International Journal of Forecasting 37 (2021) 1092–1110

Contents lists available at ScienceDirect

International Journal of Forecasting

journal homepage: www.elsevier.com/locate/ijforecast

Stock market volatility forecasting: Do we need high-frequency data?

Štefan Lyócsa [a,b,*], Peter Molnár [c,d,e], Tomáš Výrost [a,f]

[a] Institute of Financial Complex Systems, Masaryk University, Brno, Czech Republic
[b] Faculty of Management, University of Presov, Presov, Slovakia
[c] UiS Business School, University of Stavanger, Stavanger, Norway
[d] Faculty of Finance and Accounting, Prague University of Economics and Business, Prague, Czech Republic
[e] Faculty of Economic Sciences and Management, Nicolaus Copernicus University in Toruń, Toruń, Poland
[f] Faculty of Commerce, University of Economics, Bratislava, Slovakia

ARTICLE INFO

Keywords:
Volatility
Forecasting
Realized volatility
High–low range
HAR model

ABSTRACT

The general consensus in the volatility forecasting literature is that high-frequency volatility models outperform low-frequency volatility models. However, such a conclusion is reached when low-frequency volatility models are estimated from daily returns. Instead, we study this question considering daily, low-frequency volatility estimators based on open, high, low, and close daily prices. Our data sample consists of 18 stock market indices. We find that high-frequency volatility models tend to outperform low-frequency volatility models only for short-term forecasts. As the forecast horizon increases (up to one month), the difference in forecast accuracy becomes statistically indistinguishable for most market indices. To evaluate the practical implications of our results, we study a simple asset allocation problem. The results reveal that asset allocation based on high-frequency volatility model forecasts does not outperform asset allocation based on low-frequency volatility model forecasts.

1. Introduction

Volatility modeling and forecasting are an integral part of finance and play a crucial role in various financial applications, such as risk management and hedging. Historically, the first volatility models were autoregressive heteroscedasticity (ARCH) and generalized autoregressive heteroscedasticity (GARCH) models, introduced in the 1980s by Bollerslev (1986) and Engle (1982). GARCH models and the various modifications are likely still the most popular volatility models among academics and practitioners alike; see, e.g., Klein and Walther (2016).

Almost all GARCH models are associated with daily, close-to-close returns, or with even lower-frequency data requirements.[1] However, daily squared returns are a noisy proxy for true volatility (Molnár, 2012). As recognized by Garman and Klass (1980) and Parkinson (1980), the range (the difference between the highest and lowest price of the day) leads to more precise volatility estimates and should thus lead to more precise volatility models. However, it took a relatively long time for range-based volatility estimators to become incorporated into volatility models (Alizadeh et al., 2002; Brandt & Jones, 2006; Chou, 2005; Chou & Liu, 2010; Gallant et al., 1999). These estimators are currently used both in the creation of new volatility models (Fiszeder & Fałdziński, 2019; Fiszeder et al., 2019) and in applications (Kim et al., 2019).

* Corresponding author at: Institute of Financial Complex Systems, Masaryk University, Brno, Czech Republic.
E-mail address: stefan.lyocsa@gmail.com (Š. Lyócsa).

[1] Exceptions are GARCH models that also use an additional information source, such as realized volatility estimated from high-frequency data (Hansen et al., 2012) or the high–low range (Molnár, 2016).

https://doi.org/10.1016/j.ijforecast.2020.12.001
0169-2070/

Once high-frequency data became available, researchers recognized that these data are even more informative regarding volatility, and the concept of realized volatility emerged (Andersen, Bollerslev, Diebold, Ebens, 2001; Andersen, Bollerslev, Diebold, & Labys, 2001; Barndorff-Nielsen & Shephard, 2002). Realized volatility quickly found its way into the volatility modeling and forecasting literature (Andersen et al., 2003) and has since become popular, not only in volatility models (Aalborg et al., 2019; Degiannakis et al., 2020; Haugom et al., 2014; Luo et al., 2019; Lyócsa & Molnár, 2018; Lyócsa & Todorova, 2020; Ma et al., 2017; Wang et al., 2016), but also in price forecasting (Degiannakis & Filis, 2018).

There is a general consensus in the literature that volatility models based on high-frequency data perform better at forecasting than models based solely on daily returns (Andersen et al., 2003; Chortareas et al., 2011; Horpestad et al., 2019; Koopman et al., 2005; Martens, 2001; Wei, 2012). Martens (2001) find that in forecasting daily volatility, GARCH models based on intraday returns outperform daily GARCH models. Andersen et al. (2003) show that realized volatility models outperform standard GARCH models. The same conclusion is reached by Horpestad et al. (2019). Chortareas et al. (2011) conclude that using high-frequency data enhances the performance of volatility forecasts relative to volatility models based on daily returns (GARCH and stochastic volatility models). Similarly, Koopman et al. (2005) and Wei (2012) find that realized volatility models outperform stochastic volatility and GARCH-class models. Ma et al. (2018) study whether day-ahead volatility forecasts of the Shanghai Stock Exchange Composite Index and S&P 500 Index can be improved by combining low- and high-frequency volatility models. Indeed, low-frequency volatility models might still be useful, as suggested by the results of Ma et al. (2018), who showed that forecasts from low-frequency volatility models can be combined with forecasts from high-frequency volatility models to improve the forecasting accuracy. A similar conclusion is reached by Zhang et al. (2019), who find that both GARCH models and realized volatility models are outperformed by a smart combination of these models: a model that simply selects among the forecasts from these models the one with a relatively good past performance. With regard to evaluations of volatility forecasts, Bailey and Steeley (2019) show that squared returns are a poor proxy for forecast evaluation, and that realized volatility or the (Parkinson, 1980) estimator should be used instead.

However, most of the comparisons between high-frequency volatility models and low-frequency volatility models consider only low-frequency volatility models estimated from daily returns. Instead, we also consider low-frequency models based on range-based low-frequency volatility estimators to re-evaluate this question. Our study is related to Clements and Preve (2019), who shed light on what estimation techniques, volatility transformations, and specifications are most useful when one- to 22-day-ahead stock market volatility predictions are of interest. They also consider heterogeneous autoregressive (HAR) models where, instead of the realized variation, range-based estimators are employed. This is similar to our approach. We differ in five aspects: (i) we consider a broader class of models (apart from HAR models, we also consider long-memory and realized GARCH models); (ii) given possible model choice uncertainty, we also use simple combination forecasts; (iii) in order to be able to draw a stronger conclusion, we use 18 market indices instead of three; (iv) we also combine forecasts from high- and low-frequency volatility models in order to study whether low-frequency data add value if high-frequency data are available; and (v) we model the whole-day price variation (including the overnight price variation).

The existing literature has reported the increased forecasting accuracy of high-frequency volatility models, but primarily in a day-ahead volatility forecasting setting. In this paper, we apply a set of low- and high-frequency volatility models to a sample of 18 stock market indices around the world to test how out-of-sample volatility forecasting accuracy changes with respect to an increasing forecast horizon. We show that at a forecast horizon of up to 22 trading days, the out-of-sample accuracy of low- and high-frequency models tends to be statistically indistinguishable. Compared to inexpensive and easy-to-process low-frequency data, high-frequency data might have a limited advantage for multiple-day-ahead volatility forecasts.

The remainder of this paper is structured as follows. We describe the data and volatility estimators in Section 2. Section 3 explains the volatility models, out-of-sample forecasting procedure, and forecast evaluation. Section 4 presents the results, and Section 5 concludes.

2. Data and volatility estimators

We investigate the forecasting ability of high- and low-frequency volatility models using data from 18 developed and emerging stock market indices. Our sample period covers over 20 years of data: it starts in 2000 and ends in June 2020. We use two data sources. First, we use the OxfordMan Institute's Realized Library (Heber et al., 2009)[2] for data on high-frequency realized volatility estimators. To observe some empirical regularities, we aimed at studying as many market indices as possible, but at the same time, we were interested in long time series. In our data source, the given 18 market indices had a sufficiently long time series, starting in 2000 (except the Canadian GSPTSE) and ending in June 2020. For details, see the following Table 1. Second, we use Bloomberg terminal for low-frequency daily, opening O_t, highest H_t, lowest L_t, and closing C_t values (OHLC prices, henceforth), for a given day t, of the 18 market indices. In our analysis, data are treated as is usual in the volatility literature: we ignore weekends and holidays, and treat trading days as a continuous time series.[3]

[2] https://realized.oxford-man.ox.ac.uk/data.

[3] Using a similar sample of market indices, Lyócsa and Molnár (2017) show that explicitly modeling weekends in HAR models leads to systematic, but small forecast improvements: 1.42% on average across multiple market indices and for the QLIKE loss function. However, Lyócsa and Molnár (2017) find these improvements for day-ahead forecasts. For multiple-day-ahead scenarios, such improvements (if present) are likely even smaller. We therefore follow the most common approach, which is to simply consider trading days only and treat these trading days as one time series.

Table 1
Overview of market indices.

Index	Country	Country code	Starting date	# End date	# of forecasts
AEX	Netherlands	NL	02.02.2000	25.06.2020	5195
AORD	Australia	AU	04.02.2000	25.06.2020	5149
BSESN	India	IN	03.02.2000	25.06.2020	5035
FCHI	France	FR	02.02.2000	25.06.2020	5199
FTSE	United Kingdom	GB	03.02.2000	25.06.2020	5142
GDAXI	Germany	DE	02.02.2000	25.06.2020	5168
GSPTSE	Canada	CA	04.06.2002	24.06.2020	4515
HSI	Hong Kong	HK	02.02.2000	23.06.2020	4994
IBEX	Spain	ES	03.02.2000	25.06.2020	5164
IXIC	United States	US	03.02.2000	24.06.2020	5112
KS11	South Korea	KR	03.02.2000	25.06.2020	5017
KSE	Pakistan	PK	09.02.2000	25.06.2020	4955
MXX	Mexico	MX	02.02.2000	25.06.2020	5113
N225	Japan	JP	06.03.2000	23.06.2020	4958
NSEI	India	IN	03.02.2000	25.06.2020	5031
SPX	United States	US	03.02.2000	23.06.2020	5113
SSEC	China	CN	17.02.2000	23.06.2020	4921
SSMI	Switzerland	CH	03.02.2000	25.06.2020	5106

Note: Country codes refer to ISO 3166-1 α-2 abbreviations.

In the following subsections, we describe our procedures to estimate, filter, aggregate (low-frequency estimators), and adjust volatility estimators for overnight price variation.

2.1. Low-frequency volatility estimator

Let $c_t = ln(C_t) - ln(O_t)$, $h_t = ln(H_t) - ln(O_t)$, $l_t = ln(L_t) - ln(O_t)$, and $j_t = ln(O_t) - ln(C_{t-1})$. A simple estimator of intraday volatility is c_t^2. However, range-based estimators offer higher efficiency (Molnár, 2012). Therefore, if OHLC prices are available, one should resort to the following range-based estimators. Specifically, the (Parkinson, 1980) estimator is:

$$PK_t = \frac{(h_t - l_t)^2}{4ln2} \tag{1}$$

the (Garman & Klass, 1980) estimator is:

$$GK_t = 0.511\,(h_t - l_t)^2 - 0.019\,(c_t(h_t + l_t) - 2h_t l_t) - 0.383c_t^2 \tag{2}$$

and the (Rogers & Satchell, 1991) estimator is:

$$RS_t = h_t(h_t - c_t) + l_t(l_t - c_t) \tag{3}$$

2.2. High-frequency volatility estimator

The high-frequency volatility estimator in this study is based on the standard realized volatility:

$$RV_t^* = \sum_{i=1}^{N} r_{t,i}^2 \tag{4}$$

where $r_{t,i}$ is the ith intraday return on day t. Intraday returns are calculated using a calendar sampling scheme with a five-minute frequency. Our motivation for using this particular estimator is twofold. First, our sample consists of stock market indices that are constructed from data of individual stocks. At the individual stock level, microstructure noise effects are pronounced and may require the use of other estimators, e.g., the bipower estimator of Barndorff-Nielsen and Shephard (2004) or the median realized volatility estimator of Andersen et al. (2012). At the stock market index level, the microstructure noise effects are likely mitigated. Second, Liu et al. (2015) suggest that a five-minute sampling frequency is often a good choice in forecasting studies across different asset classes. We therefore opt for the simplest approach of using the standard realized volatility.

2.3. Overnight price variation

For most practical purposes (e.g., multiple-day-ahead asset allocation), one needs to account not only for the intraday price variation, as given by estimators above, but also for price changes during the overnight, non-trading time period. Prices for the non-trading period are usually not available.[4] Given the opening price, O_t, and the closing price, C_{t-1}, from the previous trading session, the overnight price variation is given by:

$$j_t^2 = (ln(O_t) - ln(C_{t-1}))^2 \tag{5}$$

The adjustment of price variation to include the overnight component follows only after data are filtered and low-frequency estimators are aggregated.

2.4. Filtering volatility estimates

The high- and low-frequency volatility estimators, as well as the overnight price variation, are occasionally subject to rare extreme observations. Such market conditions are of interest to market participants. Yet, such extreme market conditions (9/11 and the flash crash in 2010) tend be difficult if not impossible to predict using time-series models, while one or just a few such trading days have

[4] Some equity market indices are traded almost continuously; see Lyócsa and Todorova (2020).

the potential to influence the ranking of volatility models. Moreover, estimations of volatility models might also be influenced by such outliers in a detrimental way.

