GARCH Option Pricing Model with Non-normal Innovations. *Journal of Futures Markets* 33(1): 1–28.

Carr, P. and D. Madan, 1998. Towards a Theory of Volatility Trading. In Robert Jarrow (Ed.). *Volatility Estimation Techniques for Pricing Derivatives* 417–427.

Carr, P. and L. Wu, 2009. Variance risk premiums. *Review of Financial Studies* 22(3): 1311–1341.

Christoffersen, P., R. Elkamhi, B. Feunou, and K. Jacobs, 2009. Option Valuation with Conditional Heteroskedasticity and Nonnormality. *Review of Financial Studies* 23(5): 2139–2183.

Christoffersen, P., S. Heston, and K. Jacobs, 2006. Option Valuation with Conditional

Skewness. *Journal of Econometrics* 131(1): 253–284.

Christoffersen, P. and K. Jacobs, 2004. Which GARCH Model for Option Valuation? *Management Science* 50(9): 1204–1221.

Christoffersen, P., K. Jacobs, and C. Ornthanalai, 2013. GARCH Option Valuation: Theory and Evidence. *Journal of Derivatives* 21(2): 8–41.

Demeterfi, K., E. Derman, M. Kamal, and J. Zou, 1999. A Guide to Volatility and Variance Swaps. *Journal of Derivatives* 6(4): 9–32.

Diebold, F. and R. Mariano, 1995. Comparing Predictive Accuracy. *Journal of Business & Economic Statistics* 13(3): 253–63.

Duan, J.-C., 1995. The GARCH Option Pricing Model. *Mathematical Finance* 5(1): 13–32.

Duan, J.-C. and C.-Y. Yeh, 2010. Jump and Volatility Risk Premiums Implied by VIX. *Journal of Economic Dynamics and Control* 34(11): 2232–2244.

Dumas, B., J. Fleming, and R. E. Whaley, 1998. Implied Volatility Functions: Empirical Tests. *Journal of Finance* 53(6): 2059–2106.

Engle, R. F. and V. K. Ng, 1993. Measuring and Testing the Impact of News on Volatility. *Journal of Finance* 48(5): 1749–1778.

Giacomini, R. and H. White, 2006. Tests of Conditional Predictive Ability. *Econometrica* 74(6): 1545–1578.

Glosten, L. R., R. Jagannathan, and D. E. Runkle, 1993. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *Journal of Finance* 48(5): 1779–1801.

Hansen, P. R., 2005. A Test for Superior Predictive Ability. *Journal of Business & Economic Statistics* 23(4): 365–380.

Hao, J. and J. E. Zhang, 2013. GARCH Option Pricing Models, the CBOE VIX, and Variance Risk Premium. *Journal of Financial Econometrics* 11(3): 556–580.

Heston, S. L. and S. Nandi, 2000. A Closed-Form GARCH Option Valuation Model. *Review of Financial Studies* 13(3): 585–625.

Huang, Z., C. Tong, and T. Wang, 2019. VIX Term Structure and VIX Futures Pricing with Realized Volatility. *Journal of Futures Markets* 39(1): 72–93.

Kambouroudis, D. S. and D. G. McMillan, 2016. Does VIX or Volume Improve GARCH Volatility Forecasts? *Applied Economics* 48(13): 1210–1228.

Kanniainen, J., B. Lin, and H. Yang, 2014. Estimating and Using GARCH Models with VIX Data for Option Valuation. *Journal of Banking & Finance* 43: 200–211.

Liu, Q., S. Guo, and G. Qiao, 2015. VIX Forecasting and Variance Risk Premium: A New GARCH Approach. *North American Journal of Economics and Finance* 34: 314–322.

Majewski, A. A., G. Bormetti, and F. Corsi, 2015. Smile from the Past: A General Option Pricing Framework with Multiple Volatility and Leverage Components. *Journal of Econometrics* 187(2): 521–531.

Mandelbrot, B., 1963. The Variation of Certain Speculative Prices. *Journal of Business* 36(4): 394–419.

Mencía, J. and E. Sentana, 2013. Valuation of VIX derivatives. *Journal of Financial Economics* 108(2): 367 – 391.

Nelson, D. B., 1991. Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica* 59(2): 347–370.

Patton, A. J., 2011. Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics* 160(1): 246 – 256.

