Global Finance Journal 36 (2018) 41–61

Contents lists available at ScienceDirect

Global Finance Journal

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

Multiple days ahead realized volatility forecasting: Single, combined and average forecasts

Stavros Degiannakis

Department of Economic and Regional Development, Panteion University, 136 Syggrou Av., Athens 176 71, Greece

ARTICLE INFO

JEL classifications:
C14
C32
C50
G11
G15

Keywords:
Averaging forecasts
Combining forecasts
Heterogeneous autoregressive
Intra-day data
Long memory
Model confidence set
Predictive ability
Realized volatility
Ultra-high frequency

ABSTRACT

The task of this paper is the enhancement of realized volatility forecasts. We investigate whether a mixture of predictions (either the combination or the averaging of forecasts) can provide more accurate volatility forecasts than the forecasts of a single model. We estimate long-memory and heterogeneous autoregressive models under symmetric and asymmetric distributions for the major European Union stock market indices and the exchange rates of the Euro.

The majority of models provide qualitatively similar predictions for the next trading day's volatility forecast. However, with regard to the one-week forecasting horizon, the heterogeneous autoregressive model is statistically superior to the long-memory framework. Moreover, for the two-weeks-ahead forecasting horizon, the combination of realized volatility predictions increases the forecasting accuracy and forecast averaging provides superior predictions to those supplied by a single model. Finally, the modeling of volatility asymmetry is important for the two-weeks-ahead volatility forecasts.

1. Introduction

Undoubtedly, ultra-high frequency financial data have been valuable in estimating and forecasting volatility more accurately. The long-memory autoregressive and the heterogeneous autoregressive models are representative methods of volatility forecasting.

The literature provides strong evidence that ARFIMA models introduced by Granger (1980), produce superior forecasts relative to those produced by conditional volatility GARCH models that are based on daily returns. Due both to the long memory property of volatility as well as its high persistence, the ARFIMA specification is suitable for estimating realized volatility. Among others, Andersen, Bollerslev, Diebold, and Labys (2003), Chiriac and Voev (2011), Deo, Hurvich, and Lu (2005), Koopman, Jungbacker, and Hol (2005), Martens and Zein (2002), Pong, Shackleton, Taylor, and Xu (2004) have applied various extensions of ARFIMA models to ultra-high frequency-based volatility measures.

The structure of the Heterogeneous Autoregressive model of realized volatility is based on the heterogeneous market hypothesis (Müller et al., 1997), which states that in financial markets, investors (ultra-high frequency algorithmic traders, inter-day investors, institutional investors trading on a monthly basis, etc.) interact at different frequencies. Thus, the HAR model is able to accommodate the heterogeneous beliefs of traders; different types of market participants drive volatility at different frequencies. Andersen, Bollerslev, and Diebold (2007) show that volatility for equity and bond futures is adequately expressed by a HAR-GARCH model. In forecasting ultra-high frequency constructed volatility, various extensions of the HAR model have been applied by Chen and Ghysels

E-mail address: s.degiannakis@panteion.gr.

https://doi.org/10.1016/j.gfj.2017.12.002
Received 30 June 2017; Received in revised form 5 September 2017; Accepted 10 December 2017
Available online 13 December 2017
1044-0283/

(2011), Clements, Galvão, and Kim (2008), Corsi and Reno (2012), Hua and Manzan (2013), Prokopczuk, Symeonidis, and Wese Simen (2015), Sévi (2014) and Degiannakis and Filis (2017). In general, the literature provides evidence in favor of the HAR model compared to other models such as the plain autoregressive model, the MIDAS model of Ghysels, Sinko, and Valkanov (2007), the HEAVY model of Shephard and Sheppard (2010), the ARFIMA model, etc.

Apart from modeling information of realized volatility from the past, an alternative approach is to extract the predictive information from the futures market. Such techniques have been employed mainly by policy institutions,[1] which are looking for the market expectations of the exogenous variables required for their macroeconomic model frameworks. Alquist and Kilian (2010) provided an interesting analysis of oil price forecasts based on futures prices. Their study showed that futures are not the most accurate predictor of the spot price of crude oil; even no-change forecasts tend to be more accurate.

Additionally, the implied volatility extracted from the option prices has been considered as an alternative source of measuring investor sentiment with regard to market volatility. Koopman et al. (2005) showed that models based on realized volatility (i.e. ARFIMA models) outperform models based on implied volatility. On the other hand, Fleming, Ostdiek, and Whaley (1995), Christensen and Prabhala (1998), Fleming (1998), Blair, Poon, and Taylor (2001), Giot (2003), Degiannakis (2008) and Frijns, Tallau, and Tourani-Rad (2010) provided evidence that implied volatility is more informative when stock market volatility is being investigated.

Although model-averaging methods for forecasting purposes date back to the works of Bates and Granger (1969), Granger and Newbold (1977) and Granger and Ramanathan (1984), the combination of volatility forecasts has not been broadly studied. Liu and Maheu (2009) and Wang, Ma, Wei, and Wu (2016) have investigated the impact of model averaging on realized volatility prediction accuracy, while Amendola and Storti (2008) and Hu and Tsoukalas (1999) have examined the performance of combining forecasts estimated from conditional volatility models (i.e. based on daily data). However, the performance of combined forecasts has not been explored for ultra-high frequency based volatility estimates.

This paper studies whether the combination or the averaging of realized volatility predictions increases forecasting accuracy. It brings to light two strands of mixed predictions (i) selecting forecasts from a set of candidate models according to an evaluation criterion; and (ii) the averaging of forecasts.

This forecasting evaluation exercise is not limited to one-day-ahead forecasts, as multiple-days-ahead forecasts (i.e. one-week and two-weeks-ahead forecasting horizons) gather investor interest as well. Moreover, we investigate the predictive accuracy under four different distributions for the standardized unpredictable component of the models. Briefly, our results conclude that: 1) The heterogeneous autoregressive framework works better than the long memory framework. 2) The averaged models provide superior forecasts compared to those of single models. 3) The modeling of volatility asymmetry is crucial in forecasting the ten-days-ahead realized volatility. 4) The combination of volatility forecasts according to the statistical properties of forecast errors provides us with more accurate two-weeks-ahead volatility forecasts compared to forecasts from a single model.

The remainder of this study is structured as follows. Section 2 describes the estimation of the realized volatility measures, Section 3 provides information for the dataset of the 3 stock market indices and the 3 exchange rates, while Sections 4 and 5 demonstrate the ARFIMA and HAR estimated models and the relative forecast specifications for one-day and multiple-days-ahead horizons. Sections 6 and 7 present methods of combining predictions according to model selection criteria and methods of computing the model-average forecasts, respectively. Section 8 describes a unified framework for the evaluation of all predictive methods. Section 9 reports the empirical results and suggests when we should apply the volatility forecasts of a single model, a combination of models, or the average forecast from a set of models. Section 10 concludes the paper and suggests areas for further research.

