Quantitative Finance

ISSN: 1469-7688 (Print) 1469-7696 (Online) Journal homepage: www.tandfonline.com/journals/rquf20

Options-driven volatility forecasting

Nikolas Michael, Mihai Cucuringu & Sam Howison

To cite this article: Nikolas Michael, Mihai Cucuringu & Sam Howison (2025) Options-driven volatility forecasting, Quantitative Finance, 25:3, 443-470, DOI: 10.1080/14697688.2025.2454623

To link to this article: https://doi.org/10.1080/14697688.2025.2454623

Published online: 17 Feb 2025.

Quantitative Finance, 2025
Vol. 25, No. 3, 443–470, https://doi.org/10.1080/14697688.2025.2454623

Options-driven volatility forecasting

NIKOLAS MICHAEL*†, MIHAI CUCURINGU†‡§ and SAM HOWISON‡

†Department of Statistics, University of Oxford, 24-29 St Giles', Oxford OX1 3LB, UK
‡Mathematical Institute, University of Oxford, Andrew Wiles Building, Woodstock Rd, Oxford OX2 6GG, UK
§The Alan Turing Institute, John Dodson House, 96 Euston Rd, London NW1 2DB, UK

(*Received 25 March 2024; accepted 9 January 2025*)

We augment the Heterogeneous Autoregressive Regression model for forecasting realized volatility, using various measurements for the daily, weekly, and monthly volatilities, in addition to other predictive features. The main focus is on novel methods for extracting volatility estimators using option price data. Firstly, we provide a dimensionality reduction method for implied volatility surfaces built under the Black–Scholes model, whereby we combine simple row-wise and column-wise decompositions of the implied volatility surface with principal component analysis. Secondly, we provide a method for extracting the implied volatility under the Heston and Bates models. This is achieved by a calibration of these models while assuming that some of the model parameters remain constant. We demonstrate that these augmentations result in improved daily forecasts for realized volatility in a selection of different stocks. These volatility forecasts can also be utilized to further increase predictive performance for the realized volatility of other instruments, and can be combined for forecasting VIX.

Keywords: Volatility forecasting; Options; Heston model; HAR model

1. Introduction

Forecasting volatility. Volatility forecasting is one of the most well-studied problems in finance and is considered an essential tool for various market participants, such as market makers. Notable applications of volatility forecasting include Dvorsky *et al.* (2021), who show how volatility perception can affect risk management, Kim *et al.* (2021), who incorporate volatility forecasts for bond yield forecasting, and İltüzer (2022), who use it for option pricing. Volatility forecasts can also be utilized directly as an alternative form of speculation, either via the use of option straddles or by other future strategies (see Kakushadze and Serur 2018 for various examples).

Acknowledging the importance of volatility in financial markets, the Chicago Board Options Exchange (CBOE) introduced the volatility index (VIX) in 1993, which tracks the volatility of the S&P 500 index. With recent turbulent periods, interest in volatility has been as high as ever. Notably, VIX attained its highest value yet on 16 March 2020, with the long-term average generally considered to be below 20.

On one hand, the non-stationarity, non-linearity, and low signal-to-error nature of volatility render its forecasting task quite challenging. On the other hand, volatility is generally regarded as much easier to forecast compared with future price returns (Ghysels *et al.* 1996, Andersen and Bollerslev 1998). This is partly due to volatility clustering, a phenomenon first noted by Mandelbrot (1963), where changes in price at a given time are likely to be followed by further price changes of a similar magnitude. Another important property of volatility is its path dependency, with Guyon and Lekeufack (2023) recently providing a simple yet high-performing model based on the weighted average of past returns and squared returns.

Historically, early predictive models for volatility were based on simple historical smoothing and averaging methods, albeit with rather limited success. An influential model was the Autoregressive Conditional Heteroskedasticity (ARCH) model by Engle (1982), which accounts for volatility clustering. This model has resulted in a stream of extensions and generalizations, starting with the Generalized Autoregressive Conditional Heteroskedasticity (GARCH) by Bollerslev (1986).

The Heterogeneous Autoregressive (HAR) model, by Corsi (2009), is a simple additive model that uses the daily, weekly and monthly volatilities as the only features of an OLS model. The motivation for HAR was twofold. Firstly, the Heterogeneous Market Hypothesis (HMH) states that the market is composed of different types of investors. In terms of their difference in trading frequency, traders can be approximately

*Corresponding author. Email: nikolas.michael@kellogg.ox.ac.uk

separated into three categories; intraday or high-frequency traders, investors who rebalance their portfolios weekly, and long-term investors who update their positions at a monthly frequency or longer. Investors with differing trading frequencies possibly also perceive and influence volatility at differing scales. Secondly, the HAR model appears to capture some properties of the return distributions as well, including the heavier-than-Gaussian tails and apparent long-memory structure in volatility.

An extension of the HAR model was proposed by Bollerslev *et al.* (2016), the HARQ model, including terms for the realized *quarticity* (fourth power variation), showing improvement in forecasting performance. Another important phenomenon is the leverage effect (Fischer 1976), whereby a negative return leads to increased volatility due to an increase in leverage. This results in the well-known negative correlation of returns and subsequent return volatility, which can further be insightful for forecasting volatility. Various extensions to ARCH and GARCH that account for this phenomenon have also been introduced, a review of which can be found in Będowska-Sójka and Kliber (2021).

In recent decades, with the rapid rise of machine learning, there has been a turn away from parametric models for the volatility forecasting task, as many studies, such as Bucci (2020), Christensen *et al.* (2022), Audrino and Knaus (2016), Audrino *et al.* (2020), Mittnik *et al.* (2015), and Luong and Dokuchaev (2018), have shown that black-box methods from the machine learning literature can match the performance or even outperform the classical volatility models. In particular, a plethora of deep learning methods have been proposed. Examples include Liu (2019) who use Long-Short Term Memory networks, and Chen *et al.* (2023) who propose a Hybrid Deep Neural Network by utilizing a joint temporal embedding of volatility and price. Reisenhofer *et al.* (2022) formulated HARnet, and Moreno-Pino and Zohren (2022) formulated DeepVol, both models with a dilated convolutional network structure.

Ensembles of various volatility estimators. Combining multiple estimators is a popular way of reducing the effects of measurement errors. In the context of volatility, which itself is not an observable quantity, this is a widely studied topic. Naimoli *et al.* (2022) use various volatility formulas related to the GARCH family of models to show an improvement in forecasting tail risk, Patton and Sheppard (2009) show that a weighted average of different volatility estimators outperforms any single estimator, when used as the input feature of a HAR model. Extreme range estimators have been shown to boost forecasting performance by Korkusuz *et al.* (2023).

Commonality in volatility. Similar to the commonality in equity returns, volatilities can also be separated into market and idiosyncratic volatilities. A recent study is Zhang *et al.* (2022), who used the intraday commonality of volatilities to show improved forecasting performance. Earlier, Herskovic *et al.* (2016) showed how stocks that are less correlated with the volatility of the market tend to outperform other stocks. Common volatility applications also include Fang *et al.* (2020), who used VIX as a proxy for commonality.

Implied Volatility. A well-researched topic is the role of implied volatility in forecasting of realized volatility. Various studies in the past have studied this question in detail (Canina and Figlewski 1993, Jorion 1995, Lamoureux and Lastrapes 2015). While it is a common belief that implied volatility leads realized volatility, practice shows that when forecasting future volatility, predictability remains low, with Jorion (1995) noting that implied volatility underpredicts future volatility. However, more recent and complex methodologies provide evidence in the opposite direction, such as Dai *et al.* (2020), who show a significant Granger causality from option implied volatility to stock realized volatility and Pérez-Rodríguez (2020), who use a bivariate copula model to conclude that implied volatility is a better predictor of realized volatility than vice versa. Maghyereh *et al.* (2022) demonstrate the increased predictive power of implied volatility during the pandemic.

The main reason behind the failure of implied volatilities is the high dimensionality of such a data set, which naturally points to dimensionality reduction as the remedy. An early study on the dimensionality reduction of implied volatility surfaces is that of Badshah (2009), who rely on Principal Component Analysis (PCA). Cont and da Fonseca (2002) use a Karhunen–Loeve decomposition to show how IVS surfaces can be empirically modeled using only 2-3 functional components. Dobi (2014) classify equity options in the U.S. markets based on their level of systemic and idiosyncratic risk, relying on the fluctuations of their implied volatility surfaces. They propose a dimensionality reduction model based on PCA and random matrix theory.† Avellaneda *et al.* (2020) construct an eigenportfolio based on the volatility returns, and show how the top eigenvalue is correlated to a weighting scheme based on Vega and open interest. Furthermore, they show that using a multilinear SVD on the order-4 tensor of implied volatility surfaces, is better able to track the open interest-Vega portfolio compared to the flattened PCA of returns.

Alternatively, custom features are extracted from the implied volatility surface, which usually depend on specific delta-maturity pairs. Relevant literature include the concepts of implied volatility skew, defined by Xing *et al.* (2010) as the difference between the implied volatilities of an ATM call option with an OTM put option, both with 30-day maturities. Other popular features include the CIV and PIV. The CIV (PIV) corresponds to the implied volatility of an ATM call (put) option. Such features are used by Han and Liu (2019) to predict jumps in equity returns.

A popular approach is proposed by Dumas *et al.* (2001), who use a deterministic volatility function approach to extract various features. More sophisticated approaches include Fukasawa (2011) who extract the volatility skew and term structure using an expansion formula for the Black–Scholes volatility. Lastly, Aït-Sahalia *et al.* (2021) extract a closed form for implied volatility surface features such as curvature and term-structure under a stochastic volatility model with jumps.

† In particular, the Marchenko-Pastur distribution is used to distinguish between informative and uninformative principal components.

Volatility-volume interplay. The interplay between volatility and volume has also been widely studied, with the mixture of distribution hypothesis by Economics (1976) and the sequential information arrival hypothesis by Clark (1973), giving two different interpretations of how information embedded in trading volume is informative for future changes in price. Recent studies into this interplay include Liu *et al.* (2023) who use an empirical mode decomposition method on trading volume to show improvements in volatility forecasting. There are also a number of studies that investigate this interplay for the cryptocurrency market, such as Bouri *et al.* (2019) who look into the Granger causality from trading volume to volatility.

Main contributions. In this paper, we tackle the task of forecasting daily volatility. Previously, Visser (2008) and Patton and Sheppard (2009) have both explored how to combine multiple volatility proxies into a single estimator, aiming to minimize the measurement error and maximize the predictive power of a HAR or GARCH type of model. We also base our model on the HAR model, but instead of substituting the realized volatility for a single predictor, we augment the model with additional daily, weekly and monthly proxies of volatility. To that end, we make use of historical volatility proxies, which we calculate for the monthly and weekly horizons, as well as several proxies for daily volatility. We compare and contrast the different models to the original HAR.

As previous attempts at using the implied volatility to forecast realized volatility have mostly proved to be inconclusive, we consider a dimensionality reduction approach based on an implied volatility surface data set. Instead of a structurally agnostic method, we consider a simple decomposition based on the maturity and delta characteristics of the surface, which effectively reduces dimensionality, in a easily interpretable manner. Furthermore, we translate a similar idea to the implied volatility using the Heston model, whereby we use a calibration process to recover the underlying parameters of Heston and Bates models. One of those parameters, namely the current volatility, is a parallel to the implied volatility for the Black–Scholes model, albeit one that captures the price process more accurately (without the need of an implied volatility surface), providing an alternative way of utilizing the information embedded in the option prices. Lastly, similar to Zhang *et al.* (2022), we also consider a cross-sectional model, using information from different instruments as predictors of future volatility. We build this model based on the forecasts of the single assets models, which allow for a further boost in the predictive performance of the model.

Paper outline. The rest of this paper is structured as follows; in section 2, we define the realized volatility and various measures of historical volatility. In section 3, we provide an overview of the implied volatility surface data and the dimensionality reduction methodology used to extract volatility estimators. In section 4, we explain the methodology of extracting the option implied volatility using the Heston and Bates models. In section 5, we set up the forecasting problem, and explain all candidate models. The empirical results are presented and analyzed in section 6. Lastly, we draw conclusions in section 7, and provide future avenues of research.

2. Volatility estimators

2.1. *Realized volatility*

Denote by $(S_t : t \geq 0)$, $(\mu_t : t \geq 0)$ and $(\sigma_t : t \geq 0)$ the price, mean, and instantaneous volatility processes corresponding to an underlying instrument. Under the stochastic volatility family of models, the price process is driven by

$$\frac{\mathrm{d}S_t}{S_t} = \mu_t \mathrm{d}t + \sigma_t \mathrm{d}B_t, \tag{1}$$

where $\{B_t : t \geq 0\}$ denotes a standard Brownian motion process. The integrated variance is the theoretical value of the total variability occurred over the interval $[t_1, t_2]$, calculated as

$$\mathrm{IV}_{(t_1,t_2)} = \int_{t_1}^{t_2} \sigma_t^2 \, \mathrm{d}t. \tag{2}$$

As this is a theoretical quantity, there exist a number of different estimators to approximate it. By far the most common quantity used to approximately measure the integrated variance is the realized variance, which is given for the interval $[t_1, t_2]$ and a sample distance of δt, and is calculated as

$$\mathrm{RV}_{(t_1,t_2,\delta t)} = \sum_{t=t_1+\delta t}^{t_2} \log\left(\frac{S_t}{S_{t-\delta t}}\right)^2. \tag{3}$$

It has been shown by Barndorff-Nielsen and Shephard (2002), that assuming no drift ($\mu_t = 0$) the realized variance follows

$$\mathrm{RV}_{(t_1,t_2,\delta t)} \sim N(V_{(t_1,t_2,\delta t)}, 2\mathrm{IQ}_{(t_1,t_2)}\, \delta t), \tag{4}$$

where

$$\mathrm{IQ}_{(t_1,t_2)} = \int_{t_1}^{t_2} \sigma_t^4 \, \mathrm{d}t, \tag{5}$$

is the integrated quarticity. Therefore, realized variance converges in probability to integrated variance

$$\operatorname*{plim}_{\delta t \to 0} \mathrm{RV}_{(t_1,t_2,\delta t)} = \mathrm{IV}_{(t_1,t_2)}. \tag{6}$$

Another very popular estimator for integrated variance is the Bipower Variation (BPV) introduced by Barndorff-Nielsen and Shephard (2004), and calculated as

$$\mathrm{BPV}_{(t_1,t_2,\delta t)} = \sum_{t=t_1+\delta t}^{t_2-\delta t} \left|\log\left(\frac{S_t}{S_{t-\delta t}}\right)\right| \left|\log\left(\frac{S_{t+\delta t}}{S_{tt}}\right)\right|. \tag{7}$$

Compared with realized variance, BPV is more robust to price jumps and is thus used as an additional feature as a proxy for integrated variance.

Our goal for this paper is to forecast the daily realized volatility (t_1 and t_2 corresponding to market open and close, respectively), based on 1-minute intraday returns ($\delta t = 1$ minute). Denote by $S_{d,1}, \ldots, S_{d,391}$ the intraday mid-prices at 9:30, 9:31, …, 16:30 respectively for day d. The 1-minute realized variance is then calculated as

$$\mathrm{RV}_d = \sum_{j=2}^{391} \log\left(\frac{S_{d,j}}{S_{d,j-1}}\right)^2. \tag{8}$$

2.2. Heterogeneous autoregressive (HAR) model

One of the most popular volatility forecasting models is HAR, an additive model that relies on the previous daily, weekly and monthly variances as predictors for the next day realized variance. Initially formulated by Corsi (2009), it takes the following form

$$\mathrm{RV}_d = \beta_0 + \beta_1 \mathrm{RV}_{d-1} + \beta_2 \left(\frac{1}{5} \sum_{j=1}^{5} \mathrm{RV}_{d-j} \right) + \beta_3 \left(\frac{1}{22} \sum_{j=1}^{22} \mathrm{RV}_{d-j} \right). \tag{9}$$

In contrast to other models capturing the long-memory property, HAR does not make any strong assumptions for the underlying price process. Despite its simplicity, HAR has matched or even outperformed more complex models, making it a standard benchmark for volatility forecasting.

The use of a simple autoregressive model for realized volatility, such as HAR, can entail a misspecification when assuming equation (4), as it can cause an attenuation bias, that arises from the presence of measurement errors in volatility. Instead, Bollerslev (1986) considers evolving coefficients which are linearly dependent on the square root of realized quarticity

$$\mathrm{RQ}_{(t_1,t_2,\delta t)} = \frac{t_2 - t_1}{3\delta t} \sum_{t=t_1+\delta t}^{t_2} \log \left(\frac{S_t}{S_{t-\delta t}} \right)^4, \tag{10}$$

which is a consistent estimate of the integrated quarticity in equation (2). In particular, the HARQ model of Bollerslev *et al.* (2016), is formulated as

$$\mathrm{RV}_d = \beta_0 + \left(\beta_1 + \beta_{1,0} \sqrt{\mathrm{RQ}_{d-1}} \right) \mathrm{RV}_{d-1} + \beta_2 \left(\frac{1}{5} \sum_{j=1}^{5} \mathrm{RV}_{d-j} \right) + \beta_3 \left(\frac{1}{22} \sum_{j=1}^{22} \mathrm{RV}_{d-j} \right), \tag{11}$$

and is shown to achieve improvements over the HAR model. While β_2 and β_3 can also be modeled as being non-constant and dependent on realized quarticity in the same manner as β_1, Corsi (2009) shows that the estimation errors from such misspecifications are on a much smaller magnitude than for β_1, rendering equation (11) a sufficiently good model. Our proposed models are extensions of the HAR model. To incorporate additional features representing the weekly and monthly variances, we consider a number of historical volatility estimators, as shown in section 2.3. Furthermore, we incorporate additional variance estimates at the daily level, including the BPV, and volatility proxies calibrated from option data.