Stentoft, L., 2008. American Option Pricing Using GARCH Models and the Normal Inverse Gaussian Distribution. *Journal of Financial Econometrics* 6(4): 540–582.

Wang, T., Y. Shen, Y. Jiang, and Z. Huang, 2017. Pricing the CBOE VIX Futures with the Heston-Nandi GARCH Model. *Journal of Futures Markets* 37(7): 641–659.

Whaley, R. E., 1993. Derivatives on Market Volatility: Hedging Tools Long Overdue. *Journal of Derivatives* 1(1): 71–84.

Whaley, R. E., 2009. Understanding the VIX. *Journal of Portfolio Management* 35(3): 98–105.

Notes

[1]More details about the relationship between CBOE VIX and the daily conditional variance are discussed in Section 1.3.

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

[3]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.

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

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

[6]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).

[7]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.

[8]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.

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

[10]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.

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

[12]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).

[13]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 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.

Table 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 (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.

Table 3: **Out-of-sample pricing errors**

Year	Model	FHS-VI MAE	FHS-VI RMSE	FHS-VI MAE%	FHS-VI RMSE%	FHS-VI QLIKE	Normal-VIX MAE	Normal-VIX RMSE	Normal-VIX MAE%	Normal-VIX RMSE%	Normal-VIX 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 (30). Numbers in bold are the lowest values among different models for a given year.

Table 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.

Table 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.

Table 6: **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 (30). Numbers in bold are the lowest values among different models for a given year.

Table 7: **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 (30). Numbers in bold are the lowest values among different models for a given year.

Table 8: **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.

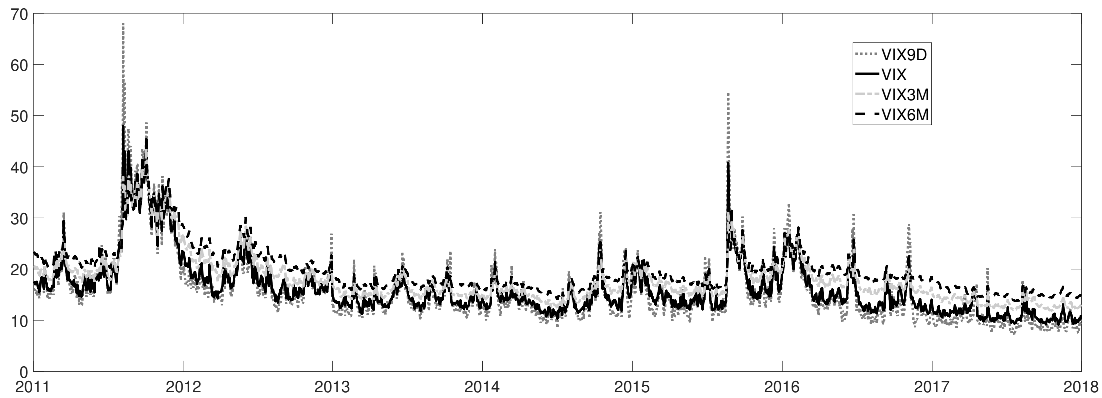
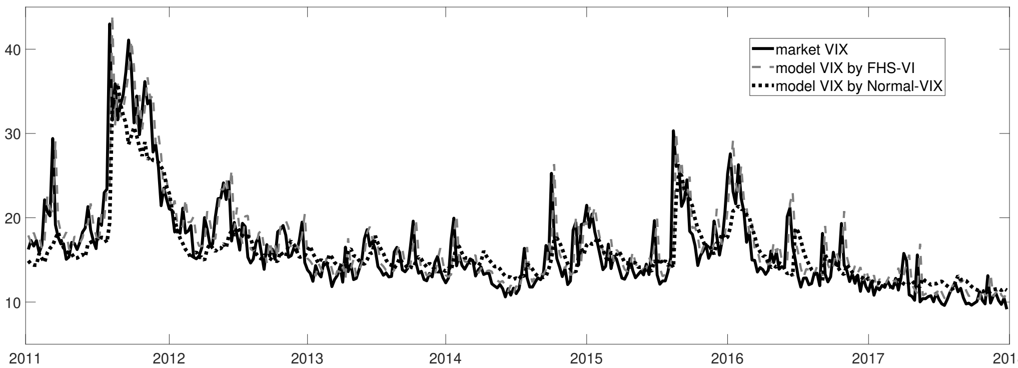
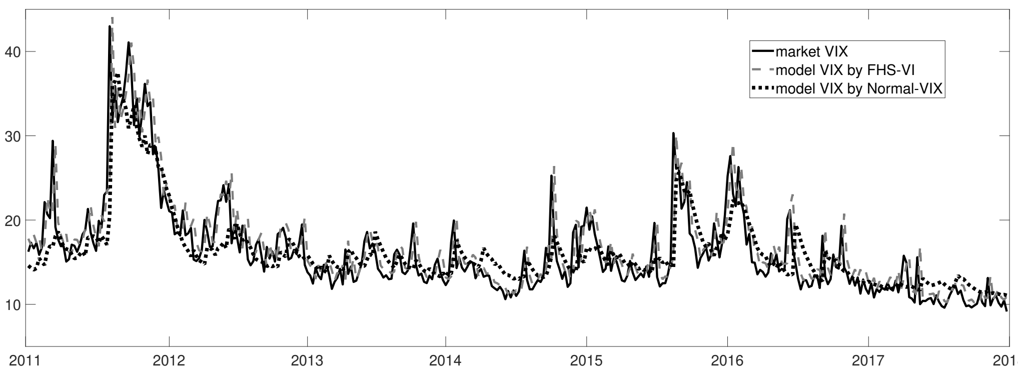