2. The realized volatility measure

The financial literature assumes that the instantaneous logarithmic price $p(t)$ of an asset follows a diffusion process $d\log p(t) = \sigma(t)dW(t)$, where $\sigma(t)$ is volatility and $W(t)$ is the Wiener process. The integrated variance $\sigma_{[t_1, t_\tau]}{}^{2(IV)}$ is the actual, but unobservable, variance over the interval $[t_1, t_\tau]$ for which we seek a proxy measure to estimate. Assuming that the number of points in time tends to infinity, $\tau \to \infty$, we are able to approximate the integrated variance as $\sigma_{[a,b]}^{2(IV)} = \int_{t_1}^{t_2} \sigma^2(t)dt + \int_{t_2}^{t_3} \sigma^2(t)dt + \ldots + \int_{t_{\tau-1}}^{t_\tau} \sigma^2(t)dt$. The realized volatility for the time interval $[t_1, t_\tau]$ which is partitioned in τ equidistance points, $RV_{[t_1,t_\tau]} = \sum_{j=1}^{\tau} (\log P_{t_j} - \log P_{t_{j-1}})^2$ converges in probability towards the integrated volatility,[2] or $\underset{\tau\to\infty}{p\lim}\,(RV_{[t_1,t_\tau]}) = \sigma_{[t_1,t_\tau]}^{2(IV)}$. Accuracy improves as the number of sub-intervals increases, or as $\tau \to \infty$, but on the other hand, at a high sampling frequency, such as $sf \to 0$, market friction is a source of noise due to market microstructure features (i.e. discreteness of the data, transaction costs, properties of the trading mechanism, bid-ask spreads, etc.). Thus, realized volatility is constructed in the highest sampling frequency which the intra-day autocovariance minimizes[3]; see e.g. Andersen, Bollerslev, Christoffersen, and Diebold

[1] The interested reader is referred to Svensson (2005) for the European Central Bank, and to the International Monetary Fund (2007) for the International Monetary Fund.

[2] The $p(t)$ is the latent efficient price, whereas P_{t_j} is the observed price. The unobserved distance between $p(t)$ and P_{t_j} is the market microstructure noise. There exist a number of estimators for the integrated volatility that possess asymptotical properties which are robust for microstructure noise and jumps. However, Sévi (2014) and Prokopczuk et al. (2015) provided empirical evidence that the modeling of jumps does not improve the forecast accuracy of the simple HAR-RV model. Thus, we construct the realized volatility estimates without taking into consideration the presence of jumps.

[3] The inter-day variance can be decomposed into the intra-day variance, $RV_t^{(\tau)}$, and the intra-day autocovariances $y_{t_i}y_{t_{i-j}}$: $y_t^2 = RV_t^{(\tau)} + 2\sum_{j=1}^{\tau-1}\sum_{i=j+1}^{\tau} y_{t_i}y_{t_{i-j}}$. As the autocovariance comprises a measurement error, its expected value equals to zero, $E(y_{t_i}y_{t_{i-j}}) = 0$, for $j \neq 0$.

Table 1
Information for the intra-day data.

Index	Number of intra-day (1 min) observations	Number of trading days	First day	Last day	*Optimal* sampling frequency, $sf^{(o)}$
CAC 40	1,403,509	2708	13th June 2000	12th January 2011	7 min
DAX 30	1,433,751	2806	3rd January 2000	12th January 2011	13 min
FTSE100	1,576,347	3128	20th August 1998	12th January 2011	7 min
EURUSD	4,622,271	3330	20th April 1998	24th January 2011	20 min
EURGBP	4,302,166	3113	4th January 1999	21st January 2011	20 min
EURJPY	4,399,091	3131	4th January 1999	24th January 2011	20 min

(2006), and Degiannakis and Floros (2015). The sequence of the sampling prices is constructed according to the previous tick method[4] of Wasserfallen and Zimmermann (1985).

In order to incorporate estimates of asset prices during the hours that the stock markets are closed to volatility, we take into consideration Hansen and Lunde's (2005) method of combining intraday volatility with closed-to-open inter-day volatility. Hansen and Lunde proposed the construction of the realized volatility measure as a weighted combination of $RV_{[t_1,t_\tau]}$ with inter-day volatility during the time that the market is closed; $(\log P_{t_1} - \log P_{t-1_\tau})^2$. Hence, we estimate $RV_t^{(\tau)} = \omega_1(\log P_{t_1} - \log P_{t-1_\tau})^2 + \omega_2 \sum_{j=1}^{\tau}(\log P_{t_j} - \log P_{t_{j-1}})^2$, such as $\min\limits_{(\omega_1,\omega_2)} E(RV_t^{(\tau)} - \sigma_{[t_1,t_\tau]}^{2(IV)})^2$. As the $\sigma_{[t_1},t_{\tau]}^{2(IV)}$ is unobservable, Hansen and Lunde (2005) provide the analytic solution of $\min\limits_{(\omega_1,\omega_2)} V(RV_t^{(\tau)})$ instead of $\min\limits_{(\omega_1,\omega_2)} E(RV_t^{(\tau)} - \sigma_{[t_1,t_\tau]}^{2(IV)})^2$, as both functions lead to the same solution. Hence, we minimize the squared distance between the realized volatility measure and integrated volatility, avoiding the need to define a specific relation between efficient prices and market microstructure noise.

3. Dataset - FTSE100, DAX30, CAC40 and euro exchange rates

The database is made up of the three most liquid euro exchange rates (with the Pound, the Dollar and the Yen) and the three major European stock indices (FTSE100, DAX30, CAC40). The Euro, the Pound, the Dollar and the Yen are the four most tradable currencies. The blue-chip FTSE 100 from the London Stock Exchange, has a market cap of €1.8 trillion, the DAX30 (a market cap of €1 trillion) from the Deutsche Boerse group, is Germany's prime index featuring many of Europe's biggest companies, and the CAC40 (market cap of €1.2 trillion) represents a capitalization-weighted measure of the 40 most significant companies listed on the Euronext Paris (formerly Paris Bourse).

Table 1 presents information for the one-minute intra-day data of the FTSE100, DAX30, and CAC40 indices, as well as for the exchange rates of Euro with the British Pound, the US Dollar and the Japanese Yen. The data are filtered for detecting data errors due to computer technical failures, typing errors, sequences of zero or non-available prices due to databases crashes, etc. Weekends and fixed and moving holidays with thin trading activity have been deleted. The selection of the optimal sampling frequency $sf^{(o)}$, is based on a trade-off between accuracy and potential biases due to market microstructure frictions (last column of Table 1).[5] The interday adjustment of Hansen and Lunde (2005) is taken into consideration. Fig. 1 plots the annualized realized volatilities, $\sqrt{252RV_t^{(\tau)}}$ and the empirical density functions of $\log\sqrt{252RV_t^{(\tau)}}$.