2.3. Historical volatility estimators

As features, we consider a number of additional volatility proxies, which are based on the historical volatility of the past L days. We denote the opening, closing, highest and lowest prices for day d as follows

$$\begin{aligned} O_d &= S_{d,1} \\ C_d &= S_{d,391} \\ H_d &= \max\left(S_{d,1}, \ldots, S_{d,391}\right) \\ L_d &= \min\left(S_{d,1}, \ldots, S_{d,391}\right) \end{aligned} \tag{12}$$

One of the most basic estimators used is the standard historical close-close volatility (HV) estimator, calculated as

$$\mathrm{HV}_d^{(L)} = \sqrt{\frac{1}{L-1} \sum_{l=0}^{L-1} \log \left(\frac{C_{d-l}}{C_{d-1-l}} \right)^2}. \tag{13}$$

As this only takes into account the closing prices, and omits the intraday variation, volatility is severely underestimated. To this end, we additionally consider some of the so-called *advanced volatility estimators*, which also take into account the opening, high and low values within a day. One of the earliest advanced volatility estimators was the high-low estimator of Parkinson (HLP) formulated in Parkinson (1980), which incorporates the high and low prices as follows

$$\mathrm{HLP}_d^{(L)} = \sqrt{\frac{\sum_{l=0}^{L-1} \left[\log \left(\frac{H_{d-l}}{L_{d-l}} \right) \right]^2}{4 \log(2) L}}. \tag{14}$$

The Garman–Klass (GK) volatility estimator (Garman and Klass 1980) was formulated as an extension of the Parkinson estimator that also includes the opening and closing prices. It is calculated as

$$\mathrm{GK}_d^{(L)} = \sqrt{\frac{1}{L} \sum_{l=0}^{L-1} \left\{ \frac{1}{2} \left[\log \left(\frac{H_{d-l}}{L_{d-l}} \right) \right]^2 - (2\log(2) - 1) \left[\log \left(\frac{C_{d-l}}{O_{d-l}} \right) \right]^2 \right\}}. \tag{15}$$

Both HV, HLP and GK estimators all assume zero drift in the returns process. An extension which is able to accommodate for non-zero drifts is the volatility estimator of Rogers–Satchell (RS) (Rogers and Satchell 1991), formulated as

$$\mathrm{RS}_d^{(L)} = \sqrt{\frac{1}{L} \sum_{l=0}^{L-1} \left[\log \left(\frac{H_{d-l}}{C_{d-l}} \right) \log \left(\frac{H_{d-l}}{O_{d-l}} \right) + \log \left(\frac{L_{d-l}}{C_{d-l}} \right) \log \left(\frac{L_{d-l}}{O_{d-l}} \right) \right]}. \tag{16}$$

Note that none of the estimators discussed so far take into account the overnight jumps, which leads them to underestimate the real volatility. The Yang–Zhang extension of the GK model (GKYZ), does take into account the overnight jumps

Table 1. This table shows the efficiencies as always by equation (22), for each of the volatility estimators.

Estimator	Efficiency
Standard close-close volatility (HV)	1
Parkinson (HLP)	5.2
Garman–Klass (GK)	7.4
Rogers–Satchell (RS)	8
Yang–Zhang extension of Garman–Klass (GKYZ)	8
Yang–Zhang (YZ)	14

Note: These values are taken by Yang and Zhang (2000) and Chou *et al.* (2010).

and takes the form

$$\mathrm{GKYZ}_d^{(L)} = \sqrt{\frac{1}{L}\sum_{l=0}^{L-1}\left\{\frac{1}{2}\left[\log\left(\frac{O_{d-l}}{C_{d-l-1}}\right)\right]^2 + \frac{1}{2}\left[\log\left(\frac{H_{d-l}}{L_{d-l}}\right)\right]^2 - (2\log(2)-1)\cdot\left[\log\left(\frac{C_{d-l}}{O_{d-l}}\right)\right]^2\right\}} \tag{17}$$

We also consider the Yang–Zhang estimator (YZ) which also includes a term for the overnight jumps by combining the RS estimator with the overnight and intraday volatilities.

$$\mathrm{YZ}_d^{(L)} = \sqrt{(1-k^{(L)})\mathrm{RS}_d^{(L)} + k^{(L)}\sigma_c^2 + \sigma_o^2}, \tag{18}$$

where

$$k^{(L)} = \frac{\alpha-1}{\alpha + \frac{L+1}{L-1}}, \quad \text{with } \alpha = 1.34, \tag{19}$$

$$\sigma_o^2 = \mathrm{Var}\left(\left\{\log\left(\frac{O_{d-(L-j)}}{C_{d-(L-1-j)}}\right)\right\}_{j=0}^{L-1}\right), \tag{20}$$

$$\sigma_c^2 = \mathrm{Var}\left(\left\{\log\left(\frac{C_{d-(L-j)}}{O_{d-(L-j)}}\right)\right\}_{j=0}^{L-1}\right). \tag{21}$$

Note that the above weights and coefficients are in most cases selected to maximize the metric of *efficiency*, computed as

$$\frac{\text{Volatility of Standard Estimator}}{\text{Volatility of Estimator}}, \tag{22}$$

where 'Volatility of Standard Estimator' refers to the volatility of the standard close-close estimator. The maximum theoretical efficiencies are given in table 1, where the Yang–Zhang estimator attains the highest efficiency. The maximum theoretical efficiency reflects the reduction in the number of data points required to achieve a level of volatility under an idealized distribution.

The last feature we consider is VIX-Fix, a volatility proxy developed by Williams (2007). This proxy has been shown to successfully replicate the VIX time series when constructed for SPX, and can be easily constructed from historical data, for any desired instrument, including other indices and stocks. It is calculated as

$$\mathrm{WVF}_d = 100\frac{\max(H_d, H_{d-1}, \ldots, H_{d-21}) - L_d}{\max(H_d, H_{d-1}, \ldots, H_{d-21})}. \tag{23}$$

2.4. *Path dependent volatility*

Contrary to stochastic volatility (SV) models which have volatility as a separate stochastic process to returns, recent empirical evidence shows how volatility depends on the path of the price return process (Zumbach 2010, Chicheportiche and Bouchaud 2014). To account for these effects, we incorporate the two features from Guyon and Lekeufack (2023)

$$R_{1,d} = \sum_{d_i \le d} \mathrm{Ker}(d - d_i; h_1, \alpha_1) r_{d_i}, \tag{24}$$

$$R_{2,d} = \sqrt{\sum_{d_i \le d} \mathrm{Ker}(d - d_i; h_2, \alpha_2) r_{d_i}^2}, \tag{25}$$

where $r_{d_i} = \log(\frac{S_{d_i,391}}{S_{d_i-1,391}})$ denotes the daily close-close return for day d_i. The kernel Ker is selected to be a time shifted power law,† i.e $\mathrm{Ker}(\tau; h, \alpha) = \frac{(\tau+h)^{-\alpha}}{\int_0^\infty (\tau+h)^{-\alpha}\,d\tau}$, where the delay helps ensure that the weights do not blow up for vanishing τ. A simple linear model using equations (24) and (25) can be shown to achieve great predictive power for forecasting VIX (Guyon and Lekeufack 2023).

3. Extracting volatility estimators from implied volatility surface data

3.1. *The Black–Scholes model*

The Black–Scholes model is a simple framework for option pricing, and is often used to extract the implied volatility from the observed prices. Under this model, the volatility of returns is assumed to be constant, denoted as σ_{BS}. Denote by r the interest rate and q the dividend yield of the underlying. Under the risk-neutral probability measure $\mathbb{Q}$, the underlying return process is then driven by the diffusion

$$\mathrm{d}S_t = (r-q)S_t\mathrm{d}t + \sigma_{\mathrm{BS}}\mathrm{d}B_t. \tag{26}$$

Consider a vanilla European call option at time t‡ with a strike price of K, that expires at time T. The payoff is given by

$$C_{K,T,\sigma_{\mathrm{BS}}}^{\text{BS-Call}}(t,S) = \mathrm{e}^{-r(T-t)}\mathbb{E}^{\mathbb{Q}}\left[(S_T-K)^+ | S_t = S\right]. \tag{27}$$

† The particular values of parameters we use, is the optimal configuration for VIX as given by Guyon and Lekeufack (2023) for the volatility of SPX-500. In particular, we use $\alpha_1 = 1.06, h_1 = 0.020, \alpha_2 = 1.60, h_2 = 0.052$ as well as a look-back window of 1000 days.

‡ Note that the value of a European call option is equal to the value of an American call option for a stock that does not pay dividends. As we work with stocks that do not pay dividends, we are able to use the formula for European options.

In a Black–Scholes world where the price process is driven by equation (27), its value is

$$C^{\text{BS-Call}}_{K,T,\sigma_{\text{BS}}}(t,S_t) = S_t e^{-q(T-t)}\Phi(d_1) - Ke^{-r(T-t)}\Phi(d_2), \quad (28)$$

where

$$d_1 = \frac{\log\left(\frac{S_t}{K}\right) + \left(r - q + \frac{\sigma^2_{\text{BS}}}{2}\right)T}{\sigma_{\text{BS}}\sqrt{T-t}}, \quad d_2 = d_1 - \sigma_{\text{BS}}\sqrt{T-t},$$

and Φ denotes the cumulative distribution function of a standard normal variable. Using the call-put parity, it can be shown that a put option with the same maturity and strike price is given by

$$C^{\text{BS-Put}}_{K,T,\sigma_{\text{BS}}}(t,S_t) = Ke^{-r(T-t)}\Phi(-d_2) - S_t e^{-q(T-t)}\Phi(-d_1). \quad (29)$$

The only unobserved variable is the volatility σ_{BS}. A value for this can be obtained by solving equation (28) or equation (29) using a simple binary search algorithm.

3.2. *Implied volatility surface*

The cost for the simplicity of the BS model is its inability to capture a time-evolving variance. Furthermore, a commonly observed phenomenon is the volatility smile, in which the implied volatility differs for different options products, both in terms of $\Delta = \partial C^{\text{BS-Call}}_{K,T}/\partial S$ and maturity $T-t$. By considering a fixed set of 10 maturity values ($\{30, 60, 91, 122, 152, 182, 273, 365, 547, 730\}$ days) and 28 values for Δ ($\{-0.75, -0.70, \ldots, -0.15, -0.10, 0.10, 0.15, \ldots, 0.70, 0.75\}$), a three dimensional object denoted as the *Implied Volatility Surface* (IVS) is formed, often visualized as a surface plot.†

An early method of constructing volatility surfaces is that of Figlewski (2010), who used a spline interpolation method. We instead use the data set by OptionMetrics (2023), which is built using a kernel interpolation technique, which we summarize in the next paragraph.

We denote the implied volatility for day d, for the option contract with Δ equal to δ and maturity equal to m as $\hat{\sigma}^{(d)}_{\text{BS},m,\delta}$. As such options are not guaranteed to exist, the implied volatilities are computed for each delta-maturity pair using an interpolation method based on a kernel smoothing algorithm. Denote by N_t the number of different option products traded for day t, with implied volatilities $(\sigma_1, \ldots, \sigma_{N_d})$, Vega's $(\mathcal{V}_1, \ldots, \mathcal{V}_{N_d})$,‡ expiration times $(T_1, \ldots, T_{N_d})$ and Δ's $(\Delta_1, \ldots, \Delta_{N_d})$. We then calculate

$$\hat{\sigma}^{(d)}_{\text{BS},m,\Delta} = \frac{\sum_{j=1}^{N_t} \mathcal{V}_j \sigma_j K(\Delta_j - \Delta, \log(\frac{m_j}{m}), \mathbb{1}_{\text{sign}(\Delta)=\text{sign}(\Delta_j)})}{\sum_{i=1}^{N_t} \mathcal{V}_i K(\Delta_i - \Delta, \log(\frac{m_i}{m}), \mathbb{1}_{\text{sign}(\Delta)=\text{sign}(\Delta_i)})}, \quad (30)$$

where

$$K(x,y,z) = \frac{1}{2\pi}\exp\left(-\left[\frac{x^2}{2h_1} + \frac{y^2}{2h_2} + \frac{z^2}{2h_3}\right]\right),$$

and the choice of $h_1 = 0.05, h_2 = 0.005, h_3 = 0.001$.

In figure 1, we plot one example of an implied volatility surface (IVS), for the symbol Triumph Group Inc (TGI) for 23-January-2013. We choose this particular instrument-date pair as it features a number of common patterns observed for IVS. Firstly, put options have a higher implied volatility compared with the corresponding call options, as is the case in this example, where implied volatilities range from 0.25 to 0.45 for put options, and from 0.23 to 0.29 for call options. This stems from the behavior of investors shielding themselves against market crashes with put options. Secondly, the lowest implied volatilities are observed for ATM options (that is with $\Delta = 0.5$ or -0.5). Typically, IVS take the usual U-shape form, while in the cases shown in figure 1, the implied volatility for put options is much higher for Δ close to 0 (OTM) and implied volatility for call options is much higher for Δ close to 1 (ITM). Lastly, there is a more subtle effect, where option products with shorter maturities have a higher implied volatility.

3.3. *Dimensionality reduction for implied volatility surfaces*

The dimensionality of implied volatility surfaces is quite high at 280. An early investigation for the relationship between implied and realized volatility is Canina and Figlewski (1993), who tests each of the implied volatility values of the surface separately, without finding consistent predictability. More recent research focused on the implied volatility of indices, includes Dai *et al.* (2020), who showed a significant Granger causality between stock market implied volatility and stock volatility, and Wei *et al.* (2020), who showed a connection between the implied volatility for the gold and silver markets to the corresponding futures in the Chinese market. To this end, Principal Component Analysis (PCA) and other dimensionality reduction methods have been used to reduce the dimensionality of the surfaces (see, e.g. Dobi 2014). Disadvantages of using a black-box dimensionality algorithm include that they are structure-agnostic and that any mispricing or faulty entries can have a disruptive effect on many of the principal components (PC). In this section, we consider an alternative method of reducing the dimensionality, by decomposing implied volatility using a simple decomposition. In particular, for each day d, we wish to extract the average level of the volatility $\hat{\sigma}^{(d)}_0$, a maturity effect for maturity m denoted as $\sigma^{(d)}_{M=m}$, and a delta effect for δ denoted as

† Note that the natural characteristics of an option product are maturity and strike price. However, Δ is used in place of the strike price as it is a normalized value, taking values in $[-1, 1]$, in contrast to the unconstrained strike price. In other words, Δ is a bounded proxy for *moneyness*, S_t/K.

‡ The Greek Vega captures the sensitivity of the option to volatility and is computed as $\mathcal{V} = \frac{\partial C^{\text{BS-Call}}_{K,T,\sigma_{\text{BS}}}}{\partial \sigma_{\text{BS}}}$.

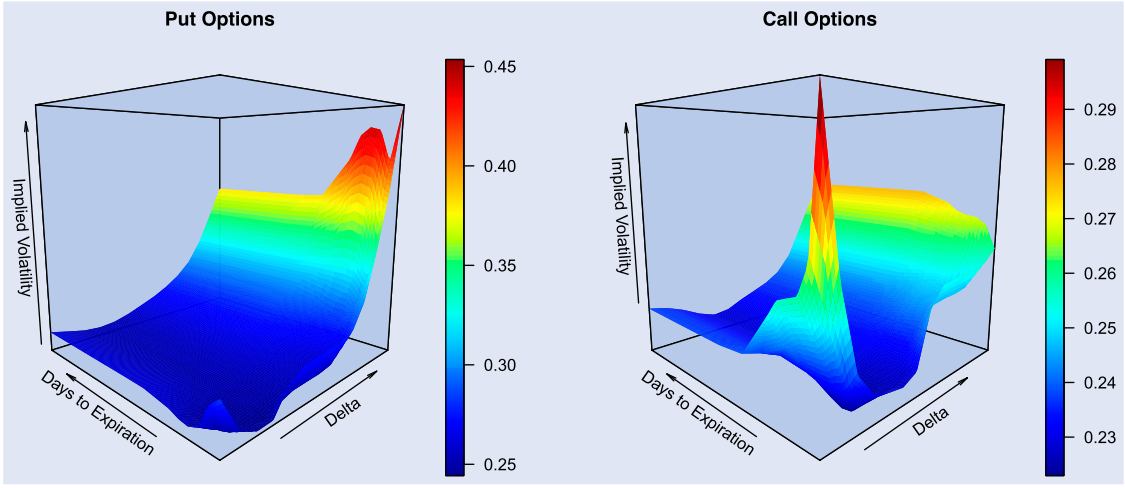

Figure 1. Example of implied volatility surface for Triumph Group Inc (TGI) - 23-January-2013. We separate the plots for call and put options.

$\sigma^{(d)}_{\Delta=\delta}$. Therefore, this decomposition takes the form

$$\hat{\sigma}^{(d)}_{\mathrm{BS},m,\delta} = \underbrace{\hat{\sigma}^{(d)}_0 + \sigma^{(d)}_{M=m} + \sigma^{(d)}_{\Delta=\delta}}_{\text{Structural Component}} + \underbrace{\epsilon^{(d)}_{m,\delta}}_{\text{Residual Component}}. \tag{31}$$

Using this decomposition, we reduce the dimensionality from 280 down to $39 + p$. We select the parameters by minimizing

$$\sum_{\delta=1}^{|\Delta|}\sum_{m=1}^{|M|}\left(\epsilon^{(d)}_{m,\delta}\right)^2. \tag{32}$$

Equation (32) has a unique solution, when we set the natural constraints

$$\sum_{\delta=1}^{|\Delta|}\sigma^{(d)}_{\Delta=\delta} = 0, \sum_{m=1}^{|M|}\sigma^{(d)}_{M=m} = 0. \tag{33}$$

We solve equation (32) by considering the Lagrangian

$$\begin{aligned}\mathcal{L} = &\sum_{\delta=1}^{|\Delta|}\sum_{m=1}^{|M|}\left(\hat{\sigma}^{(d)}_{m,\delta} - \hat{\sigma}^{(d)}_0 - \sigma^{(d)}_{\Delta=\delta} - \sigma^{(d)}_{M=m}\right)^2 \\ &+ \nu_1\sum_{m\in M}\sigma^{(d)}_{M=m} + \nu_2\sum_{\delta\in\Delta}\sigma^{(d)}_{\Delta=\delta}.\end{aligned} \tag{34}$$

The solution obtained via equation (34) is given by

$$\hat{\sigma}^{(d)}_0 = \frac{\sum_{m\in M}\sum_{\delta\in\Delta}\hat{\sigma}^{(d)}_{m,\delta}}{|\Delta||M|}, \tag{35}$$

$$\sigma^{(d)}_{\Delta=\delta} = \frac{\sum_{m\in M}\hat{\sigma}^{(d)}_{m,\delta}}{|M|} - \hat{\sigma}^{(d)}_0, \tag{36}$$

$$\sigma^{(d)}_{M=m} = \frac{\sum_{\delta\in\Delta}\hat{\sigma}^{(d)}_{m,\delta}}{|\Delta|} - \hat{\sigma}^{(d)}_0. \tag{37}$$

This decomposition of implied volatility provides a single value of the implied volatility $\hat{\sigma}^{(d)}_0$, with the rest of the features being directly interpretable, being the mean increase for each level of maturity or Δ.