To make our analysis less dependent on such events, volatility estimators[5] were subject to the rolling window filtering procedure, where values above the 99.5 percentile were replaced by the 99.5 percentile value (i.e., winsorization). The size of the rolling window was set to 1000. This filtering procedure can be used in an out-of-sample exercise. However, as in all filtering procedures, it is somewhat arbitrary,[6] and the description of the effect it has on the resulting series is therefore warranted.

With respect to the low-frequency estimators, substitutions were made in 0.64% of cases on average[7] and in 0.92% of cases at most (GPTSE, Canada). At the same time, the first-order autocorrelation coefficient increased after the filtering by 20.84% on average and by 96.2% in one case (KSE, Pakistan), where the procedure picked up significant outliers that had a detrimental effect on the otherwise large autocorrelation structure of the time series. The effect on the realized volatility (overnight price variation) was very similar, with substitution in 0.69% (0.63%) of cases on average and an increase in the autocorrelation by 19.82% (11.35%) on average. Given our assumption that the most extreme market volatility is virtually unpredictable, we consider the filtering procedure to be successful, as it led to a substantial increase in the persistence of the series, while substituting a small portion of the data.

2.5. Combination of low-frequency volatility estimators

After filtering, the three low-frequency estimators are combined via a simple average:

$$RB_t^* = 3^{-1} (PK_t + GK_t + RS_t) \tag{6}$$

The averaging approach is motivated by Patton and Sheppard (2009), who argue that the true data-generating process is unknown and might even change over time; thus, the optimal estimator is also unknown. Therefore, a suitable strategy might be to combine different estimators and in doing so, to *diversify* against estimator choice uncertainty. Using the simple average realistically assumes that no prior information about the relative accuracy of the estimators exists.

2.6. Whole-day price variation adjustment

As we are interested in multiple-day-ahead forecasts (up to 22 days), it is necessary to account for overnight price variation, j_t^2 (defined in Eq. (5)). We follow the procedure of Hansen and Lunde (2005) that captures the total variation by calculating a weighted average of the intraday and overnight price variation components, where the optimal weights $\hat{\omega}_1$ and $\hat{\omega}_2$ account both for the covariance between the two components, as well as their unequal size.[8]

5 Including semivolatilities, defined below.

6 The choice of the quantile.

7 Across the three estimators and all 18 market indices.

8 The trading period is shorter than the overnight period. Consequently, the trading period variation tends to be smaller than the overnight price variation.

The obtained *adjusted price variation* used in the subsequent volatility models is denoted by RV_t (high-frequency estimator) and RB_t (low-frequency estimator).[9]

To calculate RV_t for $t \geq 1000$, we use the (Hansen & Lunde, 2005) procedure within a window of 1000 observations ending at t, (i.e., $t-999, t-998, \dots, t$) to obtain the optimal weights for the overnight price variation ($\hat{\omega}_1^t$) and intraday price variation ($\hat{\omega}_2^t$). The adjusted price variation for a given day t is then calculated as:

$$RV_t = \hat{\omega}_1^t j_t^2 + \hat{\omega}_2^t RV_t^* \tag{7}$$

For $t < 1000$, constant optimal weights obtained from the first window of the full 1000 observations are used. The same approach is used to calculate RB_t.

3. Methodology

3.1. Forecasting models

We use six forecasting models for both low- and high-frequency data. The models belong to the class of HAR models, autoregressive fractionally integrated moving average (ARFIMA) models, and GARCH models.

3.1.1. RV-HAR, RB-HAR models

The standard HAR model of Corsi (2009) is specified as:

$$RV_{t,H} = \beta_0 + \beta_1 RV_{t-1}^D + \beta_2 RV_{t-1}^W + \beta_3 RV_{t-1}^M + z_t, \tag{8}$$

where

$$RV_{t,H} = H^{-1} \sum_{k=1}^{H} RV_{t+k-1} \tag{9}$$

is the average multiple day-ahead price variation, and where we consider one- to 22-day-ahead volatility forecasts; thus, $H = 1, 2, \dots, 22$. The daily, weekly, and monthly volatility components are:

$$RV_{t-1}^D = RV_{t-1}\ ;\ RV_{t-1}^W = 5^{-1} \sum_{k=1}^{5} RV_{t-k}\ ;\ RV_{t-1}^M = 22^{-1} \sum_{k=1}^{22} RV_{t-k} \tag{10}$$

The model is popular due to its simplicity, yet if coupled with high-frequency data, the forecasting ability tends to be much better than that of standard GARCH models (e.g. Andersen et al., 2011; Horpestad et al., 2019. The high-frequency version of the model is denoted by RB-HAR. Using RB_t instead of RV_t leads to the RB-HAR model, which uses only low-frequency data.

3.1.2. RV-HAR-l, RB-HAR-L models

The RV-HAR-L model exploits the asymmetric volatility effect, which posits that volatility tends to be higher when the market falls. The adjusted HAR model includes

9 We denote the non-adjusted estimators of intraday price variation by RV_t^* and RB_t^*.

two terms that capture the sign of volatility effects (Corsi & Renò, 2012) and also the size effect (Horpestad et al., 2019). Specifically, let $R_t = ln(C_t) - ln(C_{t-1})$; the RV-HAR-L is specified as:

$$RV_{t,H} = \beta_0 + \beta_1 RV^D_{t-1} + \beta_2 RV^W_{t-1} + \beta_3 RV^M_{t-1} + \gamma_1 |R_{t-1}| + \gamma_2 |R_{t-1}| \times I(R_{t-1} < 0) + z_t \tag{11}$$

where $I(R_t < 0)$ is a signaling function that returns 1 if the condition is true, and 0 otherwise. The asymmetric effect is captured by the $\gamma_2 |R_{t-1}| \times I(R_{t-1} < 0)$ term. However, it is likely that when market volatility RV_{t-1} is higher, the absolute return $|R_{t-1}|$ also is, which in turn might lead to a significant γ_2, not only because of the existence of the asymmetric effect, but also because of the dependence between $|R_{t-1}|$ and RV_{t-1}. We therefore include the size effect $\gamma_1 \times |R_{t-1}|$.[10] Substituting RB_t leads to the RB-HAR-L model.

3.1.3. RV-HAR-SV model, RB-HAR-ASY

The RV-HAR-SV model described above aims to exploit the asymmetric volatility effect in different ways from the RV-HAR-L model. Patton and Sheppard (2015) show that the decomposition of the realized volatility into positive and negative return components might improve a model's forecasting accuracy. The positive and negative semivolatilities are:

$$PS_t = \sum_{i=1}^{N} r^2_{t,i} \times I(r_{t,i} > 0) \ ; \quad NS_t = \sum_{i=1}^{N} r^2_{t,i} \times I(r_{t,i} \leq 0) \tag{12}$$

where $I(r_{t,i} > 0)$ is a signaling function. The RV-HAR-SV model is specified as:

$$RV_{t,H} = \beta_0 + \beta_1 PS_{t-1} + \beta_2 NS_{t-1} + \beta_3 RV^W_{t-1} + \beta_4 RV^M_{t-1} + z_t \tag{13}$$

As shown by Patton and Sheppard (2015), the volatility asymmetric effect is exploited by higher persistence of NS_{t-1} as opposed to PS_{t-1}.

The same principle cannot be directly applied for a range-based HAR model. Instead, we use the following specification:

$$RB_{t,H} = \beta_0 + \beta_1 RB_{t-1} + \beta_2 RB_{t-1} \times I(R_{t-1} < 0) + \beta_3 RB^W_{t-1} + \beta_4 RV^M_{t-1} + z_t \tag{14}$$

If $\beta_2 > 0$, volatility tends to be higher after the market declines.

All HAR models are estimated using weighted least squares (WLS) with the goal of making the parameter estimates less sensitive to extreme volatility levels. In particular, we use the following weighting scheme: $w_t = RV^{-1}_{t,H}$.[11]

HAR model forecasts are performed using the direct method. That is, the given H-day-ahead realized (range-based) volatility estimate is directly modeled within a given estimation sample. Using the estimated coefficients and the last-known right-hand-side values, the next H-periods-ahead volatility is predicted.

3.1.4. ARFIMA-class models

As volatility exhibits long-memory properties, a natural candidate to model volatility is to use an ARFIMA-class model; see, e.g., Baillie (1996), Granger and Joyeux (1980) and Hosking (1981). We use the following specification:

$$RV_t = \mu_0 + z_t; \quad (1 - \phi L)(1 - L)^d z_t = (1 + \theta L)\epsilon_t \tag{15}$$

where L is the lag operator, and $0 < d < 1$ is the fractional integration parameter. The ARFIMA model is given by:

$$f\epsilon_t = \sigma_t \eta_t; \quad \eta_t \sim iid(0, 1) \tag{16}$$

where η_t follows the (Johnson, 1949a, 1949b) distribution, which is flexible to account for the asymmetries and heavy tails often observed with equity returns (Choi & Nam, 2008). The volatility of the error term, σ_t, changes over time. For modeling the volatility of realized volatility, we consider the standard GARCH model of Bollerslev (1986) and the resulting model is denoted by RV-ARFIMA-GARCH:

$$\sigma^2_t = \omega + \alpha\epsilon^2_{t-1} + \beta\sigma^2_{t-1} \tag{17}$$

Modeling RB_t in the mean equation leads to the RB-ARFIMA-GARCH model. Alternatively, we use the EGARCH model of Nelson (1991) in the volatility equation, in which case:

$$ln(\sigma^2_t) = \omega + \zeta RV_{t-1} + \alpha s_{t-1} + \gamma\left(|s_{t-1}| - E|s_{t-1}|\right) + \beta ln(\sigma^2_{t-1}) \tag{18}$$

where s_t represents standardized innovations. The resulting model is denoted by RV-ARFIMA-EGARCH, or, when range-based estimators are used, the model is denoted by RB-ARFIMA-EGARCH.

Following (Baillie et al., 2012), let

$$\pi(L) = 1 - \sum_{j=1}^{\infty} \pi_j L^j \equiv (1 - \phi L)(1 - L)^d (1 + \theta L)^{-1} \tag{19}$$

$$\psi(L) = 1 + \sum_{j=1}^{\infty} \psi_j L^j \equiv (1 + \theta L)(1 - \phi L)^{-1}(1 - L)^{-d} \tag{20}$$

A stationary long-memory ARFIMA process with $|d| < 0.5$ may be alternatively expressed as an infinite AR or MA process as

$$\pi(L)z_t = \epsilon_t, \qquad z_t = \psi(L)\epsilon_t \tag{21}$$

10 An absolute value of return can be considered as an alternative volatility measure, as returns high in absolute value are likely associated with high volatility. In other words, significant γ_2 also captures the impact of past volatility. The asymmetric effect is captured by the $\gamma_2|R_{t-1}| \times I(R_{t-1} < 0)$ term, as this term captures how the impact of $|R_{t-1}|$ differs between positive and negative returns.

11 Clements and Preve (2019) discuss several alternative estimation techniques: OLS, LAD, and four WLS techniques of the form $w_t = x_t^{-1}$. Here, x_t is one of the following: (i) conditional volatilities of OLS residuals, (ii) OLS fitted values (as in Patton and Sheppard (2015)), (iii) the square root of realized quarticity that down-weights observations when the measurement error is large, or (iv) simply RV_t, which is our method of choice. In fact, given several market indices and multiple forecast horizons, Clements & Preve, 2019 find that $w_t = RV^{-1}_{t,H}$ leads to the most accurate forecasts. We therefore follow their approach.

A predictor for a single step ($H = 1$) at time t may be obtained as

$$z_{t,H} = \sum_{j=0}^{\infty} \pi_{1+j} y_{t-j} \tag{22}$$

Following Eq. (15), the volatility forecast is calculated as $RV_{t,H} = \mu_0 + z_{t,H}$. For other forecast horizons, rolling forecasts based on Eq. (22) are calculated (see Section 3.3).

3.1.5. RV-garch, RB-GARCH models

(Hansen et al., 2012) propose a framework that relates the conditional volatility to the given realized (or other) measures of price variation. Given the ARMA mean equation for returns,

$$r_{i,t} = \mu_0 + z_t; \quad (1 - \phi L)z_t = (1 + \theta L)\epsilon_t \tag{23}$$

where L is the lag operator. The RV-GARCH model is defined by the volatility equation:

$$\sigma_t^2 = \omega + \beta\sigma_{t-1}^2 + \gamma RV_{t-1} \tag{24}$$

and the measurement equation:

$$RV_t = \xi + \varphi\sigma_t^2 + \tau(\epsilon_t) + u_t; \; u_t \sim N(0, \lambda), \tag{25}$$

where $\tau(\epsilon_t) \equiv \tau_1\epsilon_t + \tau_2(\epsilon_t^2 - 1)$ accounts for possible asymmetric shocks.