The logarithmic transformation of realized volatility has an ogive empirical distribution which approximates the Gaussian distribution. The average value of the annualized standard deviation for the three stock indices is 18.8% (see Table 2). The mean of the annualized standard deviation for the three euro exchange rates is 10.1%. The maximum annualized volatility observed for the FTSE100 index was 167%, on Friday, October 10, 2008 (Global financial crisis of October 2008). On Friday, October 10, the stock markets crashed across Europe and Asia. London, Paris and Frankfurt dropped 10% in the first hour of trading and this also happened when Wall Street opened for trading. Since 1987, global markets have experienced some of their worst weeks in memory, and indeed in some cases, since the Wall Street Crash of 1929. The median value of annualized volatility ranges from 13.3% for the FTSE100 to 17.8% for the CAC40. On the other hand, realized standard deviations of exchange rates do not fluctuate over time at a similar magnitude. The median value of annualized volatility ranges from 7.5% for the Euro/Pound rate to 10.2% for the Euro/Yen rate. Maximum annualized volatility is observed for the Euro/Yen exchange rate to 74%, on October 24, 2008 (when recession fears caused great turbulence in the Euro/Yen rate). Table 3 provides descriptive statistics of the annualized logarithmic realized volatility. Sample skewness is positive in all cases. The average of the skewness of log-standard deviations across the stock indices decreases to 0.3 compared to 2.9 for the realized standard deviations. As far as kurtosis is concerned, the average value for the log-volatilities, across the stock indices, is 3.1 compared to 19.5 for the realized standard deviations. Therefore, although the kurtosis of the indices exceeds the normal value of three, the logarithmic transformation case is obviously much closer to the assumption of normality. Normality approximation is very good for the log-volatilities of the exchange rates as well.

[4] Based on the previous tick method (e.g. employ the most recently published price), we obtain a volatility measure that does not converge in probability to zero (see e.g. Hansen & Lunde, 2006).

[5] We follow Andersen et al. (2006) who proposed the construction of the volatility signature plot.

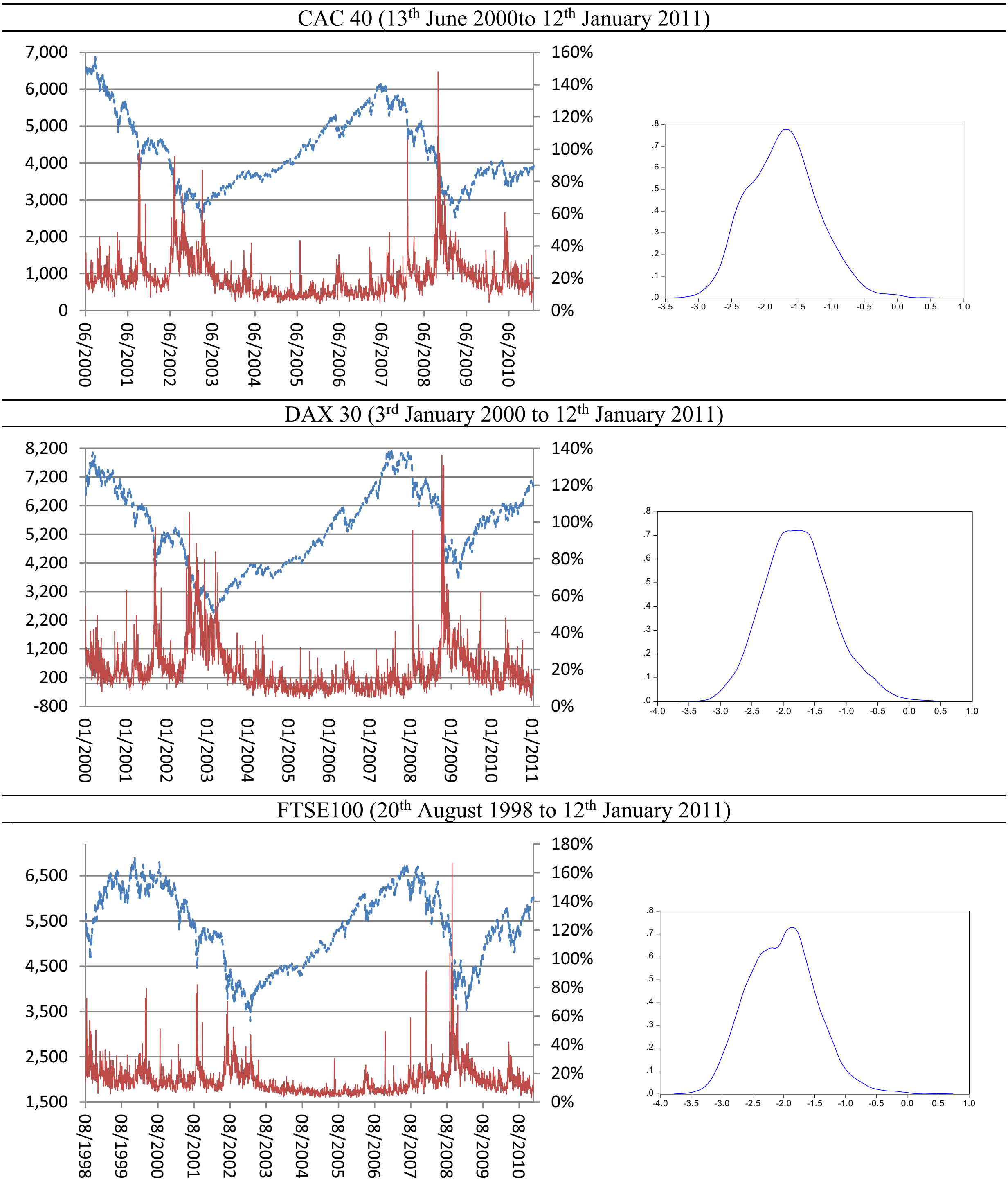

Fig. 1. The annualized realized volatility, $\sqrt{252RV_t^{(\tau)}}$, the daily prices P_t, and the empirical density function of $\log \sqrt{252RV_t^{(\tau)}}$.
The figures in the left column present the daily prices (dash line presented in the LHS axis) and the annualized realized volatility (solid line presented in the RHS axis). The figures in the right column present the empirical density function of $\log \sqrt{252RV_t^{(\tau)}}$.