4. Extracting volatility estimators under stochastic volatility models

In this section, we describe how to extract a volatility estimator using the Heston model, via calibrating its parameters. As we are only interested in tracking the current volatility and its changes, the other parameters can be fixed. We first describe the Heston model, discuss the sensitivity of option prices to its parameters, and briefly explain the calibration process. We then justify the decision to fix some of its parameters, discuss an extension to the Bates model, and lastly provide some implementation details.

4.1. Heston model

Unlike the Black–Scholes model, volatility under the Heston model is not assumed to be constant. Instead, volatility is a separate random process σ_t^2, driven by a diffusion process which is correlated with the price process S_t. In particular, the dynamics are described by the diffusions

$$\mathrm{d}S_t = \mu S_t\mathrm{d}t + S_t\sigma_t\mathrm{d}B_t, \tag{38}$$

$$\mathrm{d}\sigma_t^2 = \kappa(\bar{V} - \sigma_t^2)\mathrm{d}t + \sigma_V\sigma_t\mathrm{d}B^1_t, \tag{39}$$

where B, B^1 are standard Brownian motion processes with correlation ρ, i.e. $\mathrm{Cov}(\mathrm{d}B_t, \mathrm{d}B^1_t) = \rho$. Overall, the Heston model is governed by the following five parameters: the mean reversion speed of volatility (κ); long-term volatility average ($\bar{V}$); volatility of volatility (σ_V); correlation between the two diffusion processes (ρ); the current/initial volatility at time t ($\sigma_{t,0}$).

4.2. Sensitivity analysis of the Heston model parameters

In later sections, we explain that precise calibration of all parameters in the Heston model isn't necessary for deriving a meaningful volatility estimator. Certain parameters influence the model's behavior more than others. To identify these key drivers, we initially assess the sensitivity of option prices to each of the five parameters.

We consider the influence of each parameter for four different times to expiration ($T-t \in \{5, 15, 90, 500\}$ days) and four different strike prices ($K \in \{0.75, 0.9, 1.1, 1.25\}$) for a theoretical instrument with current spot price $S = 1$,† and setting both $r = 0$ and $q = 0$. We consider the parameter sets $\Theta_\kappa, \Theta_{\sigma_{t,0}^2}, \Theta_{\bar{V}}, \Theta_{\sigma_V}, \Theta_\rho$, each containing five candidate parameter values.

We calculate the set of sensitivities for κ as

$$\begin{aligned} I_\kappa(T,K) &= \{|C^{\text{Heston–Call}}_{T,K,(\kappa_1,\sigma_{t,0},\bar{V},\sigma_V,\rho)}(t,S_t) - C^{\text{Heston–Call}}_{T,K,(\kappa_2,\sigma_{t,0},\bar{V},\sigma_V,\rho)}(t,S_t)| : \\ &\quad (\kappa_1,\kappa_2,\sigma_{t,0},\bar{V},\sigma_V,\rho) \in \Theta_\kappa \times \Theta_\kappa \\ &\quad \times \Theta_{\sigma_{t,0}} \times \Theta_{\bar{V}} \times \Theta_{\sigma_V} \times \Theta_\rho\}, \end{aligned} \tag{40}$$

and define $I_{\sigma_{t,0}}(T,K), I_{\bar{V}}(T,K), I_{\sigma_V}(T,K), I_\rho(T,K)$ in a similar manner.

In figure 2, we plot a grid of box and whisker plots showing the set of sensitivities for each of the parameters. For low maturity options, the current volatility $\sigma_{t,0}$ is the most influential parameter, while the long term volatility $\bar{V}$ and mean reversion speed κ have a reduced significance. For higher maturities, the long term volatility $\bar{V}$ becomes the most influential parameter instead. For a maturity period of 500 days, κ also becomes a more influential parameter than $\sigma_{t,0}$. $\sigma_{t,0}$ also appears to have more influence for options that are closer to being ATM, than options with $K = 0.75$ or 1.25. Crucially, it is evident that the price primarily reacts to only three parameters: $\sigma_{t,0}, \bar{V}$, and κ. ρ and σ_V exhibit noticeably lower influence across all maturities and strike prices, with median sensitivities being close to zero for all maturity/srike price combinations.

4.3. *Full calibration of the Heston model*

Unlike the Black–Scholes model, which can be calibrated with one market price and a simple binary search, the Heston model's parameter space is five-dimensional. In particular, at any time t, the parameter vector $\theta = (\kappa, \sigma_{t,0}, \bar{V}, \sigma_V, \rho)$ needs to be determined. A semi-closed solution can be obtained for vanilla option prices (Heston 2015), with the price of a European call option being given by

$$\begin{aligned} C^{\text{Heston–Call}}_{T,K,\theta}(t,S_t) &= S_t P_1(t,S_t,T,K,\theta) \\ &\quad - \mathrm{e}^{-r(T-t)} K P_2(t,S_t,T,K,\theta), \end{aligned} \tag{41}$$

where

$$\begin{aligned} P_1(t,S_t,T,K,\theta) &= \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left(\frac{\mathrm{e}^{-\mathrm{i}u\log K}}{S_t \mathrm{i}u \mathrm{e}^{r(T-t)}} \phi^{(\text{Heston})}(\theta; u - \mathrm{i}, T-t)\right) \\ &\quad \times \mathrm{d}u, \end{aligned} \tag{42}$$

$$\begin{aligned} P_2(t,S_t,T,K,\theta) &= \frac{1}{2} + \frac{1}{\pi}\int_0^\infty \mathrm{Re}\left(\frac{e^{-\mathrm{i}u\log K}}{\mathrm{i}u} \phi^{(\text{Heston})}(\theta; u, T-t)\right) \mathrm{d}u. \end{aligned} \tag{43}$$

In the above, i corresponds to the imaginary unit and ϕ refers to the characteristic function of $\log(S_t)$. There are multiple formulations for ϕ, but we choose to follow Schoutens *et al.* (2004), whose formulation gives numerically stable results:

$$\begin{aligned} \phi^{(\text{Heston})}(\theta; u, T-t) &= \exp\left\{\mathrm{i}u\left(\log(S_t) + r(T-t)\right) + \frac{\kappa\bar{V}}{\sigma_V^2}\right. \\ &\quad \times \left[(\xi - g)(T-t) - 2\log\left(\frac{1 - g_2\mathrm{e}^{-g(T-t)}}{1-g_2}\right)\right] \\ &\quad \left. + \frac{\sigma_{t,0}^2}{\sigma_V^2}(\xi - g)\frac{1-\mathrm{e}^{-g(T-t)}}{1-g_2\mathrm{e}^{-g(T-t)}}\right\}, \end{aligned} \tag{44}$$

where

$$\xi = \kappa - \sigma_V \rho \mathrm{i}u, \tag{45}$$

$$g = \sqrt{\xi^2 + \sigma_V^2(u^2 + \mathrm{i}u)}, \tag{46}$$

$$g_1 = \frac{\xi + g}{\xi - d}, \quad g_2 = \frac{\xi - g}{\xi + g}. \tag{47}$$

Denote by N_d the number of different option products traded for day d, and denote their maturity times by $(T_1, \ldots, T_{N_d})$, strike prices by $(K_1, \ldots, K_{N_d})$, mid-quote prices by $(C_1, \ldots, C_{N_d})$, traded volumes by $(\text{Volume}_1, \ldots, \text{Volume}_{N_d})$ and quote spreads by $(\text{Spread}_1, \ldots, \text{Spread}_{N_d})$. The vector of the residuals between theoretical and observed prices is given by

$$\begin{aligned} e &= (e_1, \ldots, e_{N_d}) \\ &= \left(C^{\text{Heston–Call}}_{T_1,K_1,\theta}(t,S_t) - C_1, \ldots, C^{\text{Heston–Call}}_{T_{N_d},K_{N_d},\theta}(t,S_t) - C_{N_d}\right). \end{aligned} \tag{48}$$

By choosing a suitable weighting scheme $\{w_i\}_{i=1}^{N_d}$ for the different option products (as discussed in the next section), we can then calibrate the Heston model using the following objective function

$$\begin{aligned} &\min_{\theta=(\kappa,\sigma_{d,0},\sigma_V,\bar{V},\rho)} \sum_{i=1}^{N_d} w_i e_i \\ &= \min_{\theta=(\kappa,\sigma_{d,0},\sigma_V,\bar{V},\rho)} \sum_{i=1}^{N_d} w_i \left[C_i - C^{\text{Heston–Call}}_{T_i,K_i,\theta}(t,S_t)\right]^2. \end{aligned} \tag{49}$$

A full description of how to explicitly compute the gradient of the right-hand side of equation (49) is given by Cui *et al.* (2015). Alternatively, a difference method can be used to calculate each partial derivative. A solution can then be obtained using a Levenberg–Marquardt approach

† Note that the implied volatility surface data used is constructed based on maturity and Delta, while the grid created for the sensitivity analysis is based on maturity and strike price, which is more suitable for demonstration purposes. As the price is fixed at $S = 1$, there is a 1–1 relationship between strike prices and deltas.

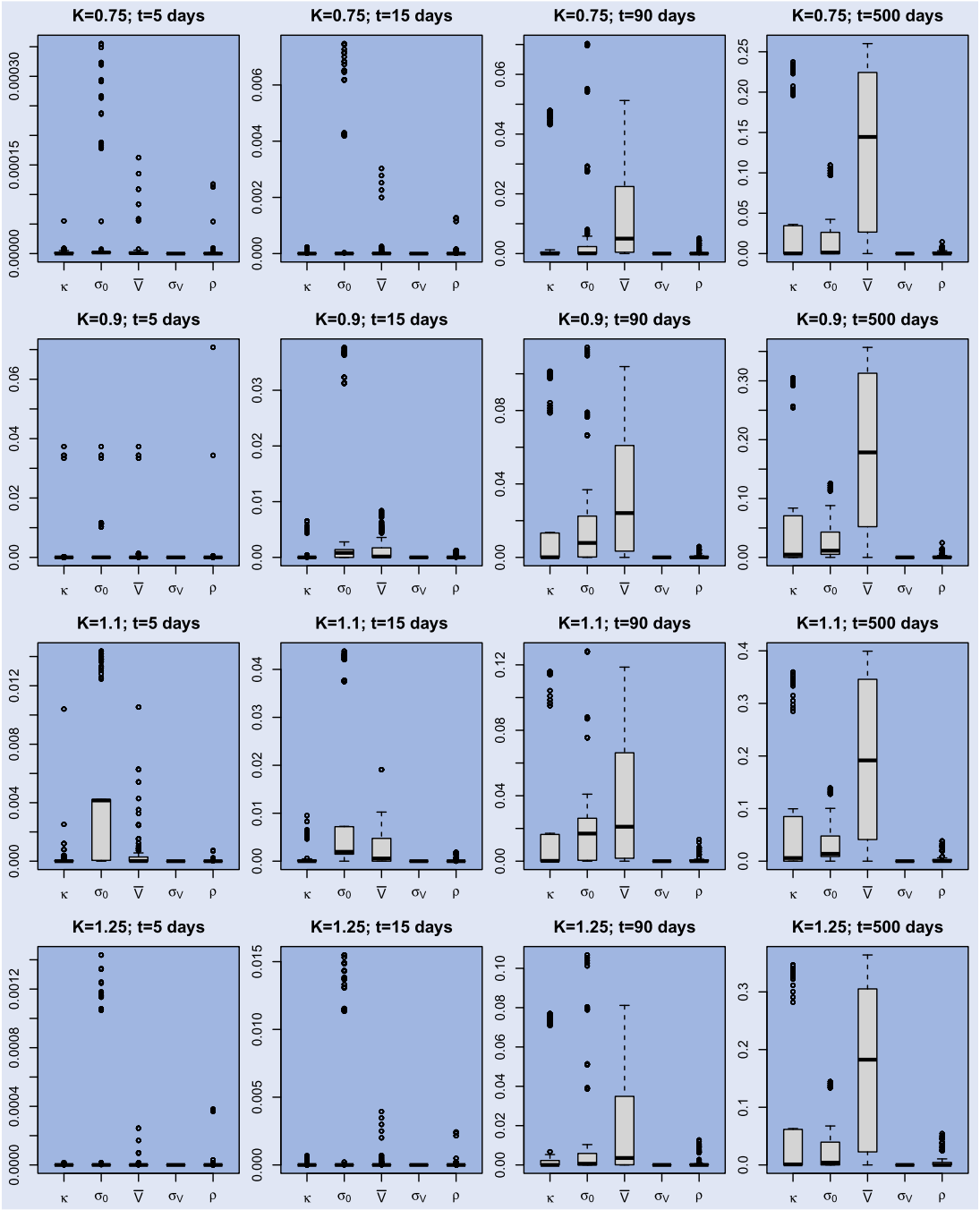

Figure 2. Sensitivity of the call option price to each of the Heston model's parameters, for a total of 16 (strike price, time to expiration) combinations. Each plot takes the form of a Box-Whisker plot, demonstrating the distribution of sensitivities, which are calculated as explained in equation (40).

(Moré 1978). We carry out the full calibration using a combination of a gradient method and Particle Swarm Optimization. We provide further details in appendix 1.

4.4. Partial calibration of the Heston model

In the previous section, we provided details on how to fully calibrate the Heston model. However, this is generally a hard problem for a number of reasons. Firstly, the estimation of equation (27) for the Heston model involves the estimation of the integrals in equations (42) and (43), making it non-negligible in terms of computing complexity. Secondly, due to the non-convexity of the problem, there is no guarantee that a gradient descent approach will converge to the global minimum. In cases where the number of products is low, multiple parameter sets can yield equally good fits, especially when there are less than five option products. Finally, the success of gradient methods for the Heston model is known to be quite sensitive to the initial set of parameters (Mrázek and Pospíšil 2017).

The aforementioned reasons render the calibration process a difficult task. While we borrow methodology from the calibration literature, our goal here is inherently different. In our case, an exact calibration of the problem is not required, as we are only interested in obtaining a proxy of the volatility

based on option data; in other words, we only wish to track the evolution of $\sigma_{d,0}$. Therefore, we are content with keeping the other four parameters fixed, deviating from the standard calibration procedure (an illustration of the process is given in figure 3).

Initially we run a full calibration algorithm as described in appendix 1. For each stock, we fix the values of ρ and σ_V to be equal to the median values of the calibrated parameters as obtained from the full calibration of each stock. We consider the ranges of values that κ and σ_V take. Between these extreme values we take a total of 5 values equally spaced on a logarithmic scale (as volatilities have highly skewed distributions). This creates a 5x5 grid of (κ, σ_V) pairs. For each of the pairs, we calibrate with respect to $\sigma_{d,0}$, while keeping the other four parameters fixed, by finding

$$\sigma^{(d)}_{\text{Heston}}(\kappa, \sigma_V) = \min_{\sigma_0} \sum_i w_i \left[C_i - C^{\text{Heston}-Call}_{T_i,K_i,(\kappa,\sigma_{d,0},\rho,\sigma_V,\bar{V})}(d, S_d) \right]^2. \quad (50)$$

Once we calculate this grid of values for pairs $(\kappa, \sigma_V) \in \Theta_\kappa \times \Theta_{\sigma_V}$, we can then proceed to find the pair that minimizes the error function, and thus the value of volatility that corresponds to the minimum value of the loss function.

$$(\kappa^*, \sigma_V^*) = \underset{\kappa\in\Theta_\kappa, \sigma_V\in\Theta_{\sigma_V}}{\text{argmin}} \sum_i w_i \times \left[C_i - C^{\text{Heston}-Call}_{T_i,K_i,\left(\kappa,\sigma^{(d)}_{\text{Heston}}(\kappa,\sigma_V),\rho,\sigma_V,\bar{V}\right)}(d, S_d) \right]^2, \quad (51)$$

where

$$\sigma^{(d)}_{\text{Heston}} = \sigma^{(d)}_{\text{Heston}}(\kappa^*, \sigma_V^*). \quad (52)$$

Secondly, we carry out PCA on a vectorized form of the 5x5 grid of volatility values, after subtracting $\sigma^{(d)}_{\text{Heston}}$ by all values. For all instruments investigated, we find that two principal components suffice to capture 95% of variability.

4.5. *Partial calibration of the Bates model*

We also consider the Bates model (Bates 1996), which enhances the Heston framework by incorporating a jump-diffusion component, allowing for differentiation between high volatility and ad-hoc abrupt price movements. The Bates model is a combination of the Heston and Merton models. The dynamics of the volatility process under the Bates model is still described by equation (39), similar to the Heston model. The price process is similar to equation (38), but incorporates a jump component

$$\mathrm{d}S_t = \mu S_t \mathrm{d}t + S_t \sqrt{V_t} \mathrm{d}B^1_t + S_t \left(\mathrm{e}^{J_t} - 1\right) \mathrm{d}N_t, \quad (53)$$

where N_t is a Poisson process with intensity λ, and $J_t \sim N(\mu_J, \sigma_J)$. Therefore, in addition to the five parameters present in the Heston model, there are three additional parameters to be calibrated.

Once again, as we are only interested in tracking changes in volatility, an accurate calibration of the model is not required. In fact, it is natural to decouple the jumps from the rest of the process, by detecting the price jumps, prior to the calibration. While the field of intraday jump detection is vast and evolving, detection of daily jumps is a significantly easier task, as it is almost equivalent to the task of outlier detection. Thus, we use the following simple sliding window algorithm to determine whether day d is a jump point. Based on the L previous daily returns, we calculate the mean $\text{HM}^{(L)}_d = \sum_{l=0}^{L-1} \log(\frac{C_{d-l}}{C_{d-(l-1)}})$ and standard deviation $\text{HV}^{(L)}_d$ (equation (13)). If the z-score

$$\frac{|\log(\frac{C_{d-L}}{C_{d-(L-1)}}) - \text{HM}^{(L)}_d|}{\text{HV}^{(L)}_d}$$

is greater than some pre-determined threshold, then the daily price process of day d is classified as having a jump.