For forecasting, we follow (Hansen et al., 2012), and note that by combining the last two equations, it is possible to obtain a VARMA(1,1) structure:

$$\begin{bmatrix} \sigma_t^2 \\ RV_t \end{bmatrix} = \begin{bmatrix} \beta & \gamma \\ \varphi\beta & \varphi\gamma \end{bmatrix} \begin{bmatrix} \sigma_{t-1}^2 \\ RV_{t-1} \end{bmatrix} + \begin{bmatrix} \omega \\ \xi + \varphi\omega \end{bmatrix} + \begin{bmatrix} 0 \\ \tau(\epsilon_t) + u_t \end{bmatrix} \tag{26}$$

which (e.g., for $H = 1$) may be used for forecasting

$$\begin{bmatrix} \sigma_{t+H}^2 \\ RV_{t+H} \end{bmatrix} = \begin{bmatrix} \beta & \gamma \\ \varphi\beta & \varphi\gamma \end{bmatrix}^H \begin{bmatrix} \sigma_t^2 \\ RV_t \end{bmatrix} + \sum_{j=0}^{H-1} \begin{bmatrix} \beta & \gamma \\ \varphi\beta & \varphi\gamma \end{bmatrix}^j \left\{ \begin{bmatrix} \omega \\ \xi + \varphi\omega \end{bmatrix} + \begin{bmatrix} 0 \\ \tau(\epsilon_{t+H-j}) + u_{t+H-j} \end{bmatrix} \right\} \tag{27}$$

Substituting RB_t for RV_t leads to the RB-GARCH model. As in ARFIMA-class models, for other forecast horizons, rolling forecasts are calculated (see Section 3.3).[12]

3.2. Forecast combinations

The idea behind employing forecast combinations is to mitigate model choice uncertainty (Timmermann, 2006). In our setting, one could argue that the combination of low-frequency volatility forecasts could lead to forecasts that are more accurate than those given by individual, high-frequency volatility forecasting models. Therefore, ignoring forecast combinations could result in underestimating the accuracy of low-frequency volatility models. However, similar arguments hold for high-frequency volatility models. We therefore employ combination forecasts for both low- and high-frequency volatility model forecasts. Furthermore, the recent study of Ma et al. (2018) shows that, in a day-ahead volatility forecasting setting, combining low- and high-frequency forecasts leads to improved forecasting accuracy. We therefore study whether this also holds for multiple market indices and, more importantly, for longer forecast horizons. In order to do so, we also combine forecasts from individual high- and low-frequency volatility models.

We use three combination forecasts. Let $FV_{t,H}^{(m)}$ denote the value of the forecasted volatility from individual model $m = 1, 2, \ldots, M$, with $M = 6$ (six low- and six high-frequency volatility models), where (m) indicates that the forecasted values are ordered from the lowest to highest forecasts. First, the *trimmed mean* forecast of size $p = 1$ is given by:

$$FV_{t,H}^{trim} = \frac{1}{(M-2)} \sum_{m=2}^{M-1} FV_{t,H}^{(m)} \tag{28}$$

which is a simple average after removing the highest and the lowest prediction.

Next, we use a *weighted average* across forecasting models, where each model receives a higher weight if it resulted in lower forecast errors in the past (calibration sample):

$$FV_{t,H}^{\delta} = \sum_{m=1}^{M} FV_{t,H}^{(m)} w_{t,H}^{*(m)} \tag{29}$$

$$w_{t,H}^{*(m)} = \frac{(w_{t,H}^m)^{-1}}{\sum_{m=1}^{M} (w_{t,H}^m)^{-1}}, \quad w_{t,H}^m = \frac{1}{CS} \sum_{j=t-CS}^{t-1} \delta^{j-t+2} L_{j,H}^m \tag{30}$$

where δ is the discount factor (Ma et al., 2018; Stock & Watson, 2004), $L_{j,H}^m$ is the value of the loss function, and CS is the size of the calibration sample, set to $CS = 200$. We use $\delta = 0.95$, which gives higher weight to recent forecast errors in the calibration sample.[13]

Finally, given the weights in Eq. (30), we use the forecasts from the models (lowest weighted forecast errors) that performed the best over the past CS observations. This is the *recent best* approach. Our main results are reported under the QLIKE loss function (defined below).

3.3. Forecasting procedure and predictions

We employ a rolling window forecasting algorithm. The size of the estimation window is $E = 1000$ observations, which is similar to the value found in other volatility forecasting studies, e.g., Patton and Sheppard (2015). The rolling drift parameter is set to 1, meaning that all the model parameters are updated daily using the last 1000 known observations. Note that the first evaluated forecast starts with the 1201st observation, as 200 observations are used in the calibration sample to calculate the combination forecasts.

[12] GARCH-class models are estimated using the (Ghalanos, 2020) rugarch package in R.

[13] Note that the weights sum to 1 and the forecast error realized 150 observations before t received a weight of just 0.001 when $\delta = 0.95$. However, if the calibration sample contains a period of larger forecast errors (crisis period), such forecast errors still influence the model rankings.

3.4. Forecast evaluations

To evaluate the forecasts,[14] we use the asymmetric QLIKE loss function, which provides a consistent model ranking even in the presence of a noisy proxy (Patton, 2011) of the following form (Christoffersen, 2011):

$$L_{t,H}(m) = RV_{t,H}/FV_{t,H} - ln(RV_{t,H}/FV_{t,H}) - 1 \tag{31}$$

Our main results are based on the QLIKE loss function, but the mean square forecast error (MSFE) is a viable alternative as well (Patton, 2011). We offer a short discussion of our findings under the MSFE loss function in a separate subsection.

An important aspect of our study is that the *volatility proxy* is the *realized volatility*, thus putting the low-frequency volatility forecasts at a disadvantage and strengthening our argument that low- and high-frequency volatility model forecasts are similar for longer forecast horizons.

To summarize, we compare the average loss function for each of the 18 stock market indices, for one- to 22-day-ahead forecast horizons and across 21 models: six individual models and three combinations of *either* only low- or only high-frequency volatility models, and three combinations of all individual low- and high-frequency volatility models $(2 \times (6+3) + 3)$. The statistical evaluation of the forecasting accuracy is performed via the model confidence set of Hansen et al. (2011), which sequentially evaluates all forecasts until only models with equal forecasting accuracy remain. We use the T_{max} statistic from (Hansen et al., 2011), where the critical values are bootstrapped (stationary bootstrap with random block lengths) using 2000 samples and $\alpha = 0.15$. We apply the test on a set of 21 models, for each market index and for each forecasting horizon separately.

4. Results

4.1. Estimates of annualized realized volatility across stock markets

The statistical characteristics of the annualized realized volatilities in Table 2 show that volatilities are subject to well-known stylized facts, e.g., excess volatility and persistence. For example, the distribution of volatilities is skewed to the right and shows a high level of kurtosis, which is consistent with the occurrence of days of excess volatilities. Next, volatilities show considerable persistence that declines slowly with time. Our estimates are in line with these observations, as the average first-order autocorrelation (across both estimators and market indices) is 0.67 (see Table 2), while at the 5th, 22nd, and 100th lag, the persistence is still at 0.53, 0.31, and 0.14, respectively, suggesting that volatility has long memory.[15]

A comparison of volatility levels across stock markets shows that the highest levels of volatility are found for emerging stock markets in India and China, and for the technology index NASDAQ. Lower levels are found for developed markets in Australia, Canada, and Switzerland. Among the emerging markets, lower volatility is observed for Pakistan and Mexico. The differences in the volatility levels can be considerable. Broader stock market indices and those with higher market capitalization are likely to show lower levels of volatility than market indices in emerging markets.

4.2. Realized and range-based volatility: An empirical comparison

As range-based estimators are unbiased, we would expect small differences in the averages (across time) between realized volatility and range-based estimators. This indeed appears to be the case, as realized volatilities are on average 1.34% higher than range-based estimators.[16] This is not much given that for five market indices, the range-based estimator is in fact higher than the realized volatility estimator. At the same time, given that realized volatility is more efficient, range-based estimators should be noisier; thus, empirically, we would expect range-based estimators to have larger volatility. The results in Table 2 (column SD) confirm this intuition. Sudden spikes of range-based estimators are also visible in Fig. 1. Across all 18 market indices, the range-based estimators have a higher standard deviation. Moreover, they lead to much higher skewness and kurtosis.

The volatility series tend to be persistent, a feature exploited by autoregressive models. Given that realized volatilities are less noisy estimates, we would expect them to show larger persistence than range-based estimators. The results in Table 2 largely confirm this intuition, as the persistence of the range-based estimators is always lower, with the relative difference increasing for higher lag orders. Specifically, the persistence of realized volatility is 24.6%, 30.2%, 38.4%, and 55.6% higher than the persistence of range-based estimators.

Given these observations, we would therefore expect that the realized volatility models would perform better for market indices where the differences in the magnitude of the volatility estimates and persistence are larger. Having established baseline volatility characteristics, we now turn our attention to the empirical part, an out-of-sample study. In the next section, we compare forecasts generated by models using either realized volatility (RV) or range-based volatility (RB) estimates, i.e., RV-models and RB-models.

[14] Given 18 market indices, 21 models, and 22 forecast horizons, the forecasts occasionally lead to implausibly large or small forecasts. Such issues are not uncommon in the volatility forecasting literature. We follow (Bollerslev et al., 2016) and use an 'insanity filter' that makes our analysis more realistic. All negative volatility forecasts are replaced by the minimum of the target variable (either RV_t or RB_t) found over the estimation window. Forecasts that are higher than the 99.9 percentile of the target variable over the estimation window are replaced by that given quantile.

[15] Persistence at higher than the first order is not reported in Table 2.

[16] Calculated from the mean volatility in Table 2.

Table 2
Summary statistics of realized volatility, range-based volatility, and daily returns.

Index	High-frequency data: realized volatility							Low-frequency data: range-based volatility							Daily returns						
	Mean	SD	Skew.	Kurt.	min.	max.	$\rho(1)$	Mean	SD	Skew.	Kurt.	min.	max.	$\rho(1)$	Mean	SD	Skew.	Kurt.	min.	max.	$\rho(1)$
AEX	1120.0	1794.1	4.9	35.7	21.5	22147.6	0.80	1092.2	2152.9	6.4	61.6	15.1	33171.9	0.70	−0.002	1.413	−0.224	10.481	−11.376	10.028	−0.006
AORD	542.4	971.9	7.0	75.1	10.2	16754.2	0.70	557.8	1154.7	8.1	104.6	0.0	25570.4	0.60	0.012	1.029	−0.738	11.719	−10.203	6.766	−0.064
BSESN	1288.5	2031.3	5.6	51.4	6.5	33330.9	0.70	1322.8	2758.6	8.2	109.7	16.9	57103.9	0.50	0.037	1.480	−0.373	12.371	−14.102	16.115	0.052
FCHI	1180.1	1688.8	4.9	38.7	26.0	23365.2	0.80	1168.3	2029.7	5.9	53.4	12.6	30073.8	0.70	−0.003	1.454	−0.229	9.390	−13.098	10.595	−0.029
FTSE	955.1	1605.1	5.8	51.5	21.0	24007.9	0.70	898.5	1739.5	6.5	59.9	16.0	26329.7	0.60	0.000	1.203	−0.364	11.100	−11.512	9.384	−0.039
GDAXI	1303.4	1996.2	4.7	35.4	32.3	27692.2	0.80	1323.0	2433.1	6.5	73.4	8.2	52379.2	0.70	0.011	1.495	−0.184	8.891	−13.055	10.797	−0.016
GSPTSE	624.6	1363.7	7.1	71.8	10.9	21798.6	0.80	624.1	1621.4	9.0	117.7	2.5	32766.3	0.50	0.015	1.097	−1.147	20.820	−13.176	9.370	−0.050
HSI	924.9	1287.1	6.5	70.2	47.9	21600.1	0.70	974.6	1917.4	10.9	208.6	25.5	56068.0	0.50	0.009	1.462	−0.090	10.830	−13.582	13.407	−0.010
IBEX	1314.6	1655.8	4.8	40.4	41.1	21385.6	0.80	1283.0	1986.2	5.8	56.3	27.7	34787.5	0.70	−0.008	1.488	−0.310	11.003	−15.151	13.484	−0.010
IXIC	1387.1	2236.2	4.4	29.9	27.2	23626.8	0.80	1360.3	2835.3	8.2	146.1	12.1	80639.9	0.60	0.017	1.600	−0.140	9.862	−13.149	13.255	−0.068
KS11	1168.1	1771.7	4.9	39.0	61.7	24655.7	0.80	1181.6	2305.9	8.3	127.8	15.7	59057.3	0.70	0.016	1.496	−0.546	10.084	−12.805	11.284	0.016
KSE	1095.6	1655.6	4.2	25.3	0.2	15942.4	0.60	1078.2	1849.8	5.5	64.8	0.1	40270.8	0.50	0.062	1.247	−0.340	6.896	−8.049	8.069	0.115
MXX	1238.7	1974.0	5.8	53.6	43.6	32628.5	0.60	1192.7	2147.9	7.2	81.8	18.9	42827.4	0.60	0.034	1.279	−0.057	8.493	−8.267	10.441	0.088
N225	1000.6	1421.1	5.9	53.9	8.6	20836.1	0.70	1006.0	2016.3	12.3	274.0	12.9	65281.8	0.50	0.002	1.506	−0.365	9.277	−12.111	13.235	−0.028
NSEI	1523.9	2545.1	6.2	64.0	3.2	44236.5	0.70	1438.8	3138.3	8.4	113.4	9.8	61129.5	0.50	0.037	1.472	−0.477	13.244	−13.904	16.226	0.053
SPX	923.8	1739.6	6.1	53.3	9.0	23754.6	0.80	862.4	1844.5	7.6	85.6	5.4	35421.1	0.70	0.016	1.260	−0.386	14.093	−12.765	10.957	−0.115
SSEC	1568.6	2231.0	3.9	25.1	44.7	27177.1	0.70	1511.5	2429.7	4.6	33.4	15.3	31755.8	0.60	0.011	1.550	−0.399	7.884	−9.256	9.401	0.021
SSMI	829.8	1453.4	6.6	68.3	55.1	24477.2	0.80	803.3	1748.7	9.1	146.2	14.6	47454.7	0.60	0.007	1.177	−0.362	11.336	−10.134	10.788	0.021

Notes: Daily returns are multiplied by 100[%], while realized volatility and the range-based estimator are annualized. SD denotes standard deviation, Skew. the skewness, and Kurt. the kurtosis; and $\rho(1)$ is the first-order autocorrelation coefficient. According to the serial-correlation test of Escanciano and Lobato (2009), all realized and range-based estimators have significant persistence. With respect to close-to-close returns, significant persistence is found for indices AORD, IXIC, KSE, MXX, and SPX.