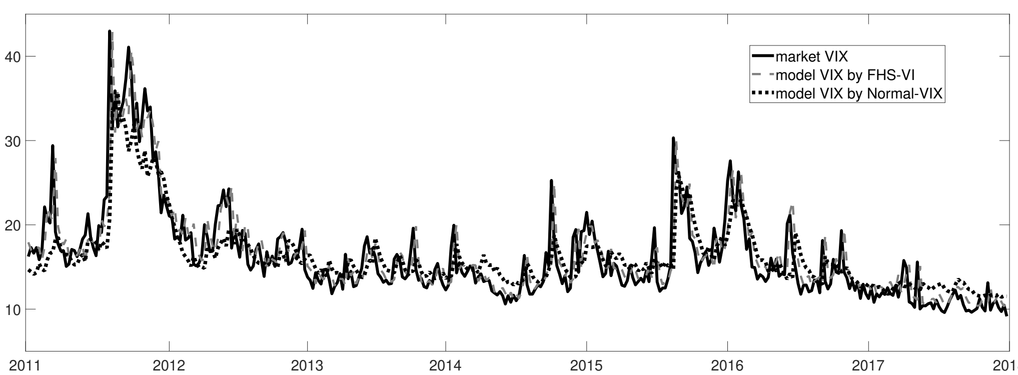
Figure 1: The dynamics of the CBOE volatility indices between 03 January, 2011 and 29 December, 2017



(a) GARCH



(b) GJR



(c) NAGARCH

Figure 2: Out-of-sample comparison of the model VIX and the CBOE VIX

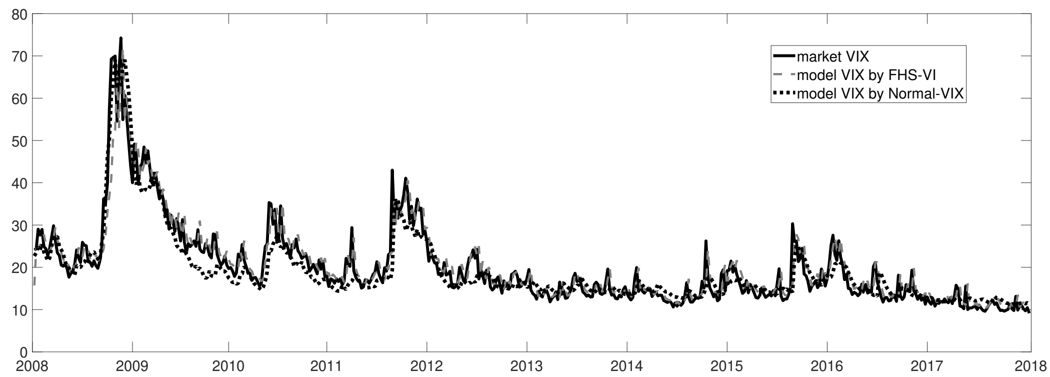


Figure 3: Out-of-sample comparison of VIX forecasts obtained using Normal-VIX, FHS-VI based on three indices, assuming GARCH, and CBOE VIX