Once all the jumps have been identified, we estimate λ, μ_J and σ_J for day d, based on the jumps detected in days $d - L', \ldots, d - 1$, using the Maximum Likelihood estimates for the Poisson and Normal distributions, respectively. We then follow the same methodology as in section 4.4, by using equation (41). With the extended parameter set $\theta = (\kappa, \sigma_{d,0}, \sigma_V, \bar{V}, \rho, \lambda, \mu_J, \sigma_J)$, and the pre-calculated values of λ, μ_J and σ_J, we replace the characteristic function in equation (44) with

$$\phi^{(\text{Bates})}(\theta; u, t) = \phi^{(\text{Heston})}(\theta; u, t) \exp \times \left\{ t\lambda \left((1 + \mu_J)^{\mathrm{i}u} e^{\frac{\mathrm{i}u}{2}(\mathrm{i}u-1)\sigma_J^2} - 1 \right) \right\}. \quad (54)$$

4.6. *Choice of weighting scheme*

A common choice of weighting scheme is taking into account the option spread, by setting

$$w_j \propto \frac{1}{F(\text{Spread}_j)}, \quad (55)$$

where F is a monotonically increasing function, usually one of $\sqrt{(\cdot)}$, $|(\cdot)|$, or $(\cdot)^2$. It is shown by Mrazek *et al.* (2016) that each of these weighting schemes achieves a lower error than a uniform scheme or schemes based on Vega. As we are dealing with daily data and our purpose is not an accurate calibration of the model, we take a slightly different approach. In particular, we use the following weighting scheme:

$$w_j \propto \exp\left(\frac{-\left|K_j - S_d\right|}{w_{\text{Moneyness}}} \right) \exp\left(\frac{-T_j}{w_{\text{Maturity}}} \right) \log\left(\text{Volume}_j\right). \quad (56)$$

The weighting is informed by the sensitivity analysis of Heston parameters illustrated in figure 2. The impact of $\sigma_{d,0}$ during the calibration is greater for products which are close to expiration and which are close to being ATM. The parameters $w_{\text{Moneyness}}$ and w_{Maturity} are empirically chosen to be 1 and 10 respectively, so that the normalized distributions of $\exp(\frac{-|K_j - S_d|}{w_{\text{Moneyness}}})$, $\exp(\frac{-(T_j - d)}{w_{\text{Maturity}}})$ and $\log(\text{Volume}_j)$ have similar values for kurtosis.

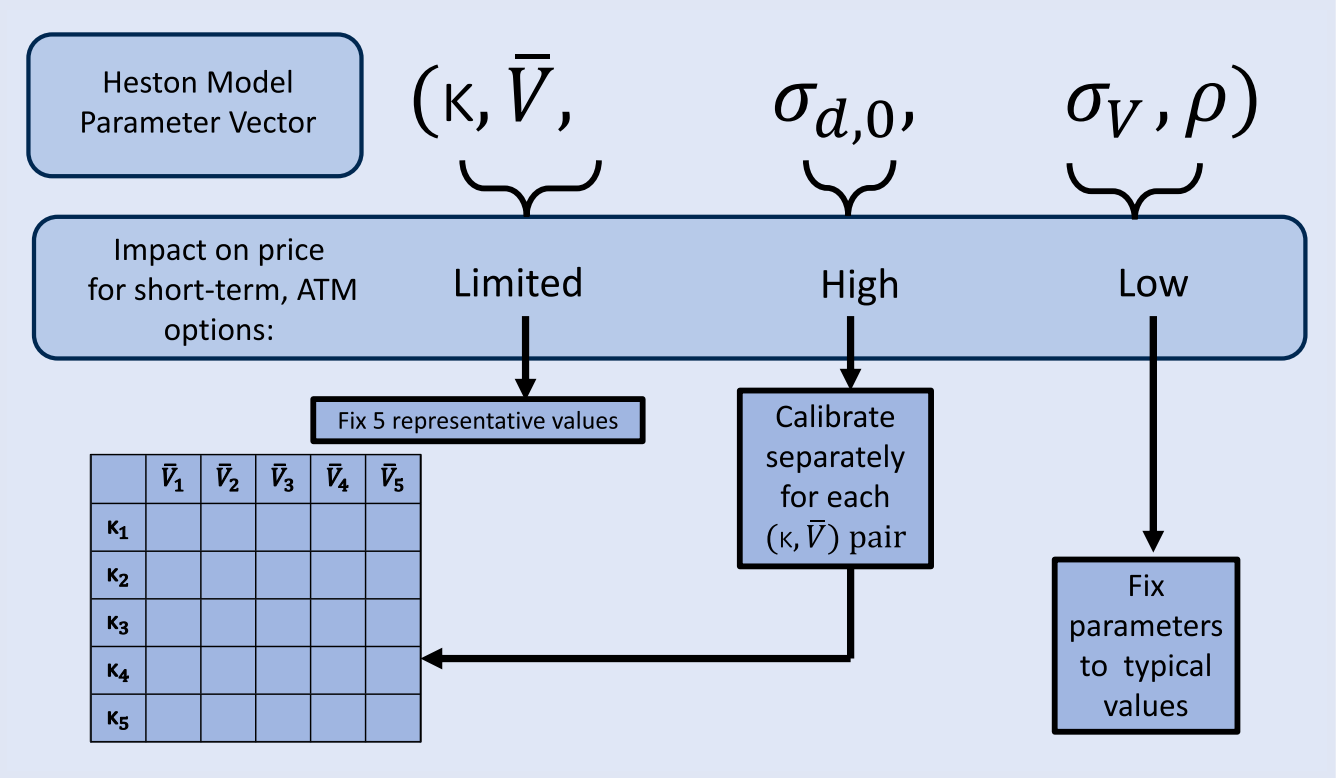

Figure 3. Pictorial illustration for the process of the extraction of the volatility estimator $\sigma^{(d)}_{\text{Heston}}$, described in section 4.4. For each instrument, the parameters (σ_V, ρ) are fixed to a singular value, while the parameters $(\kappa, \bar{V})$ take one of five predetermined values, creating 2-dimensional grid. For each combination $(\kappa, \bar{V})$ and the fixed values of (σ_V, ρ), the model is then calibrated for the optimal volatility $\sigma_{d,0}$, based on minimizing equation (49).

5. Forecasting setup

We aim to test whether augmenting a HAR-based baseline model with additional estimators of the daily, weekly and monthly variance can increase the forecasting performance. Note that as the distribution of daily realized variance has a high positive skew, we focus on forecasting log(RV) instead of RV. For each instrument, we fit a distinct model, with features built using data from its own past historical price and option data, i.e of the form

$$\log(\text{RV}_{i,d}) = \mathcal{F}_i\left([X^{\text{Base}}_{i,d-1}, X^{\text{Augm}}_{i,d-1}]\right),$$

where $\text{RV}_{i,d}$ denotes the realized variance, and $X^{\text{Base}}_{i,d}(X^{\text{Augm}}_{i,d})$ denotes the baseline (additional) matrix of predictors used to forecast it, for the ith instrument on the dth day. In figure 4, we show the different augmented feature sets we investigate in this section. Other than the natural extension of the HAR model, which is the HARQ model equation (11), this includes the additional daily, weekly and monthly volatility estimators, as well as additional predictors which are commonly used in the literature.

In this chapter we setup the forecasting task, starting with a short note in section 5.1 on the data set used. We then describe the evaluation metrics used in section 5.2, and the fitting methods used in section 5.3. In section 5.4, we focus on how to utilize the implied volatility features described in section 3. Lastly, in section 5.5, we describe all the different augmented feature sets.

5.1. Data set

In this study, we utilize data from OptionMetrics covering the period 2013–2019. In particular, we use the implied volatility surfaces, option prices, and option volume datasets from OptionMetrics. We also use intraday minutely data for equities, extracted from LOBSTER (2023), which is used to calculate daily volatility estimators and various intraday features. Lastly, we use daily summary data for equity volumes from CRSP (OptionMetrics 2023). In total, we consider a selection of 66 NYSE stocks that do not pay dividends and for which the described data is available for the specified period.

5.2. Evaluation metrics

We use the following evaluation metrics to benchmark the forecasting models under consideration. Denote the vector of forecasted values by $\hat{y}$, and the observed vector by y, which in the context of this paper mainly refers to realized volatility. We consider the following suite of standard evaluation metrics.

Root Mean Squared Error (RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{D}\sum_{d=1}^{D}\left(y_d - \hat{y}_d\right)^2}. \tag{57}$$

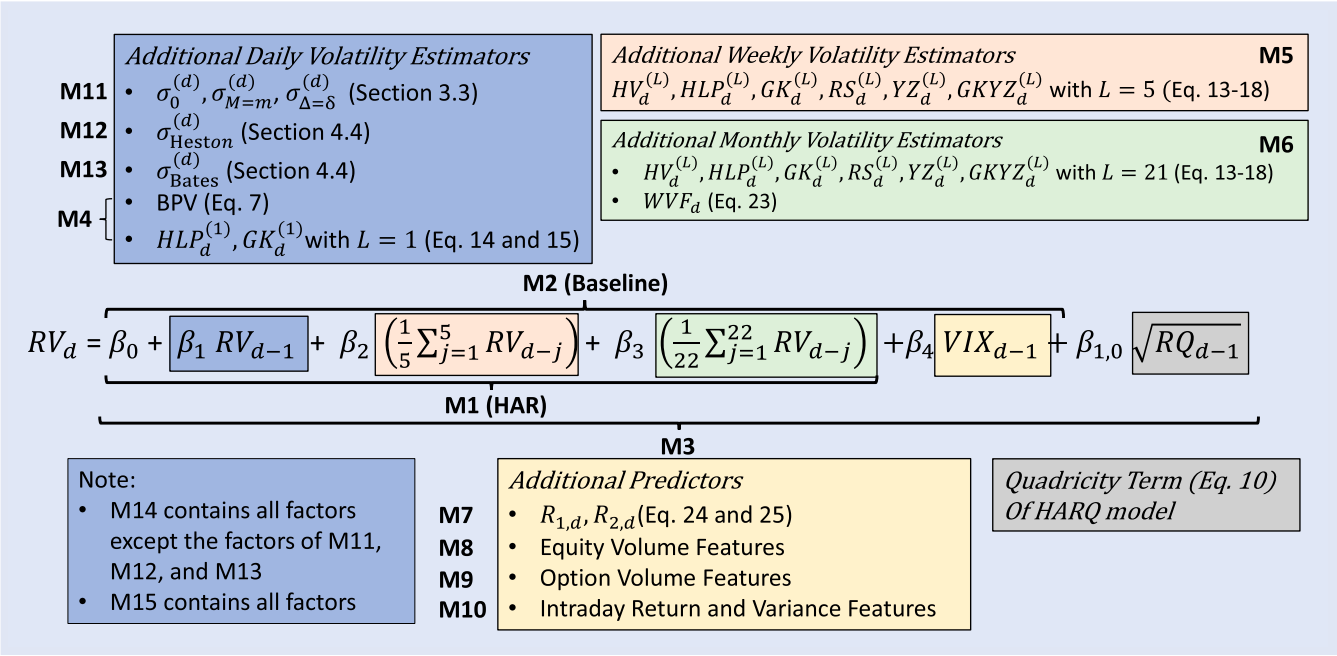

Figure 4. Pictorial illustration for the investigated features. M1 (resp., M2) corresponds to the HAR (resp., the baseline) model. The additional predictors for daily, weekly and monthly predictors are shown correspondingly, as well as a list of additional predictors for volatility.

Mean Absolute Error (MAE):

$$\text{MAE} = \frac{1}{D}\sum_{d=1}^{D} \left| y_d - \hat{y}_d \right|. \tag{58}$$

Spearman Correlation: The Spearman Correlation between $\hat{y}$ and y is the Pearson correlation between the ranks of the two vectors. This allows this metric to capture non-linear monotonic relations. The Pearson correlation between $\hat{y}$ and y is given by

$$C = \frac{\sum_{d=1}^{N}(y_d - \bar{y})(\hat{y}_d - \bar{\hat{y}})}{\sqrt{\sum_{d=1}^{N}(y_d - \bar{y})^2 (\hat{y}_d - \bar{\hat{y}})^2}}. \tag{59}$$

Quasi-Likelihood (QLIKE):

$$\text{QLIKE} = \frac{1}{D}\sum_{d=1}^{D} \frac{\exp y_d}{\exp \hat{y}_d} - (y_d - \hat{d}_t) - 1, \tag{60}$$

We also directly compare models in a hypothesis testing framework. Consider $\hat{y}^{(i)}$, the forecasts generated by model i. We compute the loss differential for day d between models i and j as

$$\text{LD}_d\,(i,j) = \text{LOSS}\left(\hat{y}_d^{(i)}, y_d\right) - \text{LOSS}\left(\hat{y}_d^{(j)}, y_d\right). \tag{61}$$

To test whether model j outperforms the baseline model i, we use the hypothesis

$$H_0 : \mathbb{E}\left[\text{LD}_d\,(i,j)\right] = 0, \tag{62}$$

For that purpose, we use the one-tailed **Diebold–Mariano (DM) test**, a popular variant of the paired t-test introduced in Diebold and Mariano (2002), which tests the null hypothesis against the alternative

$$H_1 : \mathbb{E}\left[\text{LD}_d\,(i,j)\right] > 0. \tag{63}$$

We also utilize the **Model Confidence Set (MCS)** procedure by Hansen *et al.* (2011), which is used to identify a subset $\mathcal{M}^*$ of models from a candidate set $\mathcal{M}_0 = \{M1, M2, \dots\}$ that are statistically indistinguishable from the best model. In other words, we are looking for the set $\mathcal{M}^* \subset \mathcal{M}_0$ defined as

$$\mathcal{M}^* = \{i \in \mathcal{M}_0 : \mathbb{E}\left[\text{LD}_d\,(i,j)\right] \le 0 \quad \forall j \in \mathcal{M}_0\}. \tag{64}$$

For each pair of models, we calculate

$$T_{i,j} = \frac{\text{LD}_d\,(i,j)}{\text{Var}\left(\{\text{LD}_d\,(i,j)\}_{d=1}^{D}\right)}. \tag{65}$$

Once we estimate $T_{i,j}$ for all pairs of models, the one-tailed test statistic for each model is given by

$$T_i = \max_{j \neq i} T_{i,j}. \tag{66}$$

MCS is based on sequential testing, where models that perform significantly worse than others are eliminated. In particular, at step n, a p-value p_i is computed for each model by comparing T_i to a bootstrap sample. If some models have p-values smaller than a critical value, the model with the greatest test statistic is removed by defining a new candidate set

$$\mathcal{M}_n = \mathcal{M}_{n-1} / \left\{ \underset{j \in \mathcal{M}_{n-1}}{\text{argmax}}\, T_j \right\}. \tag{67}$$

Once no further models can be removed form the candidate set, the remaining set constitutes the Model Confidence Set,

which includes models that are not significantly different from the best model.†

5.3. *Regression methods*

As the number of features we use is relatively large and the features are highly correlated, OLS regression would be inadequate. Deep learning methods are also inadequate, as the number of available data points is quite low (as this is a daily forecasting problem). We consider three widely used models: Elastic Nets, LASSO, and XGBoost, which we shortly describe. Considering three different methods, acts as a robustness check for the claim that augmenting the HAR model with additional features results in an increase in performance.

The Elastic Net (EN) is a linear regression method, which alters the objective function of Ordinary Least Squares regression to simultaneously include both ℓ_1 and ℓ_2 norms, to penalize overfitting as follows

$$\hat{\beta} = \text{argmin}_{\beta} \|y - X\beta\|^2 + \lambda \left(\alpha \|\beta\|_1 + (1-\alpha)\frac{\|\beta\|_2}{2} \right), \tag{68}$$

where β is the parameter vector and X is the feature matrix. Here, λ and α are hyperparameters chosen using cross-validation.‡ LASSO (Least Absolute Shrinkage and Selection Operator) is a special case of the EN with α set to 1 (so that only the ℓ_1 norm appears in the penalty term).

Boosting is a class of ensemble models where the weak learners usually take the form of decision trees. At each iteration, a new weak learner is fitted, whose training phase focuses on the previously misclassified data points. The eXtreme Gradient Boosting (XGBoost) is one of the most widely used variants of boosting.

5.4. *Incorporating IVS features*

While implied volatility is clearly related to realized volatility, using it in practice can be challenging and often fails to be a good predictor for realized volatility. In particular, a different set of option products may be traded each day, with a multitude of expiration dates and strike prices. Therefore, using implied volatility surfaces creates a set of 280 features which are directly comparable across days. However, this leads to a significant increase in dimensionality, with many features being highly correlated. Furthermore, in many instances, the IVS are calculated using only a handful of option products which were traded on a particular day, thus rendering the IVS a data set of lower intrinsic dimensionality.

A natural solution is to perform dimensionality reduction, such as Principal Component Analysis (PCA).§ In section 3.3, we explored a decomposition that effectively reduces dimensionality from 280 to 39. We set up a total of five feature sets that use either of the two methods or a combination of both. We describe each feature set by its additional features X^{Augm}.

- Complete (IVS-1): This feature set includes all 280 features of the implied volatility surface in vectorized form, i.e with no regard to the IVS structure.
- Complete-PCA (IVS-2): Here we perform PCA on the IVS matrix of dimensionality 280. We choose a dimensionality of $p = 7$ as it is enough to retain 90% of the variance for 94% of instruments.
- Decomposed (IVS-3): This set composes of all of the fixed maturity and delta effects obtained using the methodology explained in section 3.3, which reduces dimensionality down to $p = 39$.
- Decomposed-PCA (IVS-4): For this set, we perform PCA to reduce the dimensionality the matrix containing all of the maturity and delta effects $\sigma^{(d)}_{M=m_1}, \ldots, \sigma^{(d)}_{M=m_{10}}, \sigma^{(d)}_{\Delta=\delta_1}, \ldots, \sigma^{(d)}_{\Delta=\delta_{28}}$. With just six components, 90% of variance is retained for 96% of instruments. With the addition of $\sigma^{(d)}_0$, dimensionality is reduced to $p = 7$.
- Decomposed-Separate PCA (IVS-5): Unlike IVS-4, we use PCA on each maturity and delta effects separately. In particular, we use the first two principal components built using the maturity effects $\sigma^{(d)}_{M=m_1}, \ldots, \sigma^{(d)}_{M=m_{10}}$, the first four principal components built using the delta effects $\sigma^{(d)}_{\Delta=\delta_1}, \ldots, \sigma^{(d)}_{\Delta=\delta_{28}}$, and the IV level $\sigma^{(d)}_0$. Using these features is sufficient for retaining 90% of the variance for 91% of the instruments.

For each set, we compute the reconstruction error. Note that for IVS-1, it is equal to zero by default. For IVS-2, this is calculated as the reconstruction error for PCA.¶ For IVS-3 we calculate the reconstruction error as

$$\sqrt{\sum_{d=1}^{D} \sum_{m \in M} \sum_{\delta \in \Delta} \left(\hat{\sigma}^{(d)}_{m,\delta} - \hat{\sigma}^{(d)}_0 - \sigma^{(d)}_{\Delta=\delta} - \sigma^{(d)}_{M=m} \right)^2}. \tag{70}$$

For IVS-4 and IVS-5 the calculation is done by first reconstructing $\{\sigma^{(d)}_{\Delta=\delta_i}\}_{i=1}^{28}$ and $\{\sigma^{(d)}_{M=m_i}\}_{i=1}^{10}$ and then calculating the error using equation (70).