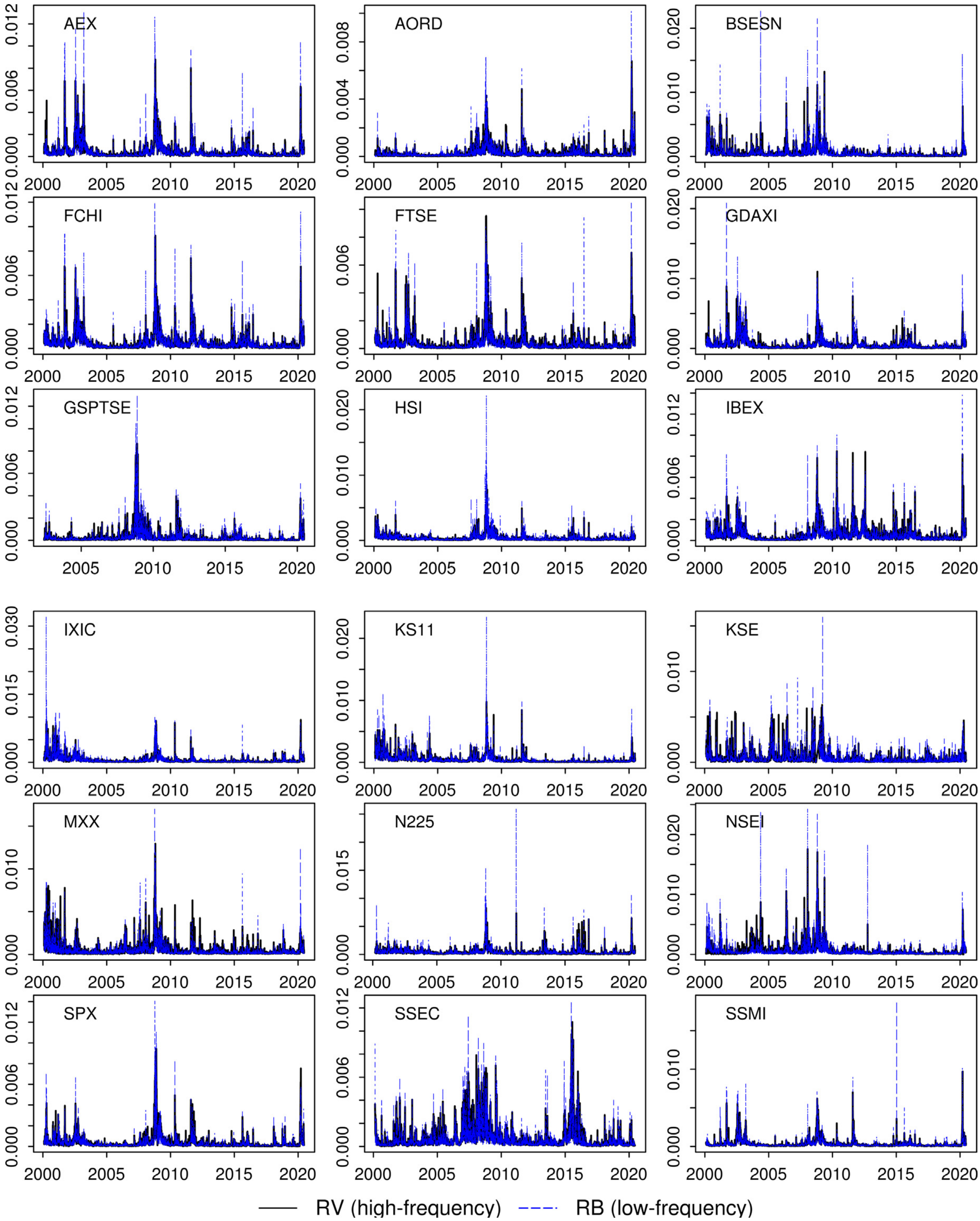

Fig. 1. Realized and range-based volatility.
Note: The lines correspond to raw (not annualized) series of realized and range-based volatility estimates.

4.3. Out-of-sample forecasts

The results from the out-of-sample study are presented in Tables 3–5. Table 3 presents the results for one-day-ahead forecasts, Table 4 for five-day-ahead forecasts, and Table 5 for 22-day-ahead forecasts. The values in the tables are the losses (QLIKE) averaged across time: lower values suggest more accurate forecasts. The columns correspond to the stock market indices. The † symbol indicates that the given forecast belongs to the set of superior forecasting models (Hansen et al., 2011), where all models in the given column are statistically compared. Each table is divided into five panels. The first two correspond to individual forecasting models. Panel A shows the results from models that utilize high-frequency data, i.e., realized volatility, and Panel B, the low-frequency data, i.e., range-based volatility. The third, fourth, and fifth panels present the results from the combination forecasts. Panel C corresponds to three combination forecasts derived from individual high-frequency models, Panel D uses combinations from individual low-frequency models, and Panel E presents results from forecasts that combine results from all individual (high- and low-frequency models) forecasting models. A visual comparison of the accuracy of the selected RV- and RB-models is presented in Fig. 2.

Among all the individual forecasting models, the best performing model is the RV-ARFIMA-EGARCH model (see the results in Tables 3–5). It consistently leads to low forecast errors across stock market indices and forecast horizons. Interestingly, RV-ARFIMA-GARCH does not perform particularly well for longer forecast horizons, meaning that allowing for asymmetry in the volatility of volatility improves forecast accuracy. Among low-frequency models, RV-ARFIMA-EGARCH again performed the best. Irrespective of whether we employed realized or range-based measures, the realized GARCH models performed poorly.

Combination forecasts, specifically the *weighted average* approach, almost never surpass the best performing individual model (RV-ARFIMA-EGARCH). However, in an out-of-sample setting, where one is unsure about the best individual forecasting model, combination forecasts are a good option. Specifically, the combinations from realized volatility models (*weighted average*) almost always belong to the set of superior models. This holds across market indices and forecast horizons. Moreover, even if the RV-ARFIMA-EGARCH model does not perform well for some market indices and forecast horizons, the weighted average of individual high-frequency model forecasts leads to competitive forecasts. Our results therefore present further empirical evidence that in real-life scenarios, combination forecasts tend to be very useful; these results are in accordance with (Lyócsa et al., 2017).

Our key observation is that the benefits of using high-frequency data diminish with an increasing forecasting horizon. For day-ahead forecasts, the set of superior models almost always includes only models with realized volatility. This result is surprisingly consistent across stock markets, with individual exceptions,[17] found only for market indices in Mexico, Japan, and India (NSEI). For five-day-ahead volatility forecasts, at least one range-based low-frequency volatility forecast model belongs to the set of superior models in eight of 18 stock market indices. For the 22-day-ahead forecast, this value improves in 17 of the 18 stock market indices (Netherlands's market index being the sole exception). That is, we almost always find at least one low-frequency volatility model that performs competitively with the high-frequency volatility models.

That said, it is also evident that the superior set of models mostly includes realized volatility models, which suggests that an analysis is subject to less model choice uncertainty when high-frequency data models are used. However, our results show that for multiple-day-ahead volatility forecasts, the combination forecasts created from range-based forecasting models are a safe bet for analysts. Specifically, the *weighted average* with discount factor $\delta = 0.95$ performs very competitively, and the *trimmed mean* much less so, while the *recent best* shows very inconsistent results across market indices and forecast horizons; it is therefore not recommended.

We have established that for day-ahead volatility forecasts, realized volatility models are superior to range-based models: the question is, to what extent. For each stock market index, we calculated the average loss across realized volatility models, and compared the result with the average loss across range-based models. Across all stock market indices, the range-based models lead to forecast errors that are 78.6% higher than the forecast errors from realized volatility models. Comparing the realized and range-based versions of weighted average combinations forecasts shows that the range-based model leads to 26.2% higher forecast errors on average. We can therefore conclude that high-frequency data are needed for short forecasting horizons.

We perform the same analysis for five- and 22-day-ahead volatility forecasts. A comparison of the average forecast errors shows that individual realized volatility models have approximately 15.9% lower forecast errors. This result illustrates a clear drop in the 'value' of high-frequency data at longer forecasting horizons. Finally, for the 22-day-ahead forecasts, individual realized volatility models have approximately 10.1% (still in favor of the realized volatility models) lower forecast errors on average.[18] These results let us conclude that for volatility forecasting purposes, the importance of high-frequency data diminishes with increasing forecasting horizons. Indeed, as noted above, for 22-day-ahead volatility forecasts, the differences between most of the realized and range-based volatility models are not statistically significant (see the model confidence set results).

4.4. Is there an added value to low-frequency volatility forecasts?

We have established that high-frequency models are useful for short-term forecasts. In a day-ahead setting, Ma

[17] Notably the RB-ARFIMA-EGARCH model.

[18] When calculating the 10.1%, the average loss for the low-frequency model excluded the $RB-ARFIMA-GARCH$ models for BSENS, GPTSE, HSE, MXX, and N225, as the average losses were clear outliers.

Table 3
Day-ahead forecast evaluation of high- and low-frequency models and combination forecasts.

	AEX	AORD	SEN	CAC	FTSE	DAX	TSE	HSI	IBEX	NASQ	KS11	KSE	MXX	N225	NSEI	SPX	SSEC	SSM
Panel A: Individual high-frequency volatility model forecasts																		
RV-HAR	0.17	0.29	0.19	0.18	0.25	0.19	0.41	0.17	0.16	0.18	0.14	0.38	0.32	0.34	0.28†	0.27	0.22	0.12
RV-HAR-SV	0.17	0.28	0.18	0.17	0.25	0.19	0.41	0.17	0.16	0.18	0.14	0.37	0.32	0.34	0.34†	0.27	0.22	0.12
RV-HAR-L	0.17	0.28	0.17	0.18	0.23	0.19	0.39	0.17	0.16	0.18	0.14	0.36	0.32	0.33	0.64†	0.26	0.22	0.12
RV-ARFIMA-GARCH	0.16	0.23	0.15†	0.16†	0.22	0.17	0.30†	0.16	0.14†	0.19	0.14	0.27†	0.24†	0.27	0.17†	0.23	0.19†	0.12
RV-ARFIMA-EGARCH	0.15†	0.22†	0.15†	0.15†	0.20†	0.16†	0.30†	0.14†	0.14†	0.15†	0.13†	0.25†	0.22†	0.24†	0.17†	0.21†	0.19†	0.11†
RV-GARCH	0.74	0.65	0.76	0.67	0.96	0.61	0.76	0.39	0.77	0.66	0.49	1.10	1.34	0.43	1.22	0.84	0.72	0.69
Panel B: Individual low-frequency volatility model forecasts																		
RB-HAR	0.39	0.54	0.45	0.36	0.47	0.41	2.00	0.41	0.34	0.56	0.30	2.43	0.57	0.49	0.87	0.59	0.40	0.35
RB-HAR-ASY	0.39	0.53	0.44	0.35	0.47	0.39	1.88	0.41	0.33	0.53	0.29	2.43	0.57	0.48	0.86	0.57	0.40	0.34
RB-HAR-L	0.39	0.58	0.51	0.36	0.49	0.46	1.88	0.42	0.34	0.56	0.29	14.17	0.59	0.49	1.16	0.65	0.42	0.36
RB-ARFIMA-GARCH	0.24	0.34	0.21	0.20	0.27	0.21	0.50	0.26	0.17	0.27	0.17	0.40	0.29	0.28	0.35	0.30	0.22	0.22
RB-ARFIMA-EGARCH	0.18	0.24	0.18	0.17	0.23	0.18	0.45	0.17	0.17	0.21	0.15	0.32	0.24†	0.26†	0.26†	0.26	0.23	0.14
RB-GARCH	0.69	0.59	0.69	0.65	0.89	0.54	0.87	0.35	0.78	0.62	0.49	1.11	1.41	0.40	1.09	0.81	0.75	0.71
Panel C: Combination of high-frequency volatility model forecasts																		
Trimmed mean	0.16	0.26	0.16	0.16	0.23	0.18	0.35	0.16	0.15	0.17	0.13	0.31	0.28	0.28	0.19†	0.24	0.20	0.12
Recent best, $\delta = 0.95$	0.15†	0.22†	0.15†	0.15†	0.21	0.16†	0.30†	0.14†	0.14†	0.16	0.13†	0.29†	0.23†	0.24†	0.17†	0.21†	0.19†	0.11†
Weighted average, $\delta = 0.95$	0.15†	0.23	0.15†	0.15†	0.21†	0.16†	0.30†	0.15	0.14†	0.16†	0.13†	0.27†	0.24†	0.25†	0.17†	0.22†	0.19†	0.11†
Panel D: Combination of low-frequency volatility model forecasts																		
Trimmed mean	0.31	0.39	0.30	0.28	0.37	0.28	0.69	0.28	0.26	0.38	0.22	0.75	0.44	0.34	0.52	0.42	0.30	0.27
Recent best, $\delta = 0.95$	0.17	0.24	0.16	0.17	0.23	0.18	0.43	0.18	0.16	0.21	0.14	0.32	0.24†	0.25†	0.28†	0.26	0.22	0.14
Weighted average, $\delta = 0.95$	0.21	0.27	0.19	0.19	0.27	0.20	0.34	0.17	0.19	0.23	0.15	0.30	0.29	0.26	0.26†	0.28	0.22	0.17
Panel E: Combination of high- and low-frequency volatility model forecasts																		
Trimmed mean	0.43	0.50	0.44	0.39	0.52	0.40	0.93	0.35	0.38	0.50	0.30	1.07	0.66	0.46	0.77	0.57	0.43	0.37
Recent best, $\delta = 0.95$	0.15†	0.22†	0.15†	0.15†	0.21	0.16†	0.30†	0.15	0.14†	0.16	0.13†	0.28†	0.23†	0.24†	0.17†	0.22†	0.19†	0.11†
Weighted average, $\delta = 0.95$	0.16	0.24	0.15	0.16	0.22	0.17	0.30†	0.15	0.15	0.17	0.13†	0.26†	0.25	0.24†	0.18†	0.23	0.19†	0.12