4. Estimation of the models

We proceed to an estimation of two widely accepted model frameworks for the annualized logarithmic realized volatility, $\log \sqrt{252RV_t^{(\tau)}}$. The first framework is the Autoregressive Fractionally Integrated Moving Average, or the ARFIMA model with time-

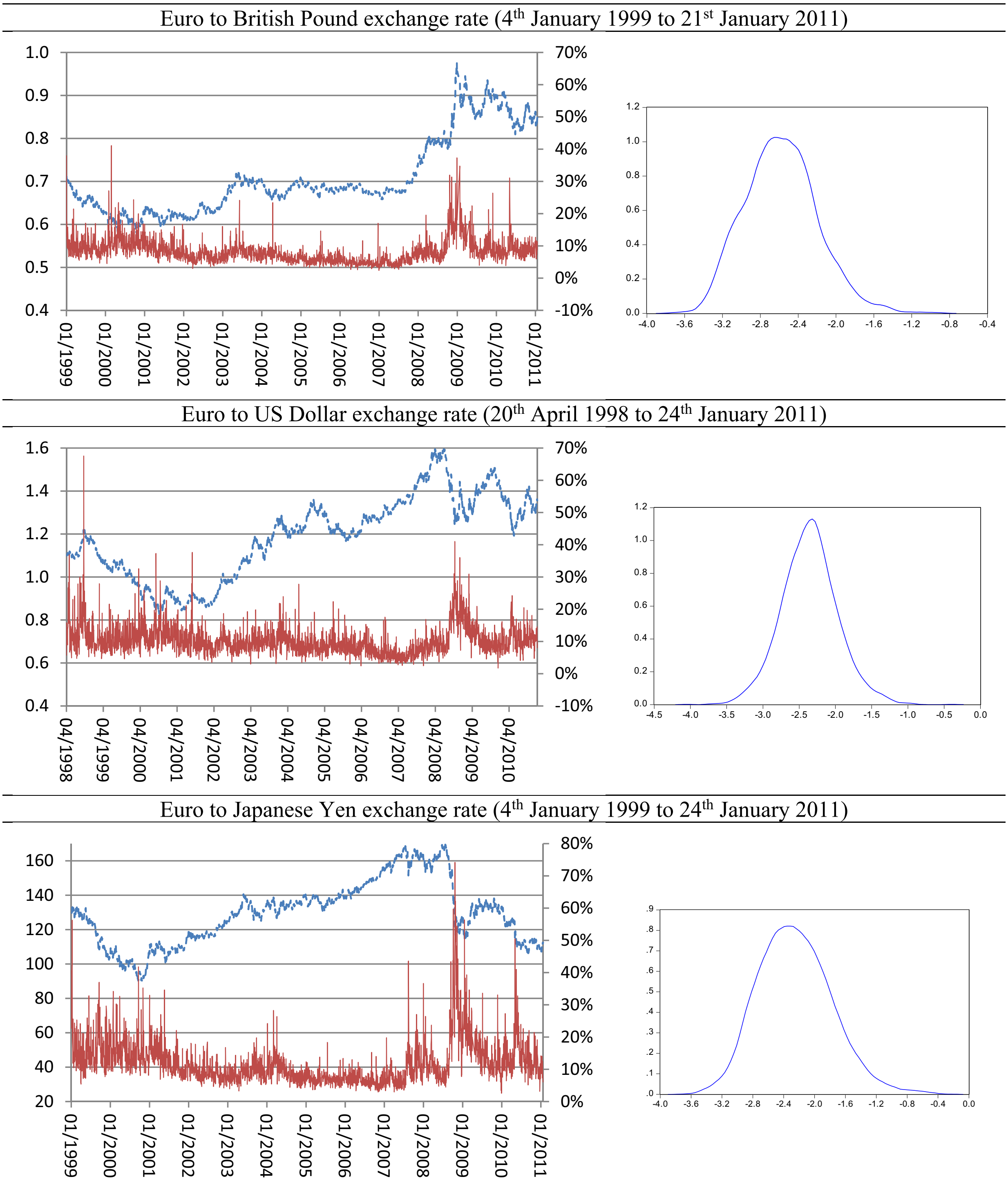

Fig. 1. (*continued*)

varying conditional innovations. The ARFIMA, initially developed by Granger (1980) and Granger and Joyeux (1980), captures the long-memory property of dependent variables. The time-variation and clustering that the volatility of realized volatility exhibits are modeled by extending the ARFIMA to the ARFIMA-GARCH framework proposed by Baillie, Bollerslev, and Mikkelsen (1996). The ARFIMA(k, d, l)-GARCH(p,q) model for $\log\sqrt{252RV_t^{(\tau)}}$ is defined as:

Table 2
Descriptive statistics of annualized one-trading-day inter-day adjusted realized daily volatility, $\sqrt{252RV_t^{(\tau)}}$.

Index	Mean[a]	Median[a]	Maximum[a]	Minimum[a]	Std.Dev[a]	Skewness	Kurtosis
CAC 40	20.6	17.8	148.1	4.1	12.5	2.5	14.6
DAX 30	20.4	17.0	136.2	3.3	13.3	2.6	14.3
FTSE100	15.4	13.3	166.9	2.9	10.2	3.5	29.7
EURUSD	10.2	9.5	67.5	1.8	4.4	2.2	16.0
EURGBP	8.2	7.5	41.1	2.4	3.6	2.2	13.1
EURJPY	11.9	10.2	74.2	2.6	6.7	2.6	14.5

[a] The numbers are expressed in percentages.

Table 3
Descriptive statistics of annualized inter-day adjusted logarithmic realized volatility, $\log\sqrt{252RV_t^{(\tau)}}$.

Index	Mean	Median	Maximum	Minimum	Std.Dev	Skewness	Kurtosis
CAC 40	2.88	2.88	5.00	1.40	0.52	0.27	3.01
DAX 30	2.85	2.83	4.91	1.18	0.55	0.34	3.16
FTSE100	2.58	2.59	5.12	1.05	0.54	0.29	3.18
EURUSD	2.25	2.25	4.21	0.56	0.39	0.13	3.65
EURGBP	2.02	2.01	3.72	0.86	0.39	0.37	3.37
EURJPY	2.35	2.33	4.31	0.94	0.48	0.39	3.31

$$
\begin{aligned}
(1 - C(L))(1 - L)^d\left(\log\sqrt{252RV_t^{(\tau)}} - \beta_0\right) &= (1 + D(L))\varepsilon_t \\
\varepsilon_t &= h_t z_t \\
h_t^2 &= a_0 + A(L)\varepsilon_t^2 + B(L)h_t^2 \\
z_t &\sim f(0, 1; \boldsymbol{\theta}),
\end{aligned}
\tag{1}
$$

where $C(L) = \sum_{i=1}^{k} c_i L^i$, $D(L) = \sum_{i=1}^{l} d_i L^i$, $A(L) = \sum_{i=1}^{q} a_i L^i$, $B(L) = \sum_{i=1}^{p} b_i L^i$ are polynomials, $f(.)$ is the density function of z_t (with $E(z_t) = 0$, $V(z_t) = 1$, $\boldsymbol{\theta}$ is the vector of the parameters which define f) and d, β_0, c_1, …, c_k, d_1, …, d_l, a_1, …, a_q, b_1, …, b_p, $\boldsymbol{\theta}$ are the parameters to be estimated. The h_t^2 can be considered as an estimate of the integrated quarticity $\sigma_t^{2(IQ)}$[6].