In table 2, we show the reconstruction error for each of these models. Using PCA after the decomposition shown in equation (1) performs better than directly applying PCA, as

† Note that for both MCS and the DM test, we use the QLIKE function, with a significance level of 5%.

‡ We use a 60-10-30 split rule for training, validation, and testing.

§ Denote by $C(X)$ the covariance matrix of X, and its eigendecomposition by $C(X) = U\Lambda U^T$, where Λ is a diagonal matrix capturing the $\lambda_1 \geq \cdots \geq \lambda_{\text{col}(X)} \geq 0$ eigenvalues of $C(X)$, and the columns of U are given by the $u_1, \ldots, u_{\text{col}(X)}$ orthonormal eigenvectors (in the same order as their corresponding eigenvalues), where $\text{col}(X)$ denotes the number of columns of matrix X. The principal component scores are the columns of XU, with each new component's variance given by its corresponding eigenvalue (component i has variance λ_i). Selecting the first p components retains a proportion of the total variance in the feature matrix equal to $\sum_{i=1}^{p} \lambda_i / \sum_{i=1}^{\text{col}(X)} \lambda_i$. Given the proportion of variance to retain $\delta_{\text{prop}} \in (0, 1)$, we set

$$p = \min \left\{ i : \frac{\sum_{k=1}^{i} \lambda_k}{\sum_{k=1}^{d} \lambda_k} \geq \delta_{\text{prop}} \right\}. \tag{69}$$

¶ We can reconstruct X using the first p eigenvectors as $\tilde{X} = XUU^T$ The reconstruction error can then be calculated as $\|X - \tilde{X}\|_F$, where $\|\cdot\|_F$ denotes the Frobenius norm.

Table 2. Comparison of dimensionality reduction methods by reconstruction error.

		Reconstruction Error			
	Dimension (p)	LQ	Median	Mean	UQ
IVS-1	280	0	0	0	0
IVS-2	7	9.524	13.995	16.954	21.303
IVS-3	39	4.253	7.100	7.834	10.478
IVS-4	7	5.483	8.109	8.967	11.272
IVS-5	7	5.263	7.745	8.799	11.543

Note: The reconstruction error is calculated as the Frobenius norm of the difference between the reconstructed surface and the original surface. The median, mean, lower quantile (LQ) and upper quantile (UQ) across instruments are given.

both IVS-4 and IVS-5 clearly outperform IVS-2. On the other hand, the performance of IVS-4 is only marginally better than IVS-5. While the reconstruction errors for both these sets are higher than that of IVS-3, the differences are rather small. We opt to use IVS-5, as it has the added benefit of its features being easier to interpret.

In table 3, we directly compare the performance of five models based on each feature set, fitted using EN, compared to the baseline model described later in section 5.5. Note that all of the models incorporate the features of a HAR model (daily, weekly and monthly volatility) for this task. IVS-1 (using all 280 features of the IVS) is outperformed by the other sets across all metrics, thus confirming the benefit of performing a dimensionality reduction step. Furthermore, IVS-4 and IVS-5 achieve the highest performance, and significantly outperform the baseline model for more than 21 instruments. Furthermore, both appear in the MCS. On the other hand, IVS-4 and IVS-5 have very similar performances. Therefore, taking into account that the factors of IVS-5 are readily interpretable, renders IVS-5 as our feature set of choice.

Based on IVS-5, we plot the first two principal component loadings for maturity, and the first four principal loadings for Δ in figure 5 for Adobe Inc. Note that these effects are mostly similar across most instruments, up to a difference in sign.

- *The first PC for maturity* assigns positive weights to low maturities (30, 60 and 91) and negative weights for high maturities, thus acting as the signed difference in volatility between short and long term options.
- In contrast to the first PC, *the second PC for maturity* has a mostly upwards trend, assigning positive weights to high maturities, and negative weights for low maturities. However, a very high positive weight is placed on the maturity of 30 days, thus substantially differentiating the importance of the short term (30-day) expiration options.
- *The 1st PC for* Δ assigns mostly positive weights to put options ($\Delta < 0$) and negative weights to call options ($\Delta > 0$), with the highest magnitudes being given for Δ's close to 0. This essentially yields a signed difference in volatility between call and put options, with a focus on OTM options.
- In contrast to the first PC, *the* 2^{nd} *PC for* Δ assigns mostly negative weights to put options ($\Delta < 0$) and positive weights to call options ($\Delta > 0$), with the highest magnitudes being assigned to Δ's close to -1 or 1. This essentially gives a signed difference in volatility between call and put options with a focus on ITM options.
- *The 3rd PC for* Δ assigns mostly positive weights to options with Δ close to 0.5 or -0.5, and negative weights for ITM or OTM options. This effect is quite clear for put options, and less so for call options, where weights are smaller and the effect for $\Delta = 0.75$ is relatively high, which distorts the overall trend of the graph.
- *The 4th PC for* Δ assigns mostly positive weights to options with Δ close to 0 and negative weights to options with Δ close to -1 or 1. But unlike the first three PCs, this effect appears to have multiple turning points.

5.5. *Description of feature sets*

In this paper, we have described several categories of features in sections 2.3, 3, and 4. We consider a number of different models built using one of the three methods described in section 5.3 and one of the feature sets described in this subsection. In particular, we define models using the same features as HAR, baseline models (on for each method) based on HAR with some additional features, or a number of different features sets used to augment the baseline feature set. In this subsection, we define a number of feature sets used to build models to determine whether each set of features (X^{Augm}) provides additional forecasting power compared to a HAR-based baseline model.

M1: It consists of all the features included in a HAR model equation (9), which is widely used as a benchmark for volatility forecasting problems.

M2: M2 is used as the baseline model for each method, against which the other models are compared. It consists of all the features included in a HAR model equation (9), as well as percentage change of realized volatility defined as $\frac{\text{RV}_d - \text{RV}_{d-1}}{\text{RV}_{d-1}}$. Furthermore, we include the value of the Volatility index (VIX) and percentage change (return) of VIX, as it was by shown Kambouroudis and Mcmillan (2015) that it offers additional forecasting power (when forecasting the realized volatility of the stock market).

The following feature sets (M3-M13) include all of the features included in the baseline set (M2) as well as the set-specific features described in the subsequent paragraphs.

M3: As discussed, the natural extension of the HAR model is the HARQ model in equation (11), i.e M3 also includes the realized quadricity feature.

M4: In section 2, we have discussed additional daily estimators that we incorporate in a model denoted as *M3*. Importantly, this includes the Bipower Variation equation (7). We also make use of the daily Parkinson (equation (14)) and Garman–Klass (equation (15)) estimators in the same manner as Będowska-Sójka and Kliber (2021), i.e we set $L = 1$. Lastly, we apply the concept of realized semivariance (Barndorff-Nielsen *et al.* 2010). In particular, the positive and

Table 3. Comparison of the performance of five different models making use of IVS data.

	MSE		MAE		R^2		QLIKE		DM test	
	Mean	% of improv.	Mean	% of improv.	Mean	% of improv.	Mean	% of improv.	% of rejections	MCS
IVS-1	0.7123	43	0.4905	43	0.2321	50	0.7656	57	16	
IVS-2	0.6995	57	0.4788	52	0.2487	55	0.7621	57	21	
IVS-3	0.7093	55	0.4852	54	0.2416	59	0.7635	57	18	
IVS-4	0.6989	62	0.4723	63	0.2520	64	0.7693	61	24	*
IVS-5	0.6996	57	0.4796	59	0.2496	59	0.7619	61	21	*

Note: All models are fitted using EN. For each model, we compare the performance in terms of the out-of-sample R^2, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). For each performance metric, we show the mean and the percentage of cases where the null model is outperformed. In addition, for models IVS-2, IVS-3, IVS-4, and IVS-5 we perform a Diebold–Mariano test comparing the residuals of each model with the residuals of the HAR model, which provides (in the last column) the number of instruments for which the null hypothesis is rejected (out of 66).

negative semivariances are calculated as

$$\mathrm{RV}_d^+ = \sum_{j=2}^{391} \log\left(\frac{S_{d,j}}{S_{d,j-1}}\right)^2 \mathbb{1}\left(S_{d,j} \geq S_{d,j-1}\right). \tag{71}$$

$$\mathrm{RV}_d^- = \sum_{j=2}^{391} \log\left(\frac{S_{d,j}}{S_{d,j-1}}\right)^2 \mathbb{1}\left(S_{d,j} \leq S_{d,j-1}\right). \tag{72}$$

M5: In section 2.3, we have discussed various historical volatility estimators. In *M5*, we include all weekly estimators including $\mathrm{HV}^{(5)}, \mathrm{HLP}^{(5)}, \mathrm{GK}^{(5)}, \mathrm{RS}^{(5)}, \mathrm{GKYZ}^{(5)}, \mathrm{YZ}^{(5)}$ as defined in equations (14) to (18), with $L = 5$.

M6: In *M6*, we include all monthly estimators including $\mathrm{HV}^{(21)}, \mathrm{HLP}^{(21)}, \mathrm{GK}^{(21)}, \mathrm{RS}^{(21)}, \mathrm{GKYZ}^{(21)}, \mathrm{YZ}^{(21)}$ as defined in equations (14) to (18), with $L = 21$. We also include VIX-Fix (equation (23)).

M7: In section 2.4, we have discussed the path dependent estimators used by Guyon and Lekeufack (2023). Thus, we include these two features in feature set *M7*.

We also consider volume-based features, which have been previously reported in the literature to be associated with volatility, as discussed in section 1. Denote by Volume_d and OI_d the traded volume and open interest for day d respectively. Note that trading volumes are not stationary and tend to increase over time. To render the features stationary, we consider the percentage change of daily volume $\frac{\mathrm{Volume}_d - \mathrm{Volume}_{d-1}}{\mathrm{Volume}_{d-1}}$, as well as the daily turnover $\frac{\mathrm{Volume}_d}{\mathrm{OI}_d}$.

M8: In *M7* we include both the percentage change of daily volume and the daily turnover those features for each corresponding equity instrument.

M9: For *M9* we consider the daily turnover and percentage change in volume for options. We also consider those quantities restricted to call and put options respectively.

M10: Feature set *M10* includes intraday features. In particular, we calculate the intraday returns and Realized Variance using only the last $l = 5, 10, 30, 60$ and 90 minutes prior to closing. That is $\log(\frac{S_{d,j}}{S_{d,j-l}})$ and $\sum_{j=391-l}^{391} \log(\frac{S_{d,j}}{S_{d,j-1}})^2$ for l = 5, 10, 30, 60 and 90 respectively, similar to Zhang *et al.* (2022). We also include the sum of the positive minutely returns and the sum of the negative minutely returns respectively

$$\sum_{j=2}^{391} \max\left(\log\left(\frac{S_{d,j}}{S_{d,j-1}}\right), 0\right) \quad \text{and}$$
$$\sum_{d=2}^{391} \max\left(-\log\left(\frac{S_{d,j}}{S_{t,j-1}}\right), 0\right). \tag{73}$$

A major focus of the paper is on formulating volatility features using option prices. In sets M11-M13 we consider the volatility estimators discussed in the paper.

M11: In section 5.4, we explored a number of feature sets using the features built using the IVS data set. M11 incorporates the features of IVS-5.

M12: The features based on the Heston model described in section 4, are included in *M12*, as well as the percentage differences for each feature.

M13: The features based on the Bates model described in section 4, are included in *M13*, as well as the percentage differences for each feature.

M14 and **M15**: To compare the effect of features built using option prices we formulate *M14* to include all of the discussed features excluding those of *M11-M13*, and *M15* to include all features discussed (including *M11-M13*).

6. Results

In this section, we showcase the performance results by comparing each model to the baseline. In table 4, we present the performance results of models based on M1-M15 and the three different methods. For each model, we report the mean RMSE, MAE, R^2, and QLIKE. Using M2 as a baseline, we also provide the percentage of instruments for which the DM hypothesis is rejected, as well as the percentage of instruments that show an increase in the performance of each metric.

For each instrument, we classify days into normal and high-volatility periods, where a high-volatility day is defined as having observed RV higher than at least 80% of the other days (i.e it is in the top quintile). In table 5, we report the performance of each model separately for days with high and normal observed RV. We also report the classification accuracy, which is the accuracy in correctly forecasting whether volatility will be high or normal.

Table 4. Performance comparison for M1-M15.

	RMSE		MAE		R^2		QLIKE		DM Test	MCS
EN	Mean	% of improv.	Mean	% of improv.	Mean	% of improv.	Mean	% of improv.	% of rejections	Included
Model 1 (M1) - HAR	0.6236	7	0.4787	9	0.2810	5	0.7713	61	46	
Model 2 (M2) - Baseline Model	0.5986	NA	0.4577	NA	0.3225	NA	0.7759	NA	NA	
Model 3 (M3) - HARQ	0.6022	53	0.4592	53	0.3164	58	0.7754	53	31	
Model 4 (M4) - Additional Daily Volatility Estimators	1.8763	30	1.8129	32	0.3005	41	0.7726	66	49	*
Model 5 (M5) - Historical Volatility Estimators (Weekly)	0.6362	41	0.4730	43	0.3013	46	0.7743	54	39	*
Model 6 (M6) - Historical Volatility Estimators (Monthly)	0.9349	29	0.6741	34	0.2944	25	0.7697	46	37	
Model 7 (M7) - Path Dependence Features	0.6066	41	0.4619	46	0.3122	57	0.7747	54	43	
Model 8 (M8) - Equity Volume Features	0.5990	50	0.4577	52	0.3245	66	0.7753	57	41	
Model 9 (M9) - Option Volume Features	0.5989	59	0.4602	50	0.3284	73	0.7759	50	35	
Model 10 (M10) - Intraday Features	0.6006	66	0.4606	64	0.3359	73	0.7750	52	43	
Model 11 (M11) - IVS Volatility Features	0.6062	47	0.4695	44	0.3195	49	0.7750	60	50	*
Model 12 (M12) - Heston Model Volatility Features	0.5921	70	0.4542	43	0.3298	64	0.7772	43	32	
Model 13 (M13) - Bates Model Volatility Features	0.6203	54	0.4792	45	0.3207	55	0.7834	43	39	
Model 14 (M14) - All Features (excluding Option Features)	0.6204	55	0.4688	50	0.3256	62	0.7727	50	44	
Model 15 (M15) - All Features	0.6182	68	0.4769	64	0.3304	70	0.7703	66	52	*
	RMSE		MAE		R^2		QLIKE		DM Test	MCS
LASSO	Mean	% of improv.	Mean	% of improv.	Mean	% of improv.	Mean	% of improv.	% of rejections	Included
Model 1 (M1) - HAR	0.6237	7	0.4790	9	0.2810	5	0.7711	59	48	
Model 2 (M2) - Baseline Model	0.5990	NA	0.4581	NA	0.3224	NA	0.7758	NA	NA	
Model 3 (M3) - HARQ	0.6021	67	0.4588	64	0.3163	53	0.7755	51	33	
Model 4 (M4) - Additional Daily Volatility Estimators	0.9632	32	2.2640	34	0.3019	39	0.7730	62	46	*
Model 5 (M5) - Historical Volatility Estimators (Weekly)	0.6494	41	0.4754	46	0.2976	45	0.7744	55	43	
Model 6 (M6) - Historical Volatility Estimators (Monthly)	0.9449	29	0.6807	34	0.2931	25	0.7696	48	37	
Model 7 (M7) - Path Dependence Features	0.6068	45	0.4626	54	0.3119	50	0.7745	57	46	
Model 8 (M8) - Equity Volume Features	0.5978	60	0.4565	58	0.3247	70	0.7754	60	46	
Model 9 (M9) - Option Volume Features	0.5983	59	0.4594	46	0.3284	75	0.7760	41	31	
Model 10 (M10) - Intraday Features	0.5978	70	0.4583	64	0.3370	79	0.7756	45	36	
Model 11 (M11) - IVS Volatility Features	0.6062	46	0.4693	46	0.3195	52	0.7750	57	47	*
Model 12 (M12) - Heston Model Volatility Features	0.5925	70	0.4546	52	0.3298	66	0.7770	41	32	
Model 13 (M13) - Bates Model Volatility Features	0.6215	50	0.4803	43	0.3201	54	0.7835	46	40	
Model 14 (M14) - All Features (excluding Option Features)	0.6877	59	0.4831	57	0.3231	64	0.7716	52	39	
Model 15 (M15) - All Features	0.6119	68	0.4721	68	0.3297	71	0.7707	61	51	*

(*Continued*).

Table 4. Continued.

	RMSE		MAE		R^2		QLIKE		DM test	MCS
XGBoost	Mean	% of improv.	Mean	% of improv.	Mean	% of improv.	Mean	% of improv.	% of rejections	Included
Model 1 (M1) - HAR	0.6178	25	0.4596	23	0.1990	12	0.7214	73	44	
Model 2 (M2) - Baseline Model	0.6029	NA	0.4390	NA	0.2619	NA	0.7262	NA	NA	
Model 3 (M3) - HARQ	0.5840	59	0.4289	62	0.2628	54	0.7265	52	24	
Model 4 (M4) - Additional Daily Volatility Estimators	0.5962	59	0.4269	66	0.2795	70	0.7289	45	22	
Model 5 (M5) - Historical Volatility Estimators (Weekly)	0.5943	50	0.4379	43	0.2568	52	0.7282	43	18	
Model 6 (M6) - Historical Volatility Estimators (Monthly)	0.6228	18	0.4700	23	0.2364	30	0.7300	29	21	
Model 7 (M7) - Path Dependence Features	0.6098	36	0.4599	48	0.2528	45	0.7294	50	30	
Model 8 (M8) - Equity Volume Features	0.5993	48	0.4393	43	0.2617	57	0.7252	76	48	
Model 9 (M9) - Option Volume Features	0.5961	68	0.4348	59	0.2752	70	0.7268	55	40	
Model 10 (M10) - Intraday Features	0.5775	71	0.4276	66	0.2826	79	0.7310	30	18	
Model 11 (M11) - IVS Volatility Features	0.6094	37	0.4562	35	0.2382	41	0.7256	57	40	
Model 12 (M12) - Heston Model Volatility Features	0.6017	52	0.4413	45	0.2709	61	0.7300	41	21	
Model 13 (M13) - Bates Model Volatility Features	0.6011	61	0.4539	48	0.2758	62	0.7324	30	26	
Model 14 (M14) - All Features (excluding Option Features)	0.5891	61	0.4323	57	0.2874	75	0.7312	38	27	
Model 15 (M15) - All Features	0.5824	64	0.4201	61	0.3124	79	0.7213	69	43	

Note: For each model, we compare the performance in terms of out-of-sample R^2, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Quasi Likelihood (QLIKE). For each model, the mean of each metric is given as well as the percentage of cases where that model outperforms M2. Lastly, we provide the number of instruments (out of the full universe of 66 instruments) that the null hypothesis of the DM was rejected, signifying an improvement. In red (blue) we highlight the highest (second highest) performing method.