Notes: Values in the table are averages of the QLIKE loss function, where each forecast from the model (in the row) is compared to the proxy, the five-minute realized volatility. The symbol † means that the given model (in the row) was part of the superior set of models, as indicated by the model confidence test of Hansen et al. (2011), starting from the set of all individual and combination forecast models.

Table 4
Five-day-ahead forecast evaluation of high- and low-frequency models and combination forecasts.

	AEX	AORD	SEN	CAC	FTSE	DAX	TSE	HSI	IBEX	NASQ	KS11	KSE	MXX	N225	NSEI	SPX	SSEC	SSM
Panel A: Individual high-frequency volatility model forecasts																		
RV-HAR	0.20	0.22	0.19†	0.19	0.22	0.20†	0.32	0.13†	0.18†	0.26	0.16†	0.35	0.23	0.31	0.21	0.37†	0.22	0.16†
RV-HAR-SV	0.20	0.22	0.19†	0.19	0.22	0.20†	0.32	0.13†	0.18†	0.26	0.16†	0.35	0.23	0.31	0.21	0.37†	0.23	0.16†
RV-HAR-L	0.20	0.22	0.19†	0.19	0.22	0.20	0.31	0.13†	0.18†	0.26	0.16†	0.43†	0.23	0.30	0.21	0.38†	0.23	0.16†
RV-ARFIMA-GARCH	0.22	0.24	0.20†	0.20	0.26	0.21	0.33	0.15	0.19†	0.35	0.22†	0.35†	0.27	0.38	0.22	0.43†	0.24	0.19
RV-ARFIMA-EGARCH	0.18†	0.19†	0.17†	0.17†	0.20†	0.18†	0.28†	0.12†	0.20†	0.22†	0.15†	0.24†	0.19†	0.26†	0.19†	0.31†	0.20†	0.16†
RV-GARCH	0.98	0.77	0.97	0.86	1.16	0.78	0.93	0.42	0.96	0.98	0.62	1.76	1.54	0.57	1.50	1.25	0.92	0.90
Panel B: Individual low-frequency volatility model forecasts																		
RB-HAR	0.28	0.27	0.26	0.25	0.30	0.23	0.46	0.16	0.25	0.37	0.20	0.61	0.30	0.34	0.41	0.47	0.27	0.24
RB-HAR-ASY	0.28	0.26	0.25	0.25	0.30	0.23	0.44	0.16	0.25	0.36	0.20	0.60	0.29	0.34	0.40	0.46	0.27	0.24
RB-HAR-L	0.28	0.26	0.26	0.25	0.30	0.23	0.45	0.16	0.25	0.36	0.20	0.65	0.30	0.34	0.42	0.47	0.27	0.24
RB-ARFIMA-GARCH	0.36	0.33	0.31	0.27	0.32	0.24	0.73	0.28	0.25	0.42	0.20	0.43	0.38	0.38	0.37	0.55	0.26	0.31
RB-ARFIMA-EGARCH	0.22	0.21†	0.20†	0.20	0.25	0.19†	0.41	0.15†	0.22	0.26	0.16†	0.29†	0.21	0.27†	0.29	0.35†	0.25	0.20
RB-GARCH	0.90	0.68	0.86	0.84	1.09	0.68	1.05	0.37	0.94	0.84	0.63	1.41	1.61	0.52	1.29	1.14	0.92	0.88
Panel C: Combination of high-frequency volatility model forecasts																		
Trimmed mean	0.19	0.21	0.19†	0.19	0.22	0.19†	0.31†	0.13†	0.18†	0.26	0.16†	0.33†	0.22	0.30	0.21	0.36†	0.22	0.16†
Recent best, $\delta = 0.95$	0.18†	0.20†	0.17†	0.18†	0.21†	0.19†	0.29†	0.12†	0.18†	0.23†	0.15†	0.37†	0.18†	0.28†	0.18†	0.37†	0.20†	0.16†
Weighted average, $\delta = 0.95$	0.19†	0.20†	0.18†	0.18†	0.21†	0.19†	0.29†	0.12†	0.18†	0.24	0.15†	0.29†	0.20	0.28†	0.19	0.34†	0.21	0.16†
Panel D: Combination of low-frequency volatility model forecasts																		
Trimmed mean	0.27	0.26	0.25	0.24	0.29	0.22	0.41	0.15	0.24	0.34	0.19	0.49	0.29	0.31	0.38	0.45	0.26	0.23
Recent best, $\delta = 0.95$	0.22	0.22	0.20†	0.21	0.25	0.20†	0.38	0.14†	0.20	0.27	0.17†	0.30†	0.20	0.28†	0.26	0.41†	0.23	0.18
Weighted average, $\delta = 0.95$	0.24	0.23	0.21	0.21	0.26	0.20	0.33	0.13†	0.22	0.29	0.17†	0.33†	0.25	0.29†	0.29	0.39†	0.24	0.19
Panel E: Combination of high- and low-frequency volatility model forecasts																		
Trimmed mean	0.38	0.35	0.35	0.33	0.42	0.30	0.54	0.20	0.33	0.45	0.25	0.69	0.45	0.39	0.52	0.58	0.35	0.32
Recent best, $\delta = 0.95$	0.19	0.20†	0.18†	0.18	0.21†	0.19†	0.30†	0.13†	0.19†	0.24	0.16†	0.37†	0.18†	0.27†	0.19†	0.38†	0.20†	0.16†
Weighted average, $\delta = 0.95$	0.20	0.21†	0.19†	0.19	0.22	0.19†	0.30†	0.12†	0.19†	0.26	0.16†	0.29†	0.21	0.27†	0.21	0.35†	0.22	0.16†

Notes: Values in the table are averages of the QLIKE loss function, where each forecast from the model (in the row) is compared to the proxy, the five-minute realized volatility. The symbol † means that the given model (in the row) was part of the superior set of models, as indicated by the model confidence test of Hansen et al. (2011), starting from the set of all individual and combination forecast models.

et al. (2018) shows that low-frequency volatility models might still be useful if their forecasts are combined with high-frequency volatility models. In Panel E of Tables 3–5, we perform such an analysis but move it into multiple-day-ahead horizons as well.

For longer forecast horizons, combinations from high- and low-frequency volatility forecasts tend to be in the set of superior models, while average losses are not lower than those achieved by high-frequency volatility forecasts. In the short term, one should resort only to high-frequency volatility forecasts, as adding low-frequency volatility forecasts to combinations might make the resulting forecasts even worse.[19] It follows, that there is very little benefit to using low-frequency forecasts if high-frequency forecasts are available.

Finally, we complement our analysis with Fig. 2, which shows how the forecast errors change with respect to the forecast horizon. For comparison purposes, we selected the *weighted average* combination forecast that discounts past forecast errors with a discount factor $\delta = 0.95$. The model leads to competitive forecasts for most market indices and forecast horizons for both high- (black line) and low-frequency (blue line) models. A dot on the line indicates that the given model belongs to the superior set of models—i.e., that the model leads to an accuracy that is statistically similar to that of the other best performing models in the set.

The differences in forecast accuracy tend to be large for short forecast horizons, and then decline with increasing

[19] Our results are therefore slightly different from those of Ma et al. (2018).

Table 5
Twenty-two-day-ahead forecast evaluation of high- and low-frequency models and combination forecasts.

	AEX	AORD	SEN	CAC	FTSE	DAX	TSE	HSI	IBEX	NASQ	KS11	KSE	MXX	N225	NSEI	SPX	SSEC	SSM
Panel A: Individual high-frequency volatility model forecasts																		
RV-HAR	0.42†	0.37†	0.34†	0.37†	0.40†	0.37†	0.56†	0.15†	0.31†	0.47†	0.23†	0.31†	0.30†	0.45†	0.35†	0.77†	0.29†	0.41†
RV-HAR-SV	0.42†	0.37†	0.34†	0.37†	0.40†	0.37†	0.56†	0.15†	0.31†	0.47†	0.23†	0.31†	0.30†	0.45†	0.35†	0.77†	0.29†	0.41†
RV-HAR-L	0.42†	0.38†	0.34†	0.37†	0.40†	0.37†	0.56†	0.15†	0.31†	0.48†	0.23†	0.31†	0.30†	0.45†	0.35†	0.77†	0.29†	0.41†
RV-ARFIMA-GARCH	0.42†	0.47	0.34†	0.36†	0.68†	0.39†	0.56†	0.22	0.31†	0.48†	0.30†	0.41	0.58	1.34	0.38†	0.71†	2.14†	0.57†
RV-ARFIMA-EGARCH	0.41†	0.34†	0.32†	0.36†	0.39†	0.34†	0.53†	0.16†	0.38†	0.41†	0.23†	0.28†	0.26†	0.39†	0.32†	0.75†	0.27†	0.40†
RV-GARCH	1.63	1.28	1.42	1.36	1.79	1.24	1.57	0.51	1.32	1.51	0.81	1.52	1.93	0.89	1.94	2.15	1.17	1.52
Panel B: Individual low-frequency volatility model forecasts																		
RB-HAR	0.51	0.41†	0.36	0.42†	0.50†	0.38†	0.59†	0.15†	0.37†	0.54	0.27†	0.34†	0.36	0.46†	0.47	0.82†	0.31†	0.47†
RB-HAR-ASY	0.51	0.41†	0.36	0.42†	0.50†	0.38†	0.60†	0.15†	0.36†	0.54	0.27†	0.33†	0.36	0.46†	0.47	0.83†	0.31†	0.48†
RB-HAR-L	0.51	0.41†	0.36	0.43†	0.50†	0.38†	0.60†	0.15†	0.37†	0.54	0.27†	0.33†	0.36	0.46†	0.47	0.83†	0.32†	0.48†
RB-ARFIMA-GARCH	0.86	1.58	2.72	0.82	1.21	0.50†	1.80†	6.58	0.55†	0.79	0.27	0.57	5.46	14.45	0.81	4.03	0.44†	1.07
RB-ARFIMA-EGARCH	0.50	0.36†	0.32†	0.38†	0.49†	0.37†	0.55†	0.18	0.39†	0.44†	0.25†	0.30†	0.31†	0.41†	0.40†	0.72†	0.32†	0.46†
RB-GARCH	1.51	1.11	1.21	1.29	1.69	1.09	1.62	0.44	1.25	1.30	0.80	1.31	2.00	0.84	1.69	1.98	1.18	1.41
Panel C: Combination of high-frequency volatility model forecasts																		
Trimmed mean	0.42†	0.38†	0.34†	0.37†	0.40†	0.37†	0.56†	0.16†	0.32†	0.47†	0.23†	0.33†	0.31	0.45†	0.35†	0.74†	0.30†	0.41†
Recent best, $\delta = 0.95$	0.41†	0.39†	0.32†	0.37†	0.40†	0.39†	0.59†	0.19†	0.31†	0.45†	0.25†	0.32†	0.28†	0.50†	0.34†	0.81†	0.32†	0.53†
Weighted average, $\delta = 0.95$	0.41†	0.37†	0.33†	0.36†	0.39†	0.36†	0.55†	0.15†	0.31†	0.45†	0.23†	0.31†	0.28†	0.45†	0.33†	0.71†	0.29†	0.40†
Panel D: Combination of low-frequency volatility model forecasts																		
Trimmed mean	0.52	0.42†	0.37	0.42†	0.50†	0.39†	0.60†	0.16†	0.38†	0.54	0.27†	0.36	0.38	0.46†	0.47	0.82†	0.32†	0.48†
Recent best, $\delta = 0.95$	0.49	0.42†	0.33†	0.48†	0.46†	0.40†	0.54†	0.17†	0.35†	0.50†	0.28	0.30†	0.32	0.83†	0.39†	1.80†	0.36†	0.63†
Weighted average, $\delta = 0.95$	0.49	0.40†	0.34†	0.41†	0.47†	0.38†	0.55†	0.15†	0.36†	0.50	0.26†	0.32†	0.34	0.45†	0.43	0.77†	0.31†	0.45†
Panel E: Combination of high- and low-frequency volatility model forecasts																		
Trimmed mean	0.67	0.56	0.49	0.55	0.68	0.51	0.77†	0.22	0.51	0.68	0.33	0.52	0.60	0.58	0.64	1.03†	0.45†	0.61
Recent best, $\delta = 0.95$	0.42†	0.40†	0.32†	0.44†	0.41†	0.38†	0.54†	0.19†	0.34†	0.45†	0.25†	0.30†	0.29†	0.49†	0.34†	0.82†	0.32†	0.48†
Weighted average, $\delta = 0.95$	0.44†	0.38†	0.33†	0.38†	0.42†	0.36†	0.54†	0.15†	0.32†	0.47†	0.24†	0.31†	0.30†	0.44†	0.36†	0.73†	0.29†	0.42†