The second framework is the Heterogeneous Autoregressive, or HAR model with time-varying conditional innovations; e.g. of Corsi, Mittnik, Pigorsch, and Pigorsch (2008) and Corsi (2009). The basic idea is that market participants have a different perspective of their investment horizon. The HAR-RV-GARCH(p,q) model is an autoregressive structure of the volatilities realized over different interval sizes:

$$
\begin{aligned}
\log\sqrt{252RV_t^{(\tau)}} &= w_0 + w_1 \log\sqrt{252RV_{t-1}^{(\tau)}} + w_2\left(5^{-1}\sum_{j=1}^{5}\log\sqrt{252RV_{t-j}^{(\tau)}}\right) \\
&\quad + w_3\left(22^{-1}\sum_{j=1}^{22}\log\sqrt{252RV_{t-j}^{(\tau)}}\right) + \varepsilon_t, \\
\varepsilon_t &= h_t z_t, \\
h_t^2 &= a_0 + A(L)\varepsilon_t^2 + B(L)h_t^2 \\
z_t &\sim f(0, 1;; \boldsymbol{\theta}),
\end{aligned}
\tag{2}
$$

where $A(L) = \sum_{i=1}^{q} a_i L^i$, $B(L) = \sum_{i=1}^{p} b_i L^i$ are polynomials and w_0, …, w_3, a_1, …, a_q, b_1, …, b_p, $\boldsymbol{\theta}$ are the parameters to be estimated.

For the 3 stock market indices and the 3 exchange rates, the ARFIMA(0,d,1)-GARCH(1,1), ARFIMA(1,d,1)-GARCH(1,1), HAR-RV-GARCH(1,1) and HAR-RV-GARCH(0,1) model specifications with innovations (i.e. unexplained component of conditional mean equation) that are i) normally distributed; $z_t \sim N(0,1)$ ii) Student t distributed; $z_t \sim t(0,1;\nu)$, iii) GED distributed $z_t \sim Ged(0,1;\nu)$, and iv) skewed Student t distributed; $z_t \sim skT(0,1;\nu,g)$, are estimated. For $z_t \sim N(0,1)$, the density function is $f_{(N)}(z_t) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{z_t^2}{2}\right)$. Under the assumption of conditional Student t distributed innovations $z_t \sim t(0,1;\nu)$, the density function is:

$$
f_{(t)}(z_t;\nu) = \frac{\Gamma((\nu+1)/2)}{\Gamma(\nu/2)\sqrt{\pi(\nu-2)}}\left(1 + \frac{z_t^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}, \text{ for} \nu > 2,
\tag{3}
$$

[7]where $\Gamma(.)$ is the gamma function. With conditional GED (Generalized Error Distribution or Exponential Power distribution)

[6] The asymptotic volatility of volatility, $\sigma_{[t_1,t_\tau]}^{2(IQ)}$, is termed *integrated quarticity*: $\sigma_{[t_1,t_\tau]}^{2(IQ)} = \int_{t_1}^{t_\tau} 2\sigma^4(t)dt$, as $\sqrt{\tau}\left(RV_{[t_1,t_\tau]} - \int_{t_1}^{t_\tau}\sigma^2(t)dt\right)/\sqrt{\int_{t_1}^{t_\tau} 2\sigma^4(t)dt} \xrightarrow{d} N(01)$.

[7] $\boldsymbol{\theta} = [\nu]$

Table 4

The mean predictive squared error $10^3 MPSE_{(1)}{}^{(m)}$, of the four models for conditionally i) normally; ii) Student *t*; ii) GED; and iv) skewed Student *t* distributed innovations. The $MPSE_{(1)}{}^{(distr)}$ i) from combining the forecasts of the four models under the same distribution according to the criteria $\min_{m=1,\dots,M}(\varepsilon_{t|t-1}^{2(m)})$ and $\min_{m=1,\dots,M}(z_{t|t-1}^{2(m)})$; ii) from averaging the forecasts of the four models under the same distribution ($AV_{t+1|t}{}^{(distr)}$); and iii) of the overall averaging forecast ($AV_{t+1|t}$).

Model	For $z_t \sim N(0,1)$	For $z_t \sim t(0,1;v)$	For $z_t \sim Ged(0,1;v)$	For $z_t \sim skT(0,1;v,g)$
CAC 40				
1	43.30	44.03	44.05	43.95
2	43.01	43.68	43.85	43.57
3	43.33	43.91	44.06	43.67
4	43.21	43.68	44.06	43.55
$\min_{m=1,\dots,M}(\varepsilon_{t\|t-1}^{2(m)})$	42.99	43.39	43.57	43.49
$\min_{m=1,\dots,M}(z_{t\|t-1}^{2(m)})$	42.93	43.71	43.47	43.62
$AV_{t+1\|t}{}^{(distr)}$	43.07	43.69	43.88	43.54
$AV_{t+1\|t}$	43.53			
DAX30				
1	49.14	49.08	49.19	49.04
2	48.78	–	48.48	48.56
3	49.72	49.58	49.51	49.42
4	49.38	49.48	49.51	49.32
$\min_{m=1,\dots,M}(\varepsilon_{t\|t-1}^{2(m)})$	48.63	49.03	48.73	48.67
$\min_{m=1,\dots,M}(z_{t\|t-1}^{2(m)})$	48.84	49.00	48.74	48.90
$AV_{t+1\|t}{}^{(distr)}$	49.13	49.16	49.05	48.94
$AV_{t+1\|t}$	49.06			
FTSE100				
1	38.16	38.42	38.41	38.27
2	38.12	38.33	38.37	38.24
3	38.27	38.48	38.58	38.35
4	38.01	38.42	38.58	38.33
$\min_{m=1,\dots,M}(\varepsilon_{t\|t-1}^{2(m)})$	37.76	38.08	38.44	38.03
$\min_{m=1,\dots,M}(z_{t\|t-1}^{2(m)})$	37.92	38.22	38.39	38.09
$AV_{t+1\|t}{}^{(distr)}$	37.99	38.29	38.37	38.18
$AV_{t+1\|t}$	38.20			
EURUSD				
1	7.94	7.89	7.90	7.88
2	7.88	7.87	7.87	7.86
3	7.89	7.89	7.89	7.88
4	7.89	7.89	7.89	7.88
$\min_{m=1,\dots,M}(\varepsilon_{t\|t-1}^{2(m)})$	7.83	7.82	7.81	7.82
$\min_{m=1,\dots,M}(z_{t\|t-1}^{2(m)})$	7.83	7.83	7.81	7.83
$AV_{t+1\|t}{}^{(distr)}$	7.88	7.86	7.87	7.86
$AV_{t+1\|t}$	7.87			
EURGBP				
1	4.97	4.89	4.96	4.95
2	5.04	4.85	4.92	4.87
3	4.91	4.92	4.93	4.93
4	4.90	4.92	4.93	4.92
$\min_{m=1,\dots,M}(\varepsilon_{t\|t-1}^{2(m)})$	4.94	4.91	4.89	4.93
$\min_{m=1,\dots,M}(z_{t\|t-1}^{2(m)})$	4.94	4.92	4.91	4.95
$AV_{t+1\|t}{}^{(distr)}$	4.92	4.88	4.91	4.90
$AV_{t+1\|t}$	4.90			
EURJPY				
1	16.69	16.78	16.71	16.78
2	16.29	16.35	16.32	16.37
3	16.34	16.43	16.35	16.34
4	16.47	16.49	16.35	16.46
$\min_{m=1,\dots,M}(\varepsilon_{t\|t-1}^{2(m)})$	16.16	16.25	16.14	16.19