Table 5. Performance comparison for M1-M15, differentiating between normal and high days.

	RMSE		MAE		R^2		QLIKE			MCS	
EN	Low	High	Low	High	Low	High	Low	High	Classification Accuracy	Low	High
Model 1 (M1) - HAR	0.6078	1.5408	0.3918	0.9488	0.1635	0.0717	0.9098	2.3629	0.8054		
Model 2 (M2) - Baseline model	0.6072	1.4713	0.3910	0.8691	0.1997	0.0725	0.9440	2.9783	0.8256		
Model 3 (M3) - HARQ	0.5973	1.4693	0.3899	0.8726	0.1999	0.0687	0.9228	2.5686	0.8238		
Model 4 (M4) - Additional Daily Volatility Estimators	0.8157	2.5707	0.3990	2.9457	0.1904	0.0693	0.9223	2.7471	0.8183		
Model 5 (M5) - Historical Volatility Estimators (Weekly)	0.6084	1.5452	0.3925	0.9057	0.1962	0.0628	0.9221	2.5742	0.8209		
Model 6 (M6) - Historical Volatility Estimators (Monthly)	0.6130	1.9199	0.4005	1.2302	0.1872	0.0669	0.9234	2.5734	0.8103		
Model 7 (M7) - Path Dependence Features	0.5984	1.4794	0.3902	0.8775	0.1992	0.065	0.9328	2.7721	0.8207		*
Model 8 (M8) - Equity Volume Features	0.5977	1.4685	0.3900	0.8691	0.2017	0.0704	0.9329	2.7739	0.8267	*	
Model 9 (M9) - Option Volume Features	0.5984	2.5644	0.3913	1.3256	0.1984	0.0762	0.9395	2.9625	0.825		
Model 10 (M10) - Intraday Features	0.5962	1.4647	0.3872	0.8723	0.2148	0.0759	0.9530	3.1772	0.8235		
Model 11 (M11) - IVS Volatility Features	0.6073	1.4825	0.4014	0.8829	0.1906	0.0755	0.9238	2.5752	0.8271		*
Model 12 (M12) - Heston Model Volatility Features	0.6058	1.4529	0.3928	0.8609	0.1999	0.077	0.9658	3.3805	0.8277		
Model 13 (M13) - Bates Model Volatility Features	0.6378	1.4356	0.4197	0.8483	0.1867	0.0825	0.9638	3.1812	0.8163	*	
Model 14 (M14) - All Features (excluding Option Features)	0.5877	1.6079	0.3844	1.0018	0.2149	0.0744	0.9211	2.5590	0.8123		
Model 15 (M15) - All Features	0.6049	1.4835	0.3941	0.8897	0.2053	0.0795	0.9449	2.3570	0.8134		*
	RMSE		MAE		R^2		QLIKE			MCS	
LASSO	Low	High	Low	High	Low	High	Low	High	Classification Accuracy	Low	High
Model 1 (M1) - HAR	0.6022	1.5415	0.3920	0.9495	0.1633	0.0718	0.9098	2.5698	0.8049		
Model 2 (M2) - Baseline model	0.6011	1.4722	0.3910	0.8698	0.1997	0.0719	0.9439	2.9781	0.8252		
Model 3 (M3) - HARQ	0.5977	1.4684	0.3901	0.8720	0.1998	0.0690	0.9228	2.5689	0.8232		
Model 4 (M4) - Additional Daily Volatility Estimators	0.8372	2.7903	0.4025	9.6149	0.1884	0.0713	0.9229	2.7483	0.8199		
Model 5 (M5) - Historical Volatility Estimators (Weekly)	0.6109	1.5496	0.3932	0.9068	0.1959	0.0626	0.9222	2.5741	0.8211		
Model 6 (M6) - Historical Volatility Estimators (Monthly)	0.6117	1.8778	0.3990	1.1940	0.1865	0.0660	0.9229	2.5727	0.8135		
Model 7 (M7) - Path Dependence Features	0.5987	1.4803	0.3902	0.8790	0.1986	0.0663	0.9328	2.7715	0.8207		*
Model 8 (M8) - Equity Volume Features	0.5981	1.4690	0.3903	0.8695	0.2012	0.0705	0.9331	2.7736	0.8263	*	
Model 9 (M9) - Option Volume Features	0.5992	2.5642	0.3920	1.3255	0.1978	0.0761	0.9398	2.9624	0.8247		
Model 10 (M10) - Intraday Features	0.5952	1.4675	0.3866	0.8757	0.2143	0.0753	0.9527	3.1755	0.8206		
Model 11 (M11) - IVS Volatility Features	0.6063	1.4820	0.4004	0.8818	0.1917	0.0755	0.9235	2.5752	0.8272		*
Model 12 (M12) - Heston Model Volatility Features	0.6063	1.4528	0.3931	0.8601	0.2000	0.0756	0.9659	3.3807	0.8272		
Model 13 (M13) - Bates Model Volatility Features	0.6383	1.4361	0.4203	0.8494	0.1869	0.0827	0.9639	3.1807	0.8177	*	
Model 14 (M14) - All Features (excluding Option Features)	0.7399	1.7555	0.3932	1.0370	0.2111	0.0723	0.9206	2.5588	0.8134		
Model 15 (M15) - All Features	0.6043	1.4632	0.3946	0.8676	0.2071	0.0774	0.9455	2.5042	0.8175		*

(*Continued*).

Table 5. Continued.

	RMSE		MAE		R^2		QLIKE			MCS	
XGBoost	Low	High	Low	High	Low	High	Low	High	Classification Accuracy	Low	High
Model 1 (M1) - HAR	0.6077	1.5604	0.3959	0.9508	0.1663	0.0424	0.9114	2.3599	0.7968		
Model 2 (M2) - Baseline model	0.6113	1.5025	0.3982	0.8941	0.1806	0.0473	0.9444	2.9709	0.8080		
Model 3 (M3) - HARQ	0.6049	1.4828	0.3951	0.8822	0.1878	0.0488	0.9227	2.5641	0.8155	*	
Model 4 (M4) - Additional Daily Volatility Estimators	0.5973	1.9199	0.3922	3.0381	0.1959	0.0523	0.9241	2.7530	0.8142	*	
Model 5 (M5) - Historical Volatility Estimators (Weekly)	0.6063	1.5002	0.3967	0.8868	0.1841	0.0509	0.9236	2.5715	0.8168		
Model 6 (M6) - Historical Volatility Estimators (Monthly)	0.6350	1.5208	0.4188	0.8956	0.1597	0.0503	0.9298	2.5783	0.8049	*	
Model 7 (M7) - Path Dependence Features	0.6239	1.5073	0.4104	0.9011	0.1754	0.0527	0.9372	2.7656	0.7952		*
Model 8 (M8) - Equity Volume Features	0.6066	1.5047	0.3963	0.9030	0.1814	0.0484	0.9322	2.7647	0.8063		
Model 9 (M9) - Option Volume Features	0.6065	6.5928	0.3965	1.3475	0.1804	0.0545	0.9385	2.9554	0.8079		
Model 10 (M10) - Intraday Features	0.6173	1.4524	0.4046	0.8504	0.1985	0.0601	0.9572	3.1784	0.8215		*
Model 11 (M11) - IVS Volatility Features	0.6246	1.5074	0.4097	0.9002	0.1612	0.0514	0.9244	2.5719	0.8061		*
Model 12 (M12) - Heston Model Volatility Features	0.6168	1.4877	0.4008	0.8871	0.1838	0.0544	0.9652	3.3740	0.8058		*
Model 13 (M13) - Bates Model Volatility Features	0.6320	1.4517	0.4124	0.8557	0.1698	0.0615	0.9588	3.1767	0.8070		
Model 14 (M14) - All Features (excluding Option Features)	0.6335	1.4595	0.4222	0.8598	0.1842	0.0602	0.9296	2.5695	0.8126		
Model 15 (M15) - All Features	0.6426	1.4552	0.4274	0.8527	0.1782	0.0699	0.9521	2.7808	0.8141		*

Note: For each model, we compare the performance in terms of the out of sample R^2, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Quasi Likelihood (QLIKE). The days in the test data set are separated into 'High' and 'Low' volatility and the average performance across each type of day is given for each of the metrics. Lastly, we give the average classification accuracy of each model. In red (blue) we highlight the highest (second highest) performing method per product.

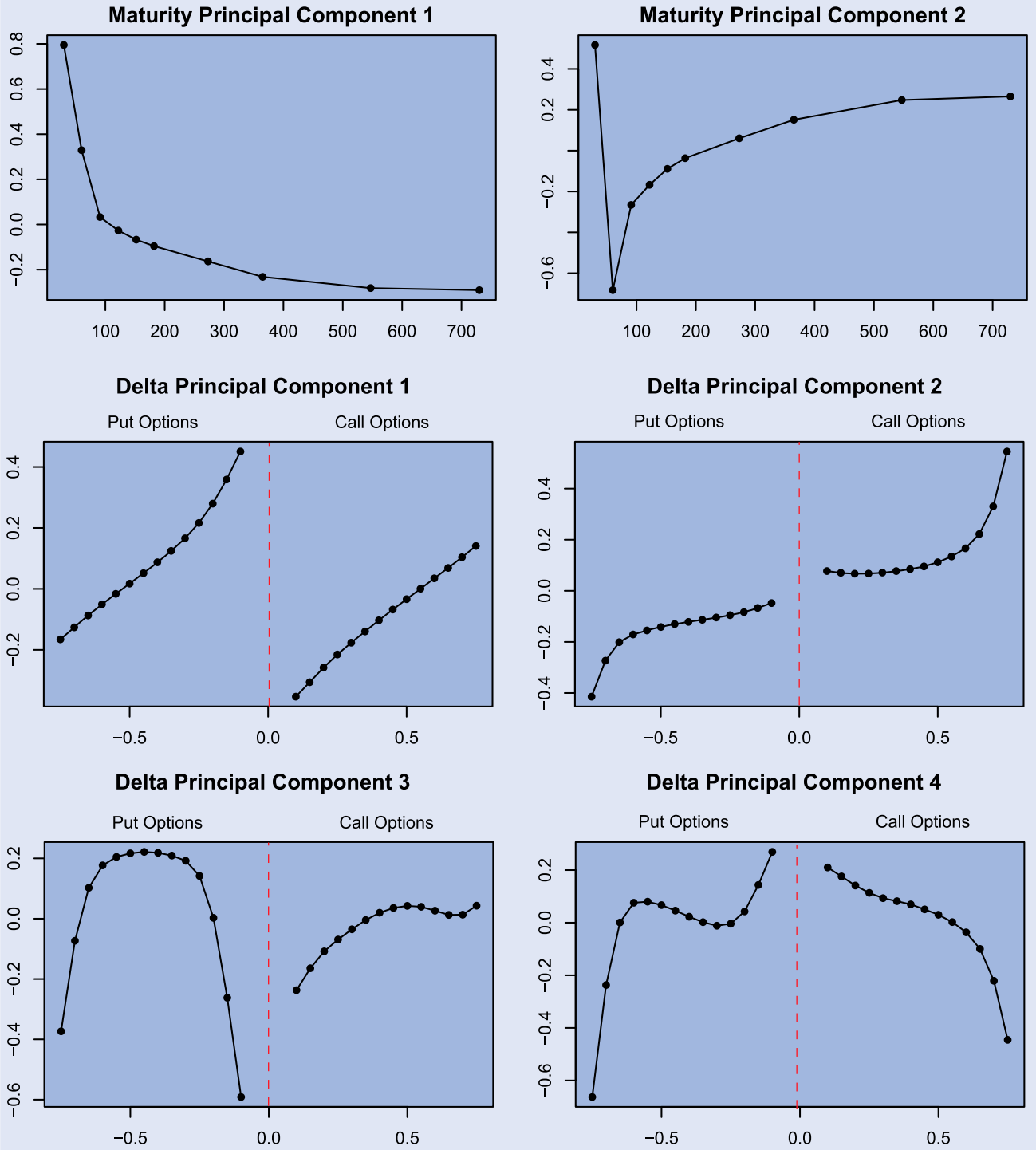

Figure 5. Plots of the first two principal components for the maturity effects, and the first four principal components for the delta effects, corresponding to model IVS-5.

In the following paragraphs, we discuss the results for each of the considered data sets shown in tables 4 and 5. In each paragraph, the effect of augmenting the HAR-based baseline model with each set of features is discussed.

Effect of Additional Daily Volatility Estimators (M3 and M4). In general, the evidence that HARQ is a significant improvement over the baseline models appears to be weak. For a large number of metrics, the percentage of improvements is close to or even below 50%, while the mean performance is below that of the baseline. This is quite different for M4, which is included in the MCS for both LASSO and EN. Many improvements can be seen across metrics, and in fact, for EN, M4 was an improvement over M2 for all instruments considered in terms of QLIKE. Augmenting the baseline model with additional volatility estimators thus appears to improve forecasting performance.

Effect of Additional Historical Volatility Estimators (M5, M6, and M7). The inclusion of weekly historical volatility estimators shows mixed results across the three models. For EN and LASSO, M5 fails to outperform the baseline in terms of

RMSE and MAE but shows improved performance in terms of QLIKE, also being included in the MCS for EN. A similar pattern holds for monthly historical volatility estimators (M6), which performs significantly worse than the baseline for most metrics, with the exception of QLIKE. Notably, M6 significantly outperforms the baseline in terms of QLIKE and scores the best in terms of mean QLIKE. Adding path dependence features (M7) does not seem to result in significant improvements. The only exception is QLIKE for the LASSO model, where the mean is slightly lower than the baseline, with a total of 46 rejections of the DM test.

Effect of Volume Features (M8 and M9). The impact of equity volume is significant in many cases. For LASSO, M8 outperforms the baseline across all metrics, being the second-best model for both RMSE and MAE, as well as improving the R^2 for 70% of the instruments and achieving 46 rejections for the DM test for QLIKE. For XGBoost, it outperforms the baseline in 76% of the cases in terms of QLIKE, with a total of 48 rejections for the DM test. For option volumes (M9), results are mixed, with the most significant improvement appearing in terms of R^2, as all three models show an increase of at least 70%.

Effect of Intraday Features (M10). An increase in performance is observed across most models and metrics, with the exception of QLIKE, where improvement rates are close to or below 50%. M10 appears to be the best or second-best performing model in several cases. It is the best-performing model in terms of RMSE, MSE, and R^2 for normal volatility days and in terms of classification accuracy when using a LASSO model.

Effect of Option Features (M11, M12, M13 and M15). The IVS features of M11 appear to only outperform the baseline in terms of QLIKE, being included in the MCS for both the LASSO and EN models, with 50 and 47 rejections of the DM test, respectively. In terms of the other metrics, it is outperformed by the baseline. This contrasts with the features from the Heston model (M12), where significant improvements can be seen across metrics, making M12 the best or second-best performing model. However, it does not outperform the baseline in terms of QLIKE. M13, which incorporates features from the Bates model, performing poorly for LASSO and EN, but showing good performance in RMSE and R^2 for XGBoost.

When comparing M14 and M15, it becomes evident that augmenting with option features significantly enhances forecasting performance. In fact, across all comparisons between the two, M14 outperforms M15 only in terms of mean MAE for EN. M15 is also the model that most consistently outperforms the baseline, with the percentage of instruments showing improvement above 60% in all cases, and in many instances, it has the highest percentage of improvements. In terms of mean QLIKE, M15 outperforms the baseline and is either the best or second-best performing model for each method. Finally, for both LASSO and EN, M15 is included in the MCS and has the highest number of rejections for the DM test.

Performance Assessment in High and Normal Volatility Regimes. In terms of QLIKE, most models outperform M2 in both normal and high-volatility regimes. For normal-volatility regimes, the MCS includes M8 and M13 for LASSO and EN, and M3, M4, and M6 for XGBoost. For high-volatility regimes, M7, M11, and M15 are included in the MCS for all methods, with M10 and M12 also in the MCS for XGBoost. The highest classification accuracy is achieved by M11 and M12 for LASSO and EN. When comparing M14 and M15, the latter improves on high-volatility regime performance and classification accuracy, while being outperformed in normal-volatility regimes.

M15 is an improvement over the baseline for more instruments than M14, and it has better mean performance for most metrics across methods. Additionally, M15 is part of the MCS in several cases. This improvement seems to result from M15's significantly better performance in high-volatility regimes, which is generally of higher importance to practitioners. This demonstrates that augmenting the HAR-based baseline with features from the option market enhances predictive power.

Feature Importance. In figure 6, we plot the feature importance by calculating the median absolute coefficient of each feature when using a LASSO method on the complete data set (M15). We further augment the complete data set (M15) with ten random features, whose entries all independently follow a standard Gaussian distribution. It can be clearly seen that a large number of features have a median coefficient higher than all of the randomly generated features. Notably, all of the different data sets used contain at least one feature scoring a higher feature importance than the ten random features.

The 15 features with the highest feature importance include the three HAR features: the previous daily, weekly, and monthly realized volatilities, as well as the realized quadricity. The top 15 features also include VIX, the sum of positive and negative price jumps (equation (73)), as well as positive and negative semivariances. The daily and monthly versions of HLP (with the weekly HLP in the top 30 features) and the Yang–Zhang monthly volatility are also included.

When it comes to features derived from option pricing, the average volatility derived from the IVS data is the second most important feature, with the volatility from the Bates model being the sixth most important feature (the volatility from the Heston model is is ranked 19th in the feature importance ranking). Lastly, the MSE feature of the Bates model is also among the 15 features with the highest importance.

In figure 7, we plot the correlation of each of the daily volatility estimators with realized volatility. We observe that the Heston volatility as well as the average IVS implied volatility are the estimators that are the least correlated with realized volatility. This is natural, as these are the only estimators that are not calculated using historical data. This observation verifies that the usage of such estimators is beneficial, as they can bring new information into the augmented HAR model.

In figure 7, we plot the correlation of the daily volatility estimators with realized volatility. We observe that the Heston, Bates and IVS implied volatility are the estimators least correlated with realized volatility. This is expected, as these are the only estimators not calculated using historical data. This observation confirms that using such estimators is beneficial, as they introduce new information into the augmented HAR model.