Notes: Values in the table are averages of the QLIKE loss function, where each forecast from the model (in the row) is compared to the proxy, the five-minute realized volatility. The symbol † means that the given model (in the row) was part of the superior set of models, as indicated by the model confidence test of Hansen et al. (2011), starting from the set of all individual and combination forecast models.

horizons. That is, high-frequency data are most important for short forecast horizons. With increasing forecast horizons, the differences tend to diminish, and in many instances, the forecasts converge. Usually, the red line, namely, the combination from all individual forecasts, lies in between the high- and low-frequency volatility forecasts, suggesting that combining forecasts from both approaches does not lead to increased forecast accuracy.[20]

4.5. Results under squared forecast errors

In many practical instances, using the QLIKE is advantageous, as it assigns more weight to volatility underestimation. Alternatively, the mean squared forecast error (MSFE),

$$\left(RV_{t,H} - FV_{t,H}\right)^2, \tag{32}$$

is a symmetric loss function that penalizes extreme under/overestimations more severely. In the presence of a noisy proxy, it also leads to consistent model rankings (Patton, 2011).

In this section, we present results for the MSFE. In order to do that, we only need to re-estimate combination forecasts, where past forecast errors are now weighted according to the MSFE instead of the QLIKE. The resulting forecasts are evaluated using the MSFE. Fig. 3 presents our main results, where we compare the weighted combination forecasts (similar to Fig. 2).[21]

Again, the results show that with increasing forecast horizons, low- and high-frequency volatility models lead to similar forecast errors. Even though for short-term forecasts, high-frequency volatility models tend to lead to lower forecast errors, for most market indices, these differences are statistically insignificant. Therefore, our general conclusion—that a low-frequency volatility model forecast still matters—appears to be even stronger than under the MSE loss function.

A closer inspection reveals that almost all indices suffer from periods of large forecast errors associated with periods of high market volatility. During such periods, no model performs well, and the MSFE from these periods is highly influential when the forecasting accuracy is evaluated across the whole sample. Consequently, it is difficult to statistically distinguish between forecast errors. That is, high- and low-frequency volatility forecasts provide statistically comparable accuracy.

4.6. Asset allocation study

In this subsection, we present a portfolio study as an empirical application, where we compare the Sharpe ratios of two asset allocation strategies. In the original mean–variance portfolio theory, the utility function of a risk-averse investor is shown to be an increasing function of expected return, and a decreasing function of return volatility, which is interpreted as a measure of risk. In the first strategy, portfolio weights are managed using a high-frequency volatility forecast model. In the second strategy, portfolio weights are managed using a low-frequency volatility forecast model.

4.6.1. Portfolio construction

As our objective is to compare the consequences of using different volatility forecasts, we opt for a rather simple portfolio model, where the allocation is restricted to only two assets: the stock market index and a risk-free asset, similar to Luo et al. (2019), Lyócsa and Todorova (2020) and Wang et al. (2016). Our approach follows a U.S. investor with risk-free rates approximated via the three-month US T-bill and perfectly hedged foreign currency positions. This investor can always hold only two assets: risk-free asset and a particular stock market index.

[20] Our conclusions here are drawn only with respect to our simple combination methods. More sophisticated methods might lead to different conclusions.

[21] Detailed tabulated results for $h = 1, 5, 22$ can be found in the Electronic Supplementary Material.

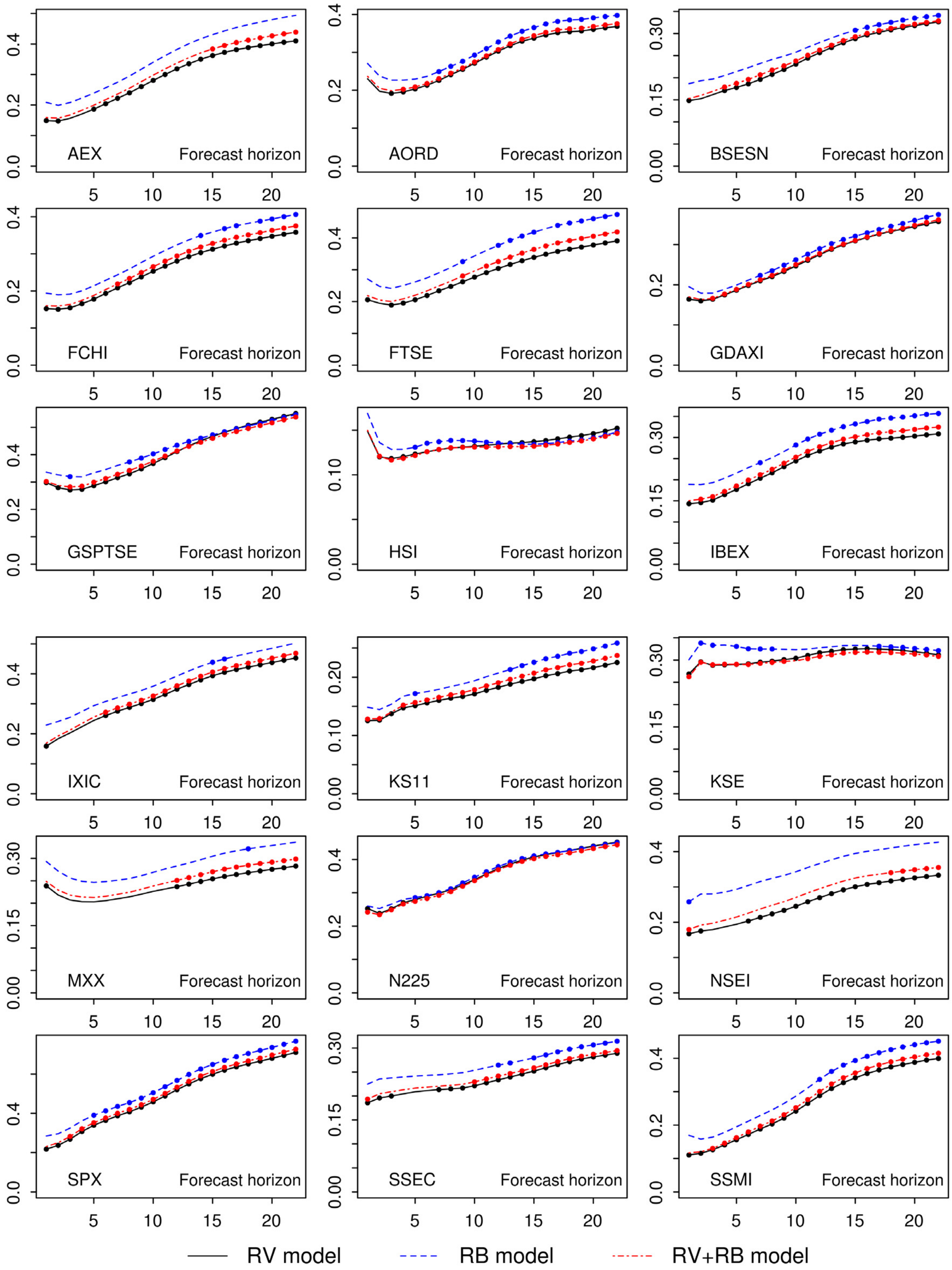

Fig. 2. Forecast comparison of QLIKE and MCS results across forecasting horizons.
Notes: The *y*-axis is the average loss achieved by a weighted average combination forecast. The black line corresponds to the *weighted average* combination forecast given by high-frequency volatility models. The blue line corresponds to the *weighted average* combination forecast given by low-frequency volatility models. The red line corresponds to the *weighted average* combination forecast given by both high- and low-frequency volatility models. A dot on the line means that the given forecast was in the superior set of models. If dots are present for a given horizon in two or more lines, the two forecasts are statistically indistinguishable. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

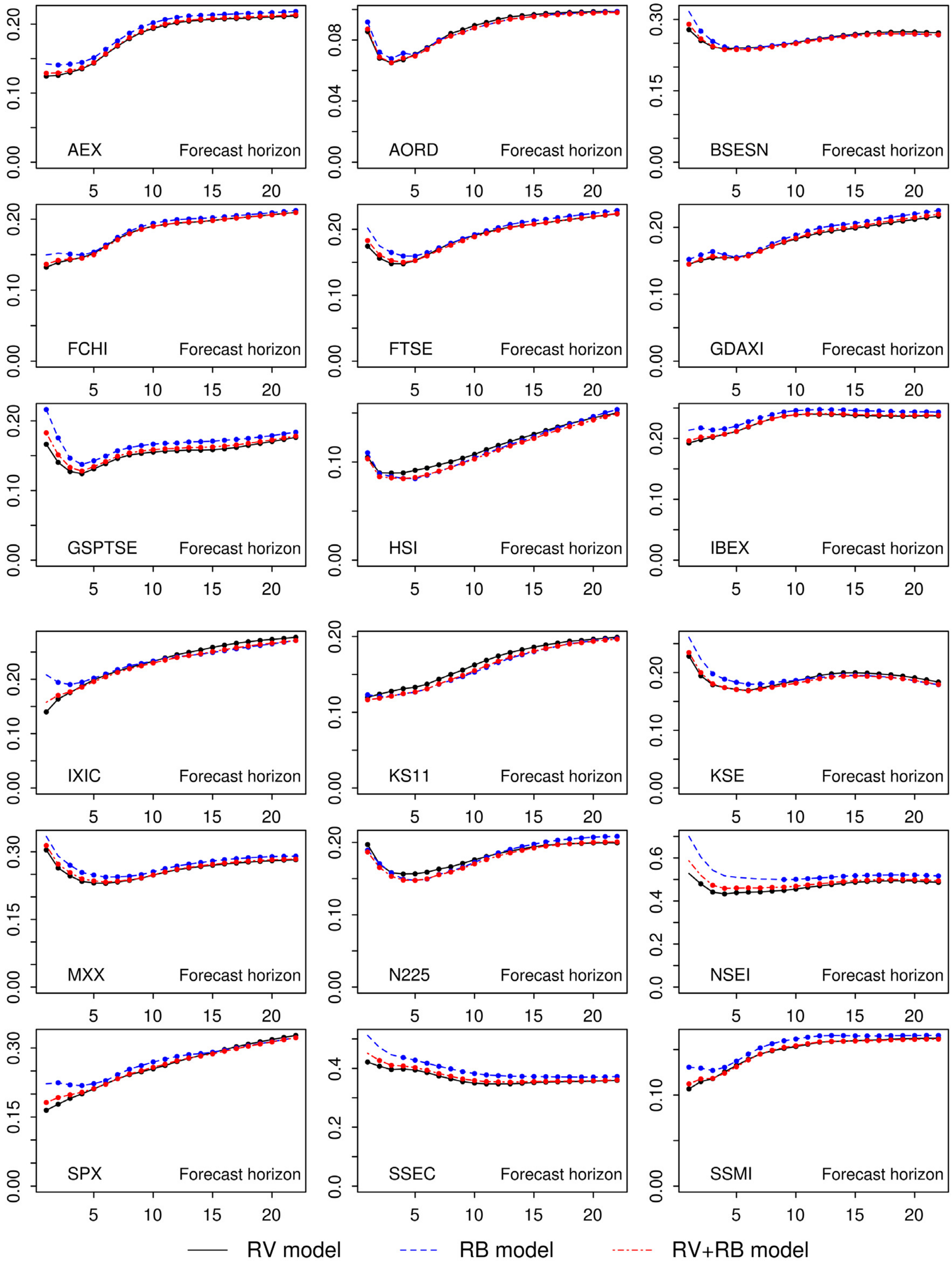

Fig. 3. Forecast comparison of MSE and MCS results across forecasting horizons.
Notes: The *y*-axis is the average loss achieved by a weighted average combination forecast. The black line corresponds to the *weighted average* combination forecast given by high-frequency volatility models. The blue line corresponds to the *weighted average* combination forecast given by low-frequency volatility models. The red line corresponds to the *weighted average* combination forecast given by both high- and low-frequency volatility models. A dot on the line means that the given forecast was in the superior set of models. If dots are present for a given horizon in two or more lines, the two forecasts are statistically indistinguishable. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The portfolio choice problem reduces to the calculation of weights $w_{i,t}$ maximizing the utility function

$$E\left[w_{i,t}(R_{i,H,t}-RF_{H,t})+RF_{H,t}\right] -\frac{\gamma}{2}D\left[w_{i,t}(R_{i,H,t}-RF_{H,t})+RF_{H,t}\right] \quad (33)$$

Here, $i = 1, 2, \ldots, 18$ refers to the stock market index, $R_{i,H,t}$ is its H-period index return at time t, and $RF_{H,t}$ is the return for holding the T-bill between $t - H + 1$ and t. We allow for borrowing, but not short sales of the stock index, by imposing $0 \leq w_{i,t} \leq 1.5$. With regard to the risk-aversion parameter, we allow for $\gamma = 3$, which represents low risk aversion, and for $\gamma = 6$, which is high risk aversion.[22] The $E[.]$ and $D[.]$ are the usual expectation and dispersion operators.