(*continued on next page*)

Table 4 (*continued*)

Model	For $z_t \sim N(0,1)$	For $z_t \sim t(0,1;v)$	For $z_t \sim Ged(0,1;v)$	For $z_t \sim skT(0,1;v,g)$
$\min_{m=1,\ldots,M}(z_{t\mid t-1}^{2(m)})$	16.29	16.38	16.27	16.29
$AV_{t+1\mid t}^{(distr)}$	16.38	16.44	16.36	16.41
$AV_{t+1\mid t}$	16.40			

Model 1: ARFIMA(0,d,1)-GARCH(1,1), Model 2: ARFIMA(1,d,1)-GARCH(1,1), Model 3: HAR-RV-GARCH(1,1), Model 4: HAR-RV-GARCH(0,1).

distributed innovations $\{z_t\}_{t=1}^{T} \sim Ged(0,1;v)$, the density function is:

$$f_{(GED)}(z_t;v) = \frac{v\exp(-0.5\ |z_t/\lambda|^v)}{\lambda 2^{\left(1+1/v\right)}\Gamma(v^{-1})},\ v>0, \tag{4}$$

where v is the tail-thickness parameter and$\lambda \equiv \sqrt{2^{-2/v}\Gamma(v^{-1})/\Gamma(3v^{-1})}$.[8] For $z_t \sim skT(0,1;v,g)$ the density function is[9]:

$$f_{(skT)}(z_t;v,g) = \begin{cases} \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{\pi(v-2)}}\left(\frac{2s}{g+g^{-1}}\right)\left(1+\frac{sz_t+\zeta}{v-2}g\right)^{-\frac{v+1}{2}} & if\ z_t < -\zeta s^{-1}, \\ \frac{\Gamma((v+1)/2)}{\Gamma(v/2)\sqrt{\pi(v-2)}}\left(\frac{2s}{g+g^{-1}}\right)\left(1+\frac{sz_t+\zeta}{v-2}g^{-1}\right)^{-\frac{v+1}{2}} & if\ z_t \geq -\zeta s^{-1}, \end{cases} \tag{5}$$

where$g$ and v are the asymmetry and tail parameters, respectively, of the distribution[10] $\zeta = \Gamma((v-1)/2)\sqrt{(v-2)}\,(\Gamma(v/2)\sqrt{\pi})^{-1}(g-g^{-1})$, and $s = \sqrt{g^2+g^{-2}-\zeta^2-1}$.

For each of the six time-series, 16 models are estimated; four model specifications combined with four distributional assumptions. The lag orders k, d, l, p, q of the models have been selected according to Schwarz's (1978) Bayesian information criterion[11]. Each of the 16 models is re-estimated every trading day t, for $\tilde{T}$ days, where $\tilde{T}$=1686, 1784, 2106, 2308, 2091, 2108 for the CAC40, DAX30, FTSE100, EURUSD, EURGBP and EURJPY realized volatility series, respectively based on a rolling sample of constant size $\breve{T}$ = 1000 days. For the ARFIMA(1,d,1)-GARCH(1,1) model, the parameter vector to be estimated at each point in time t is $(\beta_0^{(t)}, c_1^{(t)}, d^{(t)}, d_1^{(t)}, a_0^{(t)}, a_1^{(t)}, b_1^{(t)})'$. Thus, for each model the vector of parameters is re-estimated every trading day, for $t = \breve{T}, \breve{T}+1, \ldots, \breve{T}+\tilde{T}-1$ days, based on a rolling sample of constant size $\breve{T}$.

5. Realized volatility forecasting

The one-day-ahead adjusted logarithmic realized volatility,$\log(RV_{t+1\mid t}^{(\tau)})$, and the $h_{t+1\mid t}$ for the ARFIMA(1,d,1)-GARCH(1,1) model are computed as

$$\begin{aligned} &\log(\sqrt{252RV_{t+1\mid t}}) = \beta_0^{(t)}(1-c_1^{(t)}) + \\ &+ c_1^{(t)}\log(\sqrt{252RV_t}) + \sum_{j=1}^{\infty}\left(\frac{\Gamma(j+d^{(t)})}{\Gamma(d^{(t)})\Gamma(j+1)}L^{j-1}\right)\varepsilon_{t\mid t} + \sum_{j=0}^{\infty}\left(\frac{\Gamma(j+d^{(t)})}{\Gamma(d^{(t)})\Gamma(j+1)}L^{j}\right)d_1^{(t)}\varepsilon_{t\mid t} \\ &\text{and} \\ &h_{t+1\mid t} = \sqrt{a_0^{(t)} + a_1^{(t)}\varepsilon_{t\mid t}^2 + b_1^{(t)}h_{t\mid t}^2} \end{aligned} \tag{6}$$