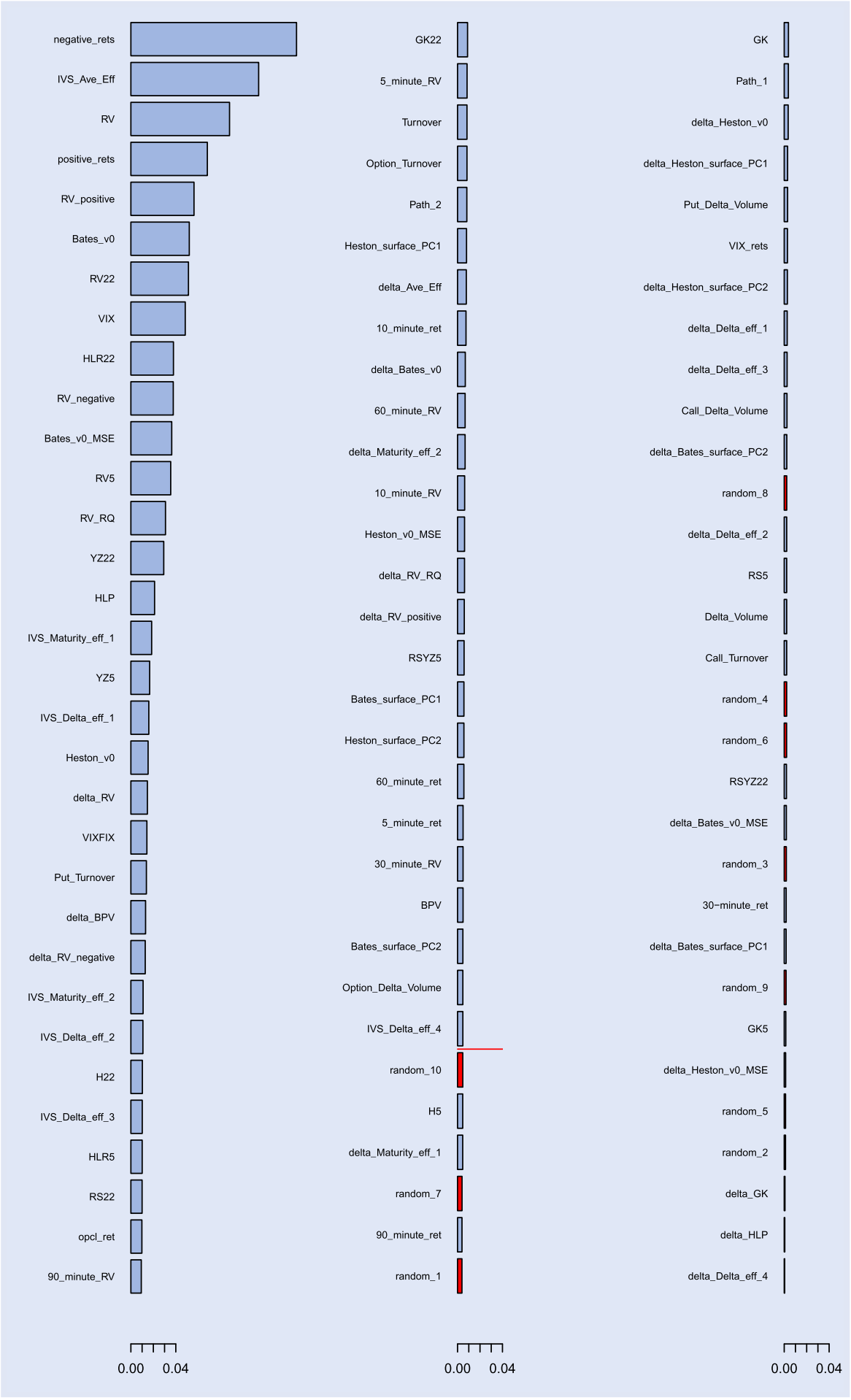

Figure 6. Barplot of the median absolute value of each feature's coefficient using an EN model, fitted with the full data set (M15). In addition, we use 10 features whose entries are all independent identically distributed N(0,1), which are shown in red color.

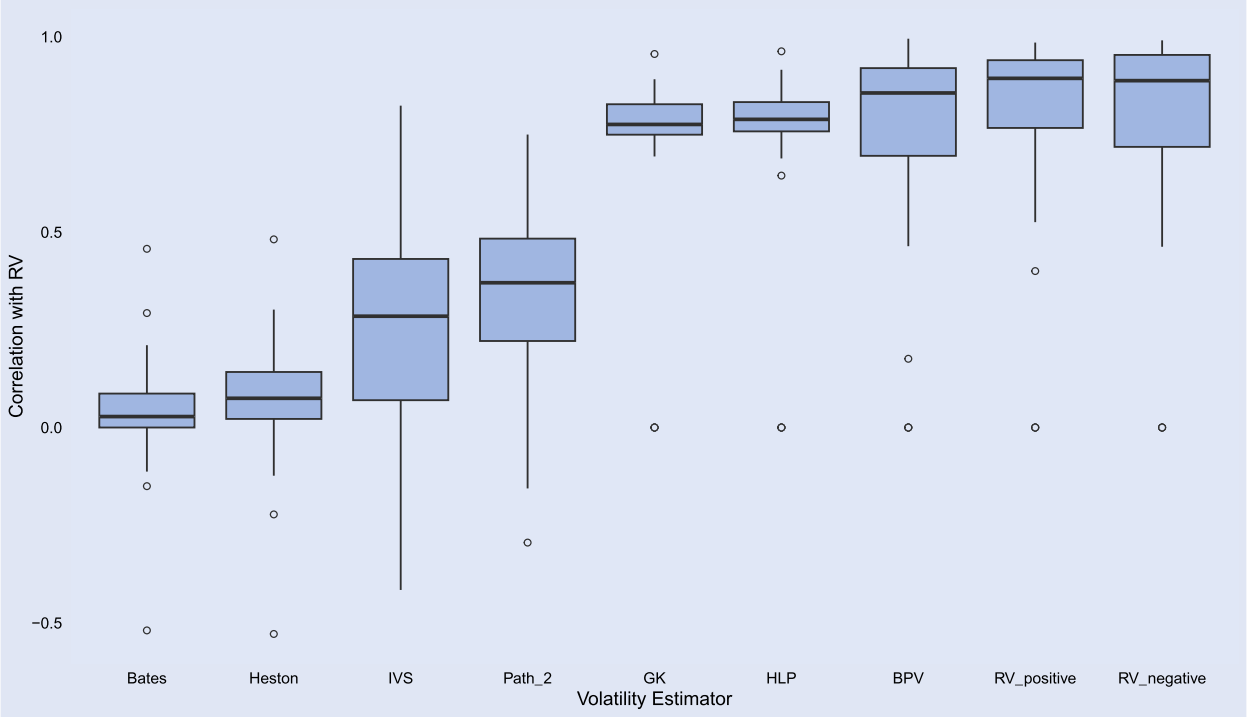

Figure 7. Box and Whisker plot of the median correlation of the different daily volatility estimators, with the corresponding value of the contemporaneous daily realized volatility.

6.1. Cross-asset model

In this section, we explore an additional volatility model based on the factors used for M15: a *cross-asset* model similar to the one used in Zhang *et al.* (2022). This model provides a different perspectives on volatility, based on the concepts of market volatility and volatility spillover.

Cross-asset models have been widely used for financial problems, such as in Cont *et al.* (2023), Lu *et al.* (2023), and Bryzgalova *et al.* (2020). High volatility in one instrument often leads to increased volatility in other instruments, a phenomenon known as volatility spillover. For the cross-asset model, we use multiple features from multiple instruments. For each asset whose volatility we predict, the model is defined as follows

$$\log(\mathrm{RV}_{i,d}) = \mathcal{F}_i\left([X^{\mathrm{Base}}_{1,d-1}, X^{\mathrm{Augm}}_{1,d-1}], \ldots, [X^{\mathrm{Base}}_{i,d-1}, X^{\mathrm{Augm}}_{i,d-1}], \ldots, [X^{\mathrm{Base}}_{N,d-1}, X^{\mathrm{Augm}}_{N,d-1}]\right).$$

As we are dealing with daily volatility and our sample size is not large enough, we are unable to use the full feature set employed in M15 for each of the instruments. Instead, we use the features from M15 for the asset's own features and then augment the model by adding two additional features for each instrument: the previous day's realized volatility and its forecast based on the M15 model. As shown in table 6, the performance of the cross-asset model is not significantly different from that of M15. It slightly outperforms M15 in terms of MAE and QLIKE, while being slightly outperformed in terms of R^2.

Forecasting VIX. While cross-asset models do not significantly improve on our ability to forecast the realized variance, we utilize them for a different task. Instead of forecasting realized variance for individual stocks as have done so far, we shortly turn our attention to forecasting VIX, a theoretical quantity based on the volatility of SPX, the index of the S&P market. This renders VIX a natural forecasting task for the cross-asset model. To this end, we compare three different models, M2 the baseline model, M15 the full model and the cross-asset model. The predictors are calculated based on data for the SPX index, as VIX is intrinsically linked to its variance. However, note that as SPX is not directly tradable, the traded volume features are excluded from M15 and the cross-asset model.

In table 7, we show the performance of the three models for VIX. We note that unlike RV forecasting, where the increase in performance between M2 and M15 was marginal, the difference in this case is substantial, with RMSE dropping from 2.415 down to 2.004. The cross-asset model further increases performance, as it outperforms M15 across all metrics.

Economic Application. As an economic application, we consider the VIX Mid-Term Futures ETF (VIXM). Instead of using VIX as the target, we instead forecast the next day open-close returns of VIXM, using the forecasts from M15 for stocks as predictors. We formulate a simple theoretical trading strategy, whereby we open a position in VIXM depending on the sign of the forecast on the open-close returns of VIXM of the next trading day, which we close on the closing of next day.

In figure 8, we plot the cumulative P&L. It can be clearly seen that following a cross-asset model based on the forecasts of the M15 model, results in a profitable strategy, which highly outperforms the baseline (longing and holding VIXM). This comes in contrast to a strategy based on forecasts from M14, which scores very poorly in terms of both Sharpe Ratio and Profit per Day (eg Profit & Loss returns). In fact,

Table 6. Comparison of the performances of different models for the forecasting of realized variance.

	RMSE		MAE		R^2		QLIKE		DM Test
	Mean	% of improv.	Mean	% of improv.	Mean	% of improv.	Mean	% of improv.	% of rejections
Model 2 (M2) - Baseline model	0.6029	NA	0.4390	NA	0.2619	NA	0.7262	NA	NA
Model 15 (M15) - All Features	0.5824	64	0.4201	61	0.3124	79	0.7213	69	43
Cross Asset Model	0.5821	67	0.4111	58	0.3109	79	0.7200	72	43

Note: For each model, we compare the performance in terms of the out-of-sample R^2, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE), and Quasi Likelihood (QLIKE). For each model, the average of these quantities is shown, as well as the percentage of cases where that model outperforms M2. Lastly, we provide the number of instruments out of 66 that the null hypothesis of the DM was rejected, signifying an improvement over M2.

Table 7. Comparison of performance of M2, M15 and the cross-asset model, for forecasting VIX.

	RMSE	MAE	R^2	DM test (d-value)
Model 2 (M2) - Baseline	2.415	1.603	0.691	NA
Model 15 (M15) - Complete	2.004	1.193	0.787	0.000
Cross-Asset Model	1.923	1.173	0.794	0.000

Note: For each model, we compare the performance in terms of the out-of-sample R^2, Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE).

both M2 and M14 are outperformed by a strategy, where VIXM is shorted. This confirms that not only are the features included in M15 beneficial for volatility forecasting, but they can also bring further economic benefits when employed for forecasting VIX.

7. Conclusion

We have augmented the well-established HAR model with various additional estimators of the daily, weekly, and monthly volatility, as well as other predictive features, to enhance the performance for daily realized volatility forecasting. We have shown significant improvements in performance, with each considered augmentation adding predictive power to the HAR-based baseline model.

A major focus has been the generation of predictors from the options market. We have provided a novel methodology to derive volatility information from the options market by assuming different models. We have shown how deriving Delta and maturity effects before applying dimensionality reduction to implied volatility surface data can result

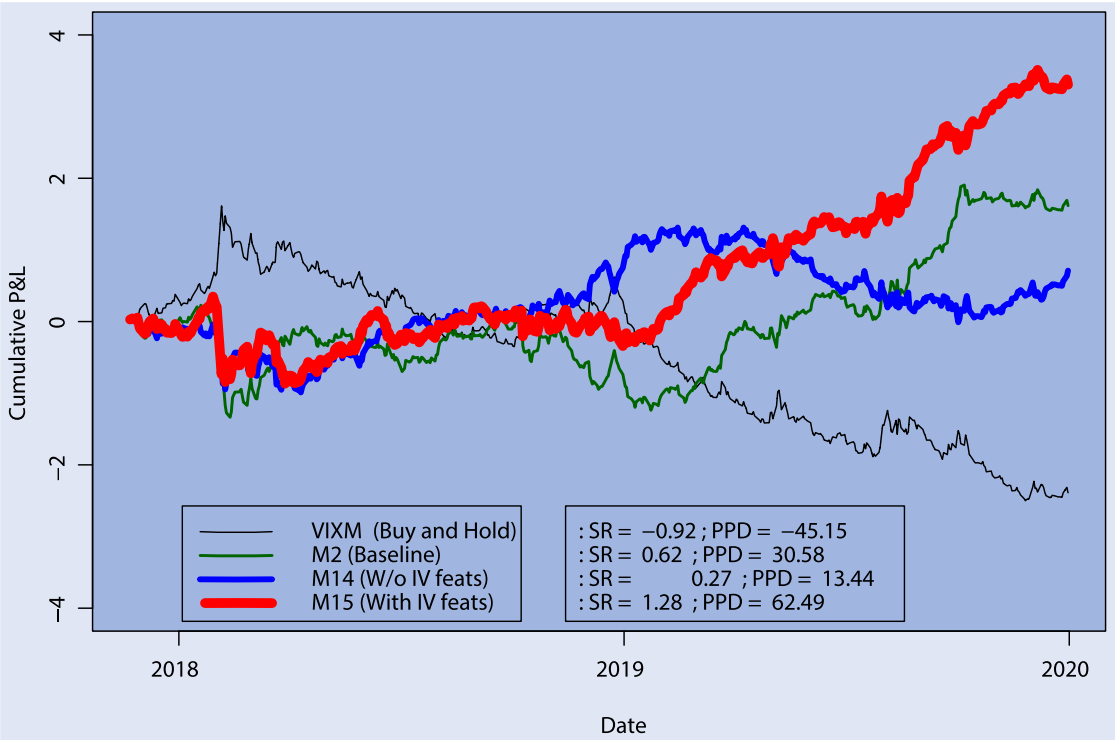

Figure 8. Cumulative P&L plot for VIXM. SR denotes the Sharpe Ratio and PPD the P&L return per Day, which are defined in appendix 2.

in more informative predictors. We have also generated features assuming the Heston and Bates models, by following a simplified calibration process.

Unlike the majority of other daily volatility proxies and trackers, the features extracted from the options market are weakly correlated to their contemporaneous realized volatility, indicating how they augment existing models in a meaningful manner, as information that does not originate from historical equity prices. Models M11-M13 (which include the option-extracted features) show a significant increase in performance with respect to most metrics. Furthermore, we observe that the extracted volatility estimators from both the IVS and the calibrated current volatility assuming non-stochastic volatility models, are both shown to have high importance. The first maturity effect, as well as the first two delta effects extracted from implied volatility surface data, also score highly in terms of feature importance.

Lastly, we demonstrated how the individual forecasts received from all the individual equity volatilities can be employed as predictors for boosting the forecasting performance of the market volatility, as they are able to offer predictive power for VIX. This augmentation appears to be useful not only for the untradable VIX index, but also for the tradable VIXM.

Future Work. Future work could extend the methodology to dividend-paying equities. Unlike equities that do not pay dividends, such equities can not be parallelized with European options, often priced via the use of binomial trees (see, e.g.Burkovska *et al.* 2016). As the resulting calibration are more unstable, further consideration is required, inviting future research.

Another angle for further research could include the calibration of the option pricing models via deep learning. For example, a promising pipeline is offered by Leite *et al.* (2021), who train an unstacked network for each of the Heston parameters using extracted features from simulated price time series data and principal component analysis. An alternative method could be the usage of the inverse map approach (see, e.g. the discussion in Itkin 2019), whereby the calibration process is modeled as a Deep Neural Network, where an increase in computational speed is achieved by replacing the intensive pricing of options with an Artificial Neural Network approximation. Despite making the pipeline more complicated, neural networks provide instant estimates for new data once trained. Furthermore, the toolbox offered by neural network methods can be easily applied, such as regularization to make the model less prone to overfitting.

Other natural extensions could include further enhancement of our model using news sentiment data, as well as the usage of more complex models such as Moreno-Pino and Zohren (2022), which was used for high frequency volatility forecasting.

Acknowledgment

We would like to express our sincere gratitude to Maria Girth for her detailed and constructive feedback.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the UK Engineering and Physical Sciences Research Council (EPSRC) grants EP/R513295/1 and EP/V520202/1.

References

Aït-Sahalia, Y., Li, C. and Li, C., Closed-form implied volatility surfaces for stochastic volatility models with jumps. *J. Econom*, 2021, **222**(1), 364–392.

Andersen, T.G. and Bollerslev, T., Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *Int. Econ. Rev. (Philadelphia)*, 1998, **39**(4), 885–905.

Audrino, F. and Knaus, S.D., Lassoing the har model: A model selection perspective on realized volatility dynamics. *Econom. Rev*, 2016, **35**(8-10), 1485–1521.

Audrino, F., Sigrist, F. and Ballinari, D., The impact of sentiment and attention measures on stock market volatility. *Int. J. Forecast*, 2020, **36**(2), 334–357.

Avellaneda, M., Healy, B., Papanicolaou, A. and Papanicolaou, G., PCA for implied volatility surfaces, 2020.

Badshah, I., Modeling the dynamics of implied volatility surfaces. *SSRN Electronic Journal*, 2009.

Barndorff-Nielsen, O.E. and Shephard, N., Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)*, 2002, **64**(2), 253–280.

Barndorff-Nielsen, O.E. and Shephard, N., Power and bipower variation with stochasticvolatility and jumps. *J. Financ. Econom.*, 2004, **2**(1), 1–37.

Barndorff-Nielsen, O.E., Kinnebrouk, S. and Shephard, N., *Measuring Downside Risk: Realised Semivariance*, edited by T. Bollerslev, J. Russell, and M. Watson, pp. 117–136, 2010 (Oxford University Press: Oxford).

Bates, D.S., Jumps and stochastic volatility: Exchange rate processes implicit in deutsche mark options. *Rev. Financ. Stud*, 1996, **9**(1), 69–107.