To calculate the optimal weights, we take (conditional) expectations on the ex ante value $R^*_{i,H,t}$ and volatility $FV_{i,H,t}$ of the index returns. Following its performance reported in Welch and Goyal (2007), we consider the simple h-period average return within the estimation window of $W = 1000$ days as the estimate $R^*_{i,H,t}$. For the volatility, we directly use the forecasted $FV_{i,H,t}$.[23] As we treat the three-month T-bill as a risk-free asset, its expected return ($RF^*_{H,t}$) is calculated as the yield at $t-H+1$, scaled for the period of H days, matching the forecast horizon. Accordingly, the covariance of the stock market index and the risk-free asset is assumed to be zero over the investment horizon. Under these assumptions, the stock market index optimal portfolio weight $w^*_{i,t}$ may be directly calculated as

$$w^*_{i,t} = \frac{R^*_{i,H,t}-RF^*_{H,t}}{\gamma FV_{i,H,t}} \quad (34)$$

The weight for the risk-free asset is then simply $1 - w^*_{i,t}$. Weights are updated after each trading day using the past 1000 observations.

4.6.2. Portfolio performance evaluation

Let T^o designate the set of time indices, excluding the subsample necessary for the calculation of volatility forecasts and expected returns defined in the previous subsection (i.e., the estimation window). The number of out-of-sample periods is given by the cardinality $|T^o|$. Additionally, let $R_{i,H,t}$ and $RF_{H,t}$ represent the H-period out-of-sample returns and yields at $t \in T^o$. We define the portfolio return and excess portfolio return at $t \in T^o$ as

$$r_{i,H,t} = w^*_{i,t}R_{i,H,t} + (1-w^*_{i,t})RF_{H,t} \quad (35)$$

$$r^e_{i,H,t} = w^*_{i,t}R_{i,H,t} + (1-w^*_{i,t})RF_{H,t} - RF_{H,t} \quad (36)$$

The average out-of-sample return and excess return are given by

$$\mu_{i,h} = \frac{1}{|T^o|}\sum_{t\in T^o} r_{i,H,t}, \qquad \mu^e_{i,h} = \frac{1}{|T^o|}\sum_{t\in T^o} r^e_{i,H,t} \quad (37)$$

Regarding the portfolio return volatility ($\sigma^2_{i,H}$), we opt for the long-run volatility estimator based on the approach of Andrews (1991), with the quadratic spectral weighing scheme and automatic bandwidth parameter selection by Newey and West (1994), in order to account for the possible autocorrelation introduced by using the moving windows in portfolio construction.

We compare the performance of the portfolios by Sharpe ratios, defined as

$$SR_{i,H} = \frac{\mu^e_{i,H}}{\sigma_{i,H}} \quad (38)$$

To formally evaluate the performance measured by Sharpe ratios for the selected strategies, we test for the null hypothesis of their equality using the robust statistical test of Ledoit and Wolf (2008) based on the studentized circular bootstrap of Politis and Romano (1994) and heteroscedasticity and autocorrelation consistent standard errors; for details, see Section 3.1 in Ledoit and Wolf (2008).

4.6.3. Evaluation of the asset allocation study

We compare asset allocation along two strategies. In the first, $FV^{HF}_{i,H,t}$ is predicted using the weighted average combination forecasts from high-frequency individual model forecasts. In the second, $FV^{LF}_{i,H,t}$ is predicted using the weighted average combination forecasts from low-frequency individual model forecasts. Annualized[24] Sharpe ratios across market indices and forecast horizons are reported in Tables 6–7.

Note that, in many instances, Sharpe ratios are negative. However, the purpose of this study is not to find the most profitable strategy, but instead to compare the economic merit of the two strategies. There are two key observations from the portfolio study. First, regardless of whether the analyst uses high- or low-frequency volatility models, Sharpe ratios are very similar across market indices and forecast horizons. As indicated by the (Ledoit & Wolf, 2008) test, statistically significant differences are almost nonexistent. Second, larger differences tend to systematically appear for short-term forecast horizons, but they are almost never significant. These results suggest that in an asset allocation framework, low-frequency volatility model forecasts are as useful as high-frequency volatility model forecasts.

4.7. Implications for future studies

In our study, the proxy for the unobserved integrated variance is the five-minute realized volatility adjusted via the (Hansen & Lunde, 2005) procedure. This also includes microstructure noise, most notably variations due to intraday price discontinuities. One might consider modeling jump robust alternatives (e.g., the bipower variation of Barndorff-Nielsen & Shephard, 2004, or the median realized volatility of Andersen et al., 2012). However, a straightforward comparison of jump robust volatility estimators with range-based daily estimators is problematic,

[22] Results for $\gamma = 1, 9$ lead to qualitatively similar results and are available upon request.

[23] Note that we forecast the price variation directly. We therefore do not use the '2' superscript for $FV_{i,H,t}$.

[24] To annualized Sharpe ratios, we follow the non-i.i.d. approach of Lo (2002).

Table 6
Sharpe ratios for $\gamma = 3$ and weighted average ($\delta = 0.95$) combination forecasts.

		AEX	AORD	SEN	CAC	FTSE	DAX	TSE	HSI	IBEX	NASQ	KS11	KSE	MXX	N225	NSEI	SPX	SSEC	SSM
$h = 1$																			
High-frequency	[Sharpe ratio]	−0.202	−0.012	**0.368**	−0.315	−0.352	0.008	0.080	**0.126**	0.005	0.617	0.157	0.287	0.254	0.067	0.320	0.479	−0.096	**−0.085**
Low-frequency	[Sharpe ratio]	−0.150	−0.054	**0.311**	−0.273	−0.316	0.011	−0.022	**0.075**	−0.042	0.649	0.136	0.340	0.184	0.054	0.273	0.522	−0.089	**−0.015**
LW test	[p-value]	0.690	0.340	**0.000**	0.320	0.440	0.270	0.090	**0.040**	0.510	0.300	0.990	0.980	0.390	0.530	0.100	0.790	1.000	**0.030**
$h = 5$																			
High-frequency	[Sharpe ratio]	−0.021	−0.019	0.064	−0.061	−0.088	−0.003	−0.227	**−0.045**	−0.038	0.106	0.038	0.046	0.177	0.033	0.055	0.089	−0.018	−0.002
Low-frequency	[Sharpe ratio]	−0.023	−0.002	0.066	−0.067	−0.119	−0.004	−0.058	**−0.018**	−0.038	0.288	0.035	0.040	0.019	0.033	0.060	0.076	0.000	−0.002
LW test	[p-value]	0.240	0.770	0.390	0.380	0.220	0.750	0.330	**0.030**	0.060	0.950	0.280	0.480	0.260	0.930	0.980	0.570	0.320	0.850
$h = 22$																			
High-frequency	[Sharpe ratio]	−0.006	−0.026	0.025	−0.014	−0.026	−0.002	−0.016	−0.003	−0.023	0.008	0.015	0.001	0.026	0.002	0.023	0.006	−0.004	−0.004
Low-frequency	[Sharpe ratio]	−0.006	−0.003	0.025	−0.014	−0.026	0.000	−0.016	−0.022	−0.023	0.094	0.005	0.013	0.052	0.003	0.024	0.017	−0.001	−0.003
LW test	[p-value]	0.850	0.080	0.770	0.950	0.620	0.720	0.890	0.150	0.930	0.870	0.750	0.860	0.080	0.550	0.200	0.680	0.790	0.830

Notes: Values correspond to annualized Sharpe ratios (non-IID annualization as in Lo, 2002). Bolded values are statistically significant Sharpe ratios as indicated by the Ledoit and Wolf (2008) circular bootstrap test (LW test) with bandwidth equal to the forecast horizon.

as the former exclude intraday jumps, while in the latter case the theoretical relationship between intraday price discontinuities and daily range-based estimators is as yet unknown. In our study, we were interested in predictions of the overall price variation over multiple days, using daily range-based estimators. We were therefore constrained to compare a daily range-based estimator with a high-frequency estimator that also covers the whole price variation.

We opted for an equal number of individual high- and low-frequency volatility models. However, the availability of intraday price variations leads to a richer set of high-frequency HAR class volatility models than low-frequency HAR class models. In our study, we used the benchmark HAR model, the semivolatility HAR model, and the asymmetric volatility (leverage) HAR model. Popular alternatives include the signed jump model of Patton and Sheppard (2015), the continuous and jump component model of Andersen et al. (2007), and the measurement error model of Bollerslev et al. (2016). Another possibility is to use more advanced estimation techniques, (see Clements & Preve, 2019) that allow for non-constant HAR model coefficients (e.g., Luo et al., 2019; Wang et al., 2016). The existing literature does not provide clear guidance into what models tend to systematically outperform the rest.[25] We therefore opted for simpler models and combination forecasts that appear to perform adequately across different asset classes (e.g., Lyócsa & Molnár, 2018; Wang et al., 2016).

The models used in this study only rely on price information from the given market index. Several studies argue that the accuracy of forecasting models can be improved when data from other markets or sources are explored. For example, Degiannakis and Filis (2017) exploit variables from stock, foreign exchange, commodity markets, and the macro-environment, Lyócsa and Molnár (2018) exploit information from related assets, Bollerslev et al. (2018) from common volatility components (Lyócsa & Todorova, 2020), and from intraday and overnight price variations in market indices. However, by using relevant exogenous variables, we would not be able to compare the usefulness of high- and low-frequency volatility estimators directly, which is the main purpose of this study. We expect that relevant exogenous variables are likely to decrease the gap between high- and low-frequency models. Our analysis might therefore be considered as a first step that establishes that all is not lost with low-frequency volatility models.

5. Conclusion

The concept of realized volatility was introduced after the emergence of high-frequency data, leading to more precise volatility estimates. With improved computation speeds and memory management, volatility models employing realized volatility measures have gained in popularity. Existing studies have shown that high-frequency volatility models outperform low-frequency models (Andersen et al., 2003; Horpestad et al., 2019; Koopman et al., 2005; Wei, 2012). However, in the existing literature, low-frequency volatility models are models estimated from daily returns, such as GARCH models. Since the highest and lowest prices of the day contain additional information about volatility, we considered low-frequency volatility models based on open, high, low, and close daily prices. Using a sample of 18 stock market indices, we compared these low-frequency volatility models to high-frequency volatility models. The results revealed that high-frequency volatility models tend to outperform low-frequency volatility models in a one- to five-day-ahead out-of-sample forecasting setting. However, at longer forecasting horizons, the differences in forecasting accuracy tend to diminish and become statistically indistinguishable.

The intuition behind these results is that volatility has a long memory and changes only gradually from day to day. Therefore, a precise estimate of the current day's volatility is very useful in predicting the volatility of the following day. However, when we are interested in volatility levels weeks ahead, a precise estimate of the current day's volatility is not that important. Since for longer forecast horizons, the benefits of working with high-frequency data might be limited, and because high-frequency data are seldom freely available, in some applications, low-frequency data might be sufficient.

We complemented our main analysis with an asset allocation study. We showed that an asset allocation strategy that uses high-frequency volatility model forecasts does not outperform (in terms of Sharpe ratios) an asset allocation strategy that uses low-frequency volatility model forecasts. This result held across market indices

[25] In a recent study on commodity market volatility, Degiannakis et al. (2020) show that it is difficult to select specific model specifications that would outperform other specifications. Similar results for the nonferrous futures markets or oil and natural gas are found in Lyócsa and Molnár (2018), Lyócsa et al. (2017).

Table 7
Sharpe ratios for $\gamma = 6$ and weighted average ($\delta = 0.95$) combination forecasts.