The $\log(RV_{t+1\mid t})$ for the ARFIMA(0,d,1)-GARCH(1,1) model is computed from Eq. (6) for $c_1^{(t)}=0$. For the HAR-RV-GARCH(1,1) model we have:

$$\begin{aligned} &\log(\sqrt{252RV_{t+1\mid t}}) = \\ &w_0^{(t)} + w_1^{(t)}\log(\sqrt{252RV_t^{(\tau)}}) + w_2^{(t)}\left(5^{-1}\sum_{j=1}^{5}\log(\sqrt{252RV_{t-j+1}^{(\tau)}})\right) + w_3^{(t)}\left(22^{-1}\sum_{j=1}^{22}\log(\sqrt{252RV_{t-j+1}^{(\tau)}})\right) + \varepsilon_{t\mid t}, \\ &\text{and} \\ &h_{t+1\mid t} = \sqrt{a_0^{(t)} + a_1^{(t)}\varepsilon_{t\mid t}^2 + b_1^{(t)}h_{t\mid t}^2} \end{aligned} \tag{7}$$

The $h_{t+1\mid t}$ for the HAR-RV-GARCH(0,1) model is computed from Eq. (7) for $b_1^{(t)}=0$. In ARFIMA-GARCH and HAR-RV-GARCH frameworks, the dependent variable is conditionally distributed as $\log\sqrt{252RV_t^{(\tau)}}\,|I_{t-1} \sim f(\mu_t, h_t^2;\Theta)$, for I_t denoting the information set available at time t and μ_t referring to the conditional mean given I_t. Therefore, the one-trading-day-ahead annualized realized volatility equals $\sqrt{252RV_{t+1\mid t}} = \exp\left(\log\sqrt{252RV_{t+1\mid t}} + \frac{1}{2}h_{t+1\mid t}^2\right)$.

[8] For more technical details on the GED, readers are referred to Box and Tiao (1973) and Johnson, Kotz, and Balakrishnan (1995).

[9] The skewed Student t distribution has been introduced by Fernandez and Steel (1998). Degiannakis (2004), Giot and Laurent (2003), Lambert and Laurent (2001) estimate model frameworks with skewed Student t distribution.

[10] $\boldsymbol{\theta} = [v, g]'$.

[11] $T = \tilde{T} + \breve{T}$.

Table 5
The *p*-values of the model confidence set for the one-day-ahead volatility forecasts.

Model	For $z_t \sim N(0,1)$	For $z_t \sim t(0,1;v)$	For $z_t \sim Ged(0,1;v)$	For $z_t \sim skT(0,1;v,g)$
CAC 40				
1	0.8300*	0.141*	0.182*	0.2095*
2	0.9426*	0.3472*	0.3472*	0.3750*
3	0.7648*	0.1628	0.0704	0.5000*
4	0.9206*	0.2821*	0.0704	0.5000*
$\min_{m=1,\dots,M}(\varepsilon_{t\|t-1}^{2(m)})$	0.9426*	0.3710*	0.5000*	0.5000*
$\min_{m=1,\dots,M}(z_{t\|t-1}^{2(m)})$	1.0000*	0.3710*	0.5000*	0.5000*
$AV_{t+1\|t}^{(distr)}$	0.9426*	0.1665*	0.1115	0.3710*
$AV_{t+1\|t}$	0.2799*			
DAX30				
1	0.6668*	0.7245*	0.6668*	0.7620*
2	0.5442*	–	1.0000*	0.8249*
3	0.1348*	0.2064*	0.2432*	0.2956*
4	0.3532*	0.2502*	0.2432*	0.4122*
$\min_{m=1,\dots,M}(\varepsilon_{t\|t-1}^{2(m)})$	0.8760*	0.5298*	0.8760*	0.8760*
$\min_{m=1,\dots,M}(z_{t\|t-1}^{2(m)})$	0.6971*	0.5298*	0.8760*	0.6668*
$AV_{t+1\|t}^{(distr)}$	0.1789	0.3540*	0.2482*	0.4122*
$AV_{t+1\|t}$	0.2698*			
FTSE100				
1	0.7213*	0.1985*	0.3507*	0.5901*
2	0.7213*	0.5058*	0.4539*	0.5614*
3	0.5614*	0.2338*	0.1507*	0.4539*
4	0.7746*	0.2822*	0.1507*	0.5614*
$\min_{m=1,\dots,M}(\varepsilon_{t\|t-1}^{2(m)})$	1.0000*	0.5614*	0.1963*	0.5901*
$\min_{m=1,\dots,M}(z_{t\|t-1}^{2(m)})$	0.7746*	0.4539*	0.2672*	0.6583*
$AV_{t+1\|t}^{(distr)}$	0.7746*	0.2656*	0.3538*	0.5614*
$AV_{t+1\|t}$	0.4539*			
EURUSD				
1	0.1413*	0.6615*	0.2659*	0.7095*
2	0.7124*	0.8461*	0.8461*	0.8461*
3	0.4022*	0.2984*	0.3698*	0.5062*
4	0.4022*	0.2722*	0.3698*	0.5062*
$\min_{m=1,\dots,M}(\varepsilon_{t\|t-1}^{2(m)})$	0.8461*	0.8461*	0.9085*	0.8461*
$\min_{m=1,\dots,M}(z_{t\|t-1}^{2(m)})$	0.7815*	0.8461*	1.0000*	0.8461*
$AV_{t+1\backslash t}^{(distr)}$	0.2710*	0.5153*	0.4020*	0.6718*
$AV_{t+1\|t}$	0.4022*			
EURGBP				
1	0.2047*	0.3359*	0.1676*	0.2396*
2	0.2479*	1.0000*	0.2958*	0.0726
3	0.2559*	0.2047*	0.1465*	0.2396*
4	0.2958*	0.2396*	0.1465*	0.2396*
$\min_{m=1,\dots,M}(\varepsilon_{t\|t-1}^{2(m)})$	0.2559*	0.2479*	0.3359*	0.2047*
$\min_{m=1,\dots,M}(z_{t\|t-1}^{2(m)})$	0.2479*	0.1793*	0.2403*	0.0758
$AV_{t+1\|t}^{(distr)}$	0.2479*	0.2958*	0.1793	0.2479*
$AV_{t+1\|t}$	0.2396*			
EURJPY				
1	0.3720*	0.2357*	0.3885*	0.2663*
2	0.6182*	0.5901*	0.5901*	0.5901*
3	0.5901*	0.3558*	0.5724*	0.5901*
4	0.3319*	0.2815*	0.5724*	0.3634*
$\min_{m=1,\dots,M}(\varepsilon_{t\|t-1}^{2(m)})$	0.6182*	0.2940*	1.0000*	0.6182*

(*continued on next page*)

Table 5 (*continued*)

Model	For $z_t \sim N(0,1)$	For $z_t \sim t(0,1;v)$	For $z_t \sim Ged(0,1;v)$	For $z_t \sim skT(0,1;v,g)$
$\min_{m=1,\dots,M}(z_{t\mid t-1}^{2(m)})$	0.4907*	0.2123*	0.4907*	0.4907*
$AV_{t+1\mid t}^{(distr)}$	0.5230*	0.1091	0.5649*	0.3988*
$AV_{t+1\mid t}$	0.2822*			

* Denotes that the model belongs to the confidence set of the best performing models. The interpretation of the MCS p-value is analogous to that of a classical p-value; a $(1-a)$ confidence interval that contains the 'true' parameter with a probability of no less than $(1-a)$.