Będowska-Sójka, B. and Kliber, A., Information content of liquidity and volatility measures. *Physica A: Stat. Mech. Appl.*, 2021, **563**, 125436.

Bollerslev, T., Generalized autoregressive conditional heteroskedasticity. *J. Econom*, 1986, **31**(3), 307–327.

Bollerslev, T., Patton, A.J. and Quaedvlieg, R., Exploiting the errors: A simple approach for improved volatility forecasting. *J. Econom*, 2016, **192**(1), 1–18.

Bouri, E., Lau, C.K.M., Lucey, B. and Roubaud, D., Trading volume and the predictability of return and volatility in the cryptocurrency market. *Fin. Res. Lett.*, 2019, **29**, 340–346.

Bryzgalova, S., Pelger, M. and Zhu, J., Forest through the trees: Building cross-sections of stock returns. Working Paper, 2020. 10.2139/ssrn.3493458

Bucci, A., Realized volatility forecasting with neural networks. *J. Financ. Econom.*, 2020, **18**(3), 502–531.

Burkovska, O., Gaß, M., Glau, K., Mahlstedt, M., Schoutens, W. and Wohlmuth, B., Calibration to american options: Numerical investigation of the de-americanization, 2016.

Canina, L. and Figlewski, S., The informational content of implied volatility. *Rev. Financ. Stud*, 1993, **6**(3), 659–681.

Chen, W.-J., Yao, J.-J. and Shao, Y.-H., Volatility forecasting using deep neural network with time-series feature embedding. *Econ. Res.-Ekon. Istraž.*, 2023, **36**(1), 1377–1401.

Chicheportiche, R. and Bouchaud, J.-P., The fine-structure of volatility feedback I: Multi-scale self-reflexivity. *Physica A: Stat. Mech. Appl.*, 2014, **410**, 174–195.
Chou, R.Y., Chou, H. and Liu, N., *Range Volatility Models and Their Applications in Finance*, pp. 1273–1281. 2010 (Springer US: Boston, MA).
Christensen, K., Siggaard, M. and Veliyev, B., A machine learning approach to volatility forecasting*. *J. Financ. Econom.*, 2022, **21**(5), 1680–1727.
Clark, P.K., A subordinated stochastic process model with finite variance for speculative prices. *Econometrica*, 1973, **41**(1), 135–155.
Cont, R. and da Fonseca, J., Dynamics of implied volatility surfaces. *Quant. Finance*, 2002, **2**(1), 45–60.
Cont, R., Cucuringu, M. and Zhang, C., Cross-impact of order flow imbalance in equity markets. *Quant. Finance*, 2023, **23**(10), 1373–1393.
Corsi, F., A simple approximate long-memory model of realized volatility. *J. Financ. Econom.*, 2009, **7**(2), 174–196.
Cui, Y., Rollin, S. d. B. and Germano, G., Full and fast calibration of the heston stochastic volatility model, 2015.
Dai, Z., Zhou, H., Wen, F. and He, S., Efficient predictability of stock return volatility: The role of stock market implied volatility. *N. Am. J. Econ. Finance*, 2020, **52**, 101174.
Diebold, F. and Mariano, R., Comparing predictive accuracy. *J. Bus. Econ. Stat*, 2002, **20**, 134–44.
Dobi, D., Modeling systemic risk in the options market. PhD Thesis, New York University, 2014.
Dumas, B., Fleming, J. and Whaley, R., Implied volatility functions: Empirical tests. *J. Finance*, 2001, **53**, 2059–2106.
Dvorsky, J., Belas, J., Gavurova, B. and Brabenec, T., Business risk management in the context of small and medium-sized enterprises. *Econ. Res.-Ekon. Istraž.*, 2021, **34**(1), 1690–1708.
Economics, T.E.C., A model of asset trading under the assumption of sequential information arrival. *J. Finance*, 1976, **31**(4), 1149–1168.
Engle, R., Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 1982, **50**(4), 987–1007.
Fang, Y., Lysy, M. and Mcleish, D., Common-factor stochastic volatility modelling with observable proxy. *Can. J. Stat.*, 2020, **48**(1), 36–61.
Figlewski, S., Estimating the implied risk-neutral density for the US market portfolio. In *Volatility and Time Series Econometrics: Essays in Honor of Robert Engle*, 2010 (Oxford University Press: Oxford).
Fischer, B., Studies in stock price volatility changes. *Proc. Am. Stat. Assoc. Bus. Econ. Stat.*, 1976, **64**, 171–181.
Fukasawa, M., Asymptotic analysis for stochastic volatility: Martingale expansion. *Finance Stoch.*, 2011, **15**, 635–654.
Garman, M.B. and Klass, M.J., On the estimation of security price volatilities from historical data. *J. Bus.*, 1980, **53**(1), 67–78.
Ghysels, E., Harvey, A. and Renault, E., Stochastic volatility. Cahiers de recherche, Centre interuniversitaire de recherche en Ãcconomie quantitative, CIREQ, 1996.
Guyon, J. and Lekeufack, J., Volatility is (mostly) path-dependent. *Quant. Finance*, 2023, **23**(9), 1221–1258.
Han, Y. and Liu, F., The information content of the implied volatility surface: Can option prices predict jumps? *SSRN Electronic Journal*, 2019.
Hansen, P.R., Lunde, A. and Nason, J.M., The model confidence set. *Econometrica*, 2011, **79**(2), 453–497.
Herskovic, B., Kelly, B., Lustig, H. and Van Nieuwerburgh, S., The common factor in idiosyncratic volatility: Quantitative asset pricing implications. *J.Financ.Econ.*, 2016, **119**(2), 249–283.
Heston, S.L., Aclosed-form solution for options with stochastic volatility with applications to bond and currency options. *Rev. Financ. Stud*, 2015, **6**(2), 327–343.
Itkin, A., Deep learning calibration of option pricing models: Some pitfalls and solutions, 2019.
İltüzer, Z., Option pricing with neural networks vs. black-scholes under different volatility forecasting approaches for bist 30 index options. *Borsa Istanb. Rev.*, 2022, **22**(4), 725–742.
Jorion, P., Predicting volatility in the foreign exchange market. *J. Finance*, 1995, **50**(2), 507–528.
Kakushadze, Z. and Serur, J.A., *151 Trading Strategies*, 2018 (Springer: Cham).
Kambouroudis, D. and Mcmillan, D., Does vix or volume improve garch volatility forecasts? *Appl. Econ*, 2015, **48**, 1–19.
Kennedy, J. and Eberhart, R., Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, Vol. 4, Perth, Australia, pp. 1942–1948, 1995.
Kim, J.-M., Kim, D.H. and Jung, H., Applications of machine learning for corporate bond yield spread forecasting. *N. Am. J. Econ. Finance*, 2021, **58**, 101540.
Kingma, D.P. and Ba, J., Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
Korkusuz, B., Kambouroudis, D. and McMillan, D.G., Do extreme range estimators improve realized volatility forecasts? Evidence from G7 stock markets. *Finance Res. Lett.*, 2023, **55**, 103992.
Lamoureux, C.G. and Lastrapes, W.D., Forecasting stock-return variance: Toward an understanding of stochastic implied volatilities. *Rev. Financ. Stud*, 2015, **6**(2), 293–326.
Leite, I.M.S., Yamim, J.D.M. and da Fonseca, L.G., The deeponets for finance: An approach to calibrate the heston model. In *Progress in Artificial Intelligence*, edited by G. Marreiros, F. S. Melo, N. Lau, H. Lopes Cardoso, and L. P. Reis, pp. 351–362, 2021 (Springer International Publishing, Cham).
Liu, Y., Novel volatility forecasting using deep learning–long short term memory recurrent neural networks. *Expert. Syst. Appl*, 2019, **132**, 99–109.
Liu, M., Choo, W.-C., Lee, C.-C. and Lee, C.-C., Trading volume and realized volatility forecasting: Evidence from the china stock market. *J. Forecast*, 2023, **42**(1), 76–100.
LOBSTER, Limit order book system - the efficient reconstructor at humboldt universität zu berlin, germany, 2023. http://LOBSTER.wiwi.hu-berlin.de.
Lu, Y., Reinert, G. and Cucuringu, M., Co-trading networks for modeling dynamic interdependency structures and estimating high-dimensional covariances in us equity markets, 2023.
Luong, C. and Dokuchaev, N., Forecasting of realised volatility with the random forests algorithm. *J. Risk Financ. Manag.*, 2018, **11**(4), 61.
Maghyereh, A., Abdoh, H. and Awartani, B., Have returns and volatilities for financial assets responded to implied volatility during the covid-19 pandemic? *J. Commod. Mark.*, 2022, **26**, 100194.
Mandelbrot, B., The variation of certain speculative prices. *J. Bus.*, 1963, **36**, 394–394.
Mittnik, S., Robinzonov, N. and Spindler, M., Stock market volatility: Identifying major drivers and the nature of their impact. *J. Bank. Financ*, 2015, **58**, 1–14.
Moré, J.J., The levenberg-marquardt algorithm: Implementation and theory. In *Numerical Analysis*, edited by G.A. Watson, pp. 105–116, 1978. (Springer Berlin Heidelberg: Berlin, Heidelberg).
Moreno-Pino, F. and Zohren, S., Deepvol: Volatility forecasting from high-frequency data with dilated causal convolutions, 2022.
Mrázek, M. and Pospíšil, J., Calibration and simulation of heston model. *Open Math.*, 2017, **15**(1), 679–704.
Mrazek, M., Pospíšil, J. and Sobotka, T., On calibration of stochastic and fractional stochastic volatility models. *Eur. J. Oper. Res*, 2016, **254**, 679–704.
Naimoli, A., Gerlach, R. and Storti, G., Improving the accuracy of tail risk forecasting models by combining several realized volatility estimators. *Econ. Model*, 2022, **107**, 105701.
OptionMetrics, Optionmetrics: Ivydb us option data. Wharton Research Data Services, 2023. https://wrds-web.wharton.upenn.edu/wrds/.
Parkinson, M.H., The extreme value method for estimating the variance of the rate of return. *J. Bus.*, 1980, **53**, 61–65.

Patton, A.J. and Sheppard, K., Optimal combinations of realised volatility estimators. *Int. J. Forecast*, 2009, **25**(2), 218–238. Forecasting Returns and Risk in Financial Markets using Linear and Nonlinear Models.

Pérez-Rodríguez, J.V., Another look at the implied and realised volatility relation: A copula-based approach. *Risk Manag.*, 2020, **22**(1), 38–64.

Reisenhofer, R., Bayer, X. and Hautsch, N., Harnet: A convolutional neural network for realized volatility forecasting, 2022.

Rogers, L. and Satchell, S., Estimating variance from high, low and closing prices. *Ann. Appl. Probab*, 1991, **1**, 504–512.

Schoutens, W., Simons, E. and Tistaert, J., A perfect calibration! now what? *Wilmott*, 2004, **2004**, 66–78.

Sharpe, W.F., Capital asset prices: A theory of market equilibrium under conditions of risk. *J. Finance*, 1964, **19**(3), 425–442.

Visser, M., Ranking and combining volatility proxies for garch and stochastic volatility models, 2008.

Wei, Y., Liang, C., Li, Y., Zhang, X. and Wei, G., Can cboe gold and silver implied volatility help to forecast gold futures volatility in china?Evidence based on har and ridge regression models. *Finance Res. Lett.*, 2020, **35**, 101287.

Williams, L., The vix fix. pp. 24–32, 2007.

Xing, Y., Zhang, X. and Zhao, R., What does the individual option volatility smirk tell us about future equity returns? *J. Financ. Quant. Anal.*, 2010, **45**(3), 641–662.

Yang, D. and Zhang, Q., Drift-independent volatility estimation based on high, low, open, and close prices. *J. Bus.*, 2000, **73**(3), 477–492.

Zhang, C., Zhang, Y., Cucuringu, M. and Qian, Z., Volatility forecasting with machine learning and intraday commonality, 2022.

Zumbach, G., Volatility conditional on price trends. *Quant. Finance*, 2010, **10**(4), 431–442.

Appendices

Appendix 1. Heston model calibration

In this section we describe how to fully fit the Heston model described in section 4, for the initial parameter estimates. Our methodology uses a combination of the particle swarm optimization (PSO) method and a gradient descend method (in this case ADAM). This way we take advantage of the faster gradient method, and employ the slower (but less prone to getting stuck in local minima) PSO only when the obtained parameter set is a bad fit.

Specifically, the parameter set for the first day of the data set is found using PSO. Then for each subsequent day a gradient descent method is used with the starting parameter vector being equal to the one fitted for the previous trading day. If the obtained RMSE (see equation equation (49)) is lower than a pre-defined threshold $\text{RMSE}_{\max}$, we accept it, otherwise we run PSO. Finally, between the parameter vectors obtained using PSO and the gradient method, we choose the one that corresponds to a lower RMSE.

In the rest of this section we give a description for each of the two methods described. Note that the function to be minimized is the RMSE as defined in equation (49).

$$f(\theta_d) = \sqrt{\frac{1}{N_d}\sum_{i=1}^{N_d} w_i r_i(\theta_d)^2}. \tag{A1}$$

For convenience, we drop the day index d.

A.1. Particle swarm optimization

Particle Swarm Optimization (Kennedy and Eberhart 1995) (PSO) is a computational optimization method, that does not require the underlying problem to be convex, and can find a solution even when there is a multitude of local minima. A closed box (Cartesian product of closed intervals) U within which a solution needs to be specified. Thus, with this restriction, we aim to solve $\min_{\theta\in U} f(\theta)$. The algorithm works by considering moving particles $p_1, \ldots, p_N \in U$. The central idea is that each of these particles explores the parameter space U, with varying velocities $v_1, \ldots, v_N > 0$. The multiplicity of particles ensures sufficient exploration, thus avoiding local optima. Each of the particles' velocities depends on the optimal solution encountered by that particle so far, while the velocity of the swarm depends on the overall optimal solution. In algorithm 1, we give the complete algorithm we use.

ADAM. Out of the large family of gradient methods, we opt to use ADAM (Kingma and Ba 2014), a stable in optimization, due its simplicity and efficiency. This is an iterative algorithm which starts with a parameter vector $\theta^{(0)}$, and iteratively carries out the following updating step

$$\theta_j^{(n)} = \theta_j^{(n-1)} - \eta \frac{\frac{m_j^{(n)}}{1-\beta_1^n}}{\sqrt{\frac{v_j^{(n)}}{1-\beta_2^{(n)}}} + \epsilon}, \tag{A2}$$

$$m_j^{(n)} = \beta_1 m_j^{(n-1)} + (1-\beta_1)\nabla_j f\left(\theta^{(n-1)}\right), \tag{A3}$$

$$v_j^{(n)} = \beta_2 v_j^{(n-1)} + (1-\beta_2)\left(\nabla_j f\left(\theta^{(n-1)}\right)\right)^2, \tag{A4}$$

where we set $m_j^{(0)} = v_j^{(0)} = 0, \forall j$. The update step is repeated until one of the following terminal criteria is met.

- $f(x) \le \epsilon_1$
- $|\theta_n - \theta_{n-1}| \le \epsilon_2$
- $n \ge n_{\max}$

We use the configuration of the original ADAM paper, with $\eta = 0.001, \epsilon = 10^{-8}, \beta_1 = 0.9, \beta_2 = 0.999$. For the terminal conditions, we use $n_{\max} = 500, \epsilon_1 = 0.01, \epsilon_2 = 10^{-4}$.

Appendix 2. Financial trading metrics

In section 6.1 we consider a hypothetical trading strategy. In this section, we shortly define the financial metrics used in figure 8. Let P_d^{open} and P_d^{close} denote the opening and closing price of an asset for day d. The daily open to close return is defined as

$$f_d = \frac{P_{d+1}^{\text{close}} - P_{d+1}^{\text{open}}}{P_{d+1}^{\text{open}}}. \tag{A5}$$

Denote by $\hat{f}_d$, the forecast for f_d. The Profit and Loss (P&L) metric captures the total profit that is achieved by following the set of forecasts $\{\hat{f}_d\}_{d=1}^D$, for a trading strategy where as the underlying instrument is longed (shorted) whenever the sign of $\hat{f}_d$ is positive (negative).

$$\text{P\&L}_d = \sum_{d=1}^{D} f_d \operatorname{sign}\left(\hat{f}_d\right). \tag{A6}$$

A measure of the performance of a strategy is the Sharpe Ratio (Sharpe 1964), defined as

$$\tilde{\text{SR}} = \frac{\text{mean}(\mathbf{P\&L})\sqrt{252}}{\text{sd}(\mathbf{P\&L})}. \tag{A7}$$

We also define the P&L return per Day (PPD) as

$$\text{PPD} = \frac{\sum_{d=1}^{D-1} \text{P\&L}_d}{D-1}. \tag{A8}$$

It should be noted that we do not consider a trading strategy in a realistic setting, as the trading costs and liquidity constraints are ignored.

Algorithm 1: Particle Swarm Optimization

Input: U (parameter space), p (number of particles), w (inertia parameter), ϕ_p, ϕ_g (velocity update parameters), $t_{\max}, \epsilon, t_{\text{conv}}$ (convergence parameters)
Output: θ^*

```
Initialize t ← 1, C1 = C2 = C3 = 0
for i ← 1 to p do
    Initialize the particle's position x_{1,i} by uniformly drawing a point from U
    Initialize the particle's velocity v_{1,i} to 0
    Initialize the particle's best known position to ρ_{1,i} ← x_{1,i}
end for
Initialize the swarm's best known position to g_1 ← argmax_{x∈{ρ_{1,1},...,ρ_{1,p}}} f(x)
while C1 + C2 + C3 = 0 (Termination Criteria are not met) do
    for i ← 1 to p do
        Draw r_p, r_g ~ U(0, 1)
        Update the particle's velocity as v_{t+1,i} ← w v_{t,i} + φ_p r_p (ρ_{t,i} − x_{t,i}) + φ_g r_g (g_t − x_{t,i})
        Update the particle's position as x_{t+1,i} ← x_t + v_{t+1,i}
        Update the particle's best known position as p_{t+1,i} ← argmax_{x∈{p_{t,i}, x_{t+1,i}}} f(x)
    end for
    Update the swarm's best known position as g_{t+1} ← argmax_{x∈{ρ_{t+1,1},...,ρ_{t+1,p}}} f(x)
    t ← t + 1;  C1 ← 1_{t≥t_max};  C2 ← 1_{f(g_{t+1})≥ε};  C3 ← 1_{g_{t+1}=g_{t+1−t_conv}}
end while
Set θ* = g_t
```