		AEX	AORD	SEN	CAC	FTSE	DAX	TSE	HSI	IBEX	NASQ	KS11	KSE	MXX	N225	NSEI	SPX	SSEC	SSM
$h = 1$																			
High-frequency	[Sharpe ratio]	−0.224	0.001	**0.355**	−0.388	−0.325	0.047	0.136	0.134	**0.028**	0.502	0.168	0.309	0.256	−0.017	0.330	0.394	−0.092	−0.122
Low-frequency	[Sharpe ratio]	−0.196	−0.026	**0.306**	−0.335	−0.292	0.088	0.022	0.089	**0.009**	0.572	0.152	0.351	0.218	−0.048	0.263	0.428	−0.095	−0.013
LW test	[p-value]	0.384	0.093	**0.005**	0.504	0.897	0.896	0.127	0.056	**0.016**	0.602	0.568	0.350	0.213	0.817	0.199	0.612	0.720	0.125
$h = 5$																			
High-frequency	[Sharpe ratio]	−0.049	−0.012	0.064	−0.079	−0.100	−0.009	−0.051	−0.004	−0.019	0.125	0.101	0.056	0.053	0.026	0.057	0.103	−0.009	−0.020
Low-frequency	[Sharpe ratio]	−0.096	−0.031	0.051	−0.081	−0.235	−0.008	−0.052	−0.015	−0.025	0.123	0.039	0.092	0.050	0.025	0.056	0.099	−0.004	−0.017
LW test	[p-value]	0.533	0.179	0.978	0.269	0.546	0.688	0.672	0.799	0.709	0.516	0.200	0.174	0.582	0.884	0.370	0.156	0.104	0.983
$h = 22$																			
High-frequency	[Sharpe ratio]	−0.007	−0.003	0.024	−0.016	−0.026	−0.003	−0.017	−0.004	−0.023	0.031	0.002	0.010	0.024	0.001	0.022	0.023	−0.001	−0.005
Low-frequency	[Sharpe ratio]	−0.060	−0.002	0.000	−0.125	−0.026	−0.003	−0.017	−0.002	−0.023	0.027	0.003	0.010	0.025	0.005	0.023	0.021	−0.001	−0.005
LW test	[p-value]	0.993	0.243	0.425	0.428	0.829	0.974	0.900	0.104	0.470	0.681	0.245	0.164	0.934	0.424	0.530	0.377	0.871	0.226

Notes: Values correspond to annualized Sharpe ratios (non-IID annualization as in Lo, 2002). Bolded values are statistically significant Sharpe ratios as indicated by the Ledoit & Wolf, 2008 circular bootstrap test (LW test) with bandwidth equal to the forecast horizon.

and forecast horizons. These results imply that in order to assess the merits of a new high-frequency volatility model, it should be tested in a multiple-day-ahead setting, benchmarked against low-frequency volatility models, and validated in an economic application.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Lyócsa and Molnár appreciate the funding support of the GACR 18-05829S.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.ijforecast.2020.12.001.

References

Aalborg, H. A., Molnár, P., & de Vries, J. E. (2019). What can explain the price, volatility and trading volume of Bitcoin? *Finance Research Letters*, *29*, 255–265.
Alizadeh, S., Brandt, M. W., & Diebold, F. X. (2002). Range-based estimation of stochastic volatility models. *The Journal of Finance*, *57*(3), 1047–1091.
Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*, *89*(4), 701–720.
Andersen, T. G., Bollerslev, T., Diebold, F. X., & Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economic*, *61*(1), 43–76.
Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2001). The distribution of realized exchange rate volatility. *Journal of the American Statistical Association*, *96*(453), 42–55.
Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, *71*(2), 579–625.
Andersen, T. G., Bollerslev, T., & Huang, X. (2011). A reduced form framework for modeling volatility of speculative prices based on realized variation measures. *Journal of Econometrics*, *160*(1), 176–189.
Andersen, T. G., Dobrev, D., & Schaumburg, E. (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics*, *169*(1), 75–93.
Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 817–858.
Bailey, G., & Steeley, J. M. (2019). Forecasting the volatility of the Australian dollar using high-frequency data: Does estimator accuracy improve forecast evaluation? *International Journal of Finance & Economics*, *24*(3), 1355–1389.
Baillie, R. T. (1996). Long memory processes and fractional integration in econometrics. *Journal of Econometrics*, *73*(1), 5–59.
Baillie, R. T., Kongcharoen, C., & Kapetanios, G. (2012). Prediction from ARFIMA models: Comparisons between MLE and semiparametric estimation procedures. *International Journal of Forecasting*, *28*(1), 46–53.
Barndorff-Nielsen, O. E., & Shephard, N. (2002). Estimating quadratic variation using realized variance. *Journal of Applied Econometrics*, *17*(5), 457–477.
Barndorff-Nielsen, O. E., & Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics*, *2*(1), 1–37.
Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, *31*(3), 307–327.
Bollerslev, T., Hood, B., Huss, J., & Pedersen, L. H. (2018). Risk everywhere: Modeling and managing volatility. *Review of Financial Studies*, *31*(7), 2729–2773.
Bollerslev, T., Patton, A. J., & Quaedvlieg, R. (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics*, *192*(1), 1–18.
Brandt, M. W., & Jones, C. S. (2006). Volatility forecasting with range-based EGARCH models. *Journal of Business & Economic Statistics*, *24*(4), 470–486.
Choi, P., & Nam, K. (2008). Asymmetric and leptokurtic distribution for heteroscedastic asset returns: the SU-normal distribution. *Journal of Empirical Finance*, *15*(1), 41–63.
Chortareas, G., Jiang, Y., & Nankervis, J. C. (2011). Forecasting exchange rate volatility using high-frequency data: Is the euro different? *International Journal of Forecasting*, *27*(4), 1089–1107.
Chou, R. Y. (2005). Forecasting financial volatilities with extreme values: the conditional autoregressive range (CARR) model. *Journal of Money, Credit and Banking*, *37*(3), 561–582.
Chou, R. Y., & Liu, N. (2010). The economic value of volatility timing using a range-based volatility model. *Journal of Economic Dynamics and Control*, *34*(11), 2288–2301.
Christoffersen, P. (2011). *Elements of financial risk management*. Academic Press.
Clements, A., & Preve, D. (2019). A practical guide to harnessing the HAR volatility model. Available at SSRN 3369484.
Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Economic*, *7*(2), 174–196.
Corsi, F., & Renò, R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics*, *30*(3), 368–380.
Degiannakis, S., & Filis, G. (2017). Forecasting oil price realized volatility using information channels from other asset classes. *Journal of International Money and Finance*, *76*, 28–49.
Degiannakis, S., & Filis, G. (2018). Forecasting oil prices: High-frequency financial data are indeed useful. *Energy Economics*, *76*, 388–402.
Degiannakis, S., Filis, G., Klein, T., & Walther, T. (2020). Forecasting realized volatility of agricultural commodities. *International Journal of Forecasting*.
Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987–1007.
Escanciano, J. C., & Lobato, I. N. (2009). An automatic portmanteau test for serial correlation. *Journal of Econometrics*, *151*(2), 140–149.
Fiszeder, P., & Fałdziński, M. (2019). Improving forecasts with the co-range dynamic conditional correlation model. *Journal of Economic Dynamics and Control*, Article 103736.
Fiszeder, P., Fałdziński, M., & Molnár, P. (2019). Range-based DCC models for covariance and value-at-risk forecasting. *Journal of Empirical Finance*, *54*, 58–76.
Gallant, A., Hsu, C.-T., & Tauchen, G. (1999). Using daily range data to calibrate volatility diffusions and extract the forward integrated variance. *The Review of Economics and Statistics*, *81*(4), 617–631.
Garman, M. B., & Klass, M. J. (1980). On the estimation of security price volatilities from historical data. *Journal of Business*, *53*(1), 67–78.
Ghalanos, A. (2020). *Introduction to the rugarch package.(Version 1.3-8): Technical report v*, Available at http://cran.r-project.org/web/packages....
Granger, C. W., & Joyeux, R. (1980). An introduction to long-memory time series models and fractional differencing. *Journal of Time Series Analysis*, *1*(1), 15–29.
Hansen, P. R., Huang, Z., & Shek, H. H. (2012). Realized GARCH: a joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, *27*(6), 877–906.
Hansen, P. R., & Lunde, A. (2005). A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Economic*, *3*(4), 525–554.
Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Haugom, E., Langeland, H., Molnár, P., & Westgaard, S. (2014). Forecasting volatility of the US oil market. *Journal of Banking & Finance*, *47*, 1–14.

Heber, G., Lunde, A., Shephard, N., & Sheppard, K. (2009). *Oxford-Man Institute's realized library, version 0.1*. Oxford-Man Institute, University of Oxford Oxford.

Horpestad, J. B., Lyócsa, Š., Molnár, P., & Olsen, T. B. (2019). Asymmetric volatility in equity markets around the world. *The North American Journal of Economics and Finance*, *48*, 540–554.

Hosking, J. (1981). Fractional differencing. Biometrika 68 165–176. *Mathematical Reviews (MathSciNet): MR614953 Zentralblatt MATH*, *464*.

Johnson, N. L. (1949a). Bivariate distributions based on simple translation systems. *Biometrika*, *36*(3/4), 297–304.

Johnson, N. L. (1949b). Systems of frequency curves generated by methods of translation. *Biometrika*, *36*(1/2), 149–176.

Kim, N., Lučivjanská, K., Molnár, P., & Villa, R. (2019). Google searches and stock market activity: Evidence from Norway. *Finance Research Letters*, *28*, 208–220.

Klein, T., & Walther, T. (2016). Oil price volatility forecast with mixture memory GARCH. *Energy Economics*, *58*, 46–58.

Koopman, S. J., Jungbacker, B., & Hol, E. (2005). Forecasting daily variability of the S&P 100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance*, *12*(3), 445–475.

Ledoit, O., & Wolf, M. (2008). Robust performance hypothesis testing with the Sharpe ratio. *Journal of Empirical Finance*, *15*(5), 850–859.

Liu, L. Y., Patton, A. J., & Sheppard, K. (2015). Does anything beat 5-minute RV? A comparison of realized measures across multiple asset classes. *Journal of Econometrics*, *187*(1), 293–311.

Lo, A. W. (2002). The statistics of sharpe ratios. *Financial Analysts Journal*, *58*(4), 36–52.

Luo, J., Klein, T., Ji, Q., & Hou, C. (2019). Forecasting realized volatility of agricultural commodity futures with infinite hidden Markov HAR models. *International Journal of Forecasting*.

Lyócsa, Š., & Molnár, P. (2017). The effect of non-trading days on volatility forecasts in equity markets. *Finance Research Letters*, *23*, 39–49.

Lyócsa, Š., & Molnár, P. (2018). Exploiting dependence: Day-ahead volatility forecasting for crude oil and natural gas exchange-traded funds. *Energy*, *155*, 462–473.

Lyócsa, Š., Molnár, P., & Todorova, N. (2017). Volatility forecasting of non-ferrous metal futures: Covariances, covariates or combinations? *Journal of International Financial Markets, Institutions and Money*, *51*, 228–247.

Lyócsa, Š., & Todorova, N. (2020). Trading and non-trading period realized market volatility: Does it matter for forecasting the volatility of US stocks?. *International Journal of Forecasting*, *36*(2), 628–645.

Ma, F., Li, Y., Liu, L., & Zhang, Y. (2018). Are low-frequency data really uninformative? A forecasting combination perspective. *The North American Journal of Economics and Finance*, *44*, 92–108.

Ma, F., Wahab, M., Huang, D., & Xu, W. (2017). Forecasting the realized volatility of the oil futures market: A regime switching approach. *Energy Economics*, *67*, 136–145.

Martens, M. (2001). Forecasting daily exchange rate volatility using intraday returns. *Journal of International Money and Finance*, *20*(1), 1–23.

Molnár, P. (2012). Properties of range-based volatility estimators. *International Review of Financial Analysis*, *23*, 20–29.

Molnár, P. (2016). High-low range in GARCH models of stock return volatility. *Applied Economics*, *48*(51), 4977–4991.

Nelson, D. B. (1991). Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, *59*(2), 347–370.

Newey, W. K., & West, K. D. (1994). Automatic lag selection in covariance matrix estimation. *Review of Economic Studies*, *61*(4), 631–653.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business*, *53*(1), 61–65.

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

Patton, A. J., & Sheppard, K. (2009). Optimal combinations of realised volatility estimators. *International Journal of Forecasting*, *25*(2), 218–238.

Patton, A. J., & Sheppard, K. (2015). Good volatility, bad volatility: Signed jumps and the persistence of volatility. *The Review of Economics and Statistics*, *97*(3), 683–697.

Politis, D. N., & Romano, J. P. (1994). The stationary bootstrap. *Journal of the American Statistical association*, *89*(428), 1303–1313.

Rogers, L. G., & Satchell, S. E. (1991). Estimating variance from high, low and closing prices. *Annals of Applied Probability*, *1*(4), 504–512.

Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of Forecasting*, *23*(6), 405–430.

Timmermann, A. (2006). Forecast combinations. *Handbook of Economic Forecasting*, *1*, 135–196.

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

Wei, Y. (2012). Forecasting volatility of fuel oil futures in China: GARCH-type, SV or realized volatility models? *Physica A. Statistical Mechanics and its Applications*, *391*(22), 5546–5556.

Welch, I., & Goyal, A. (2007). A comprehensive look at the empirical performance of equity premium prediction. *Review of Financial Studies*, *21*(4), 1455–1508.

Zhang, Y., Ma, F., Wang, T., & Liu, L. (2019). Out-of-sample volatility prediction: A new mixed-frequency approach. *Journal of Forecasting*, *38*(7), 669–680.