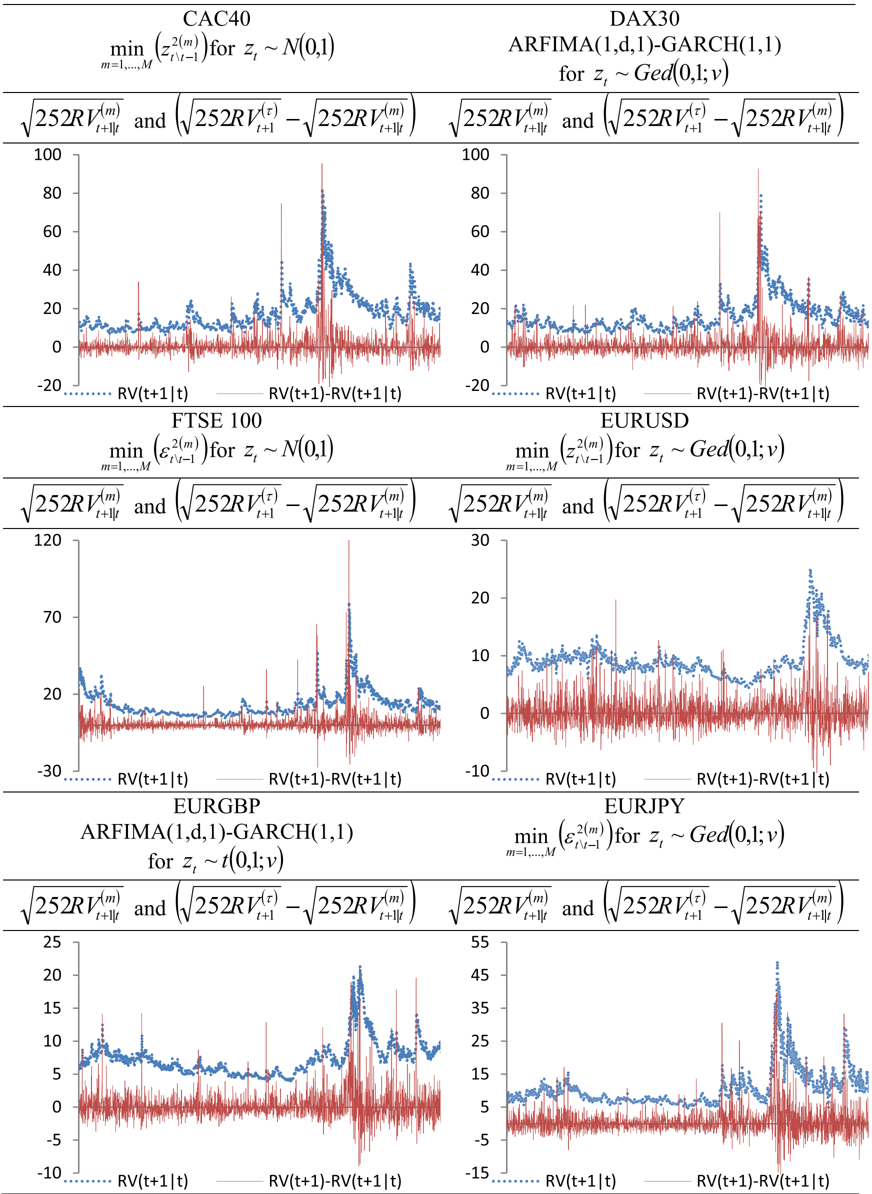

Fig. 2. The $\sqrt{252RV_{t+1\mid t}^{(m)}}$, the discrepancy between $\sqrt{252RV_{t+1}^{(\tau)}}$ and $\sqrt{252RV_{t+1\mid t}^{(m)}}$ for the forecasting methods with the minimum value of $MPSE_{(1)}^{(m)}$.

Table 6
The mean predictive squared error $10^3 MPSE_{(5)}{}^{(m)}$, of the four models for conditionally i) normally; ii) Student t; ii) GED; and iv) skewed Student t distributed innovations. The $MPSE_{(5)}{}^{(distr)}$ i) from combining the forecasts of the four models under the same distribution according to the criteria $\min_{m=1,\dots,M}(\varepsilon_{t|t-1}^{2(m)})$ and $\min_{m=1,\dots,M}(z_{t|t-1}^{2(m)})$; ii) from averaging the forecasts of the four models under the same distribution ($AV_{t+5|t}{}^{(distr)}$); and iii) of the overall averaging forecast ($AV_{t+5|t}$).

Model	For $z_t \sim N(0,1)$	For $z_t \sim t(0,1;v)$	For $z_t \sim Ged(0,1;v)$	For $z_t \sim skT(0,1;v,g)$
CAC 40				
1	54.26	55.81	56.07	55.07
2	51.34	52.16	52.76	51.55
3	39.06	39.14	39.61	38.66
4	39.17	39.49	39.61	39.20
$\min_{m=1,\dots,M}(\varepsilon_{t\|t-1}^{2(m)})$	40.75	40.90	41.40	40.57
$\min_{m=1,\dots,M}(z_{t\|t-1}^{2(m)})$	40.79	40.98	41.45	40.52
$AV_{t+5\|t}{}^{(distr)}$	42.89	43.79	44.24	43.25
$AV_{t+5\|t}$	43.53			
DAX30				
1	59.43	58.84	59.23	58.39
2	57.34	–	56.34	56.27
3	42.95	42.60	42.60	42.41
4	42.73	42.61	42.60	42.41
$\min_{m=1,\dots,M}(\varepsilon_{t\|t-1}^{2(m)})$	44.51	44.50	44.46	44.11
$\min_{m=1,\dots,M}(z_{t\|t-1}^{2(m)})$	44.60	44.55	44.45	44.19
$AV_{t+5\|t}{}^{(distr)}$	48.48	49.53	48.10	47.69
$AV_{t+5\|t}$	48.18			
FTSE100				
1	44.53	44.99	45.16	44.57
2	43.71	44.07	44.40	43.75
3	32.16	32.50	32.85	32.20
4	31.97	32.62	32.85	32.38
$\min_{m=1,\dots,M}(\varepsilon_{t\|t-1}^{2(m)})$	34.34	34.73	35.01	34.49
$\min_{m=1,\dots,M}(z_{t\|t-1}^{2(m)})$	34.34	34.80	34.98	34.42
$AV_{t+5\|t}{}^{(distr)}$	36.48	37.06	37.32	36.75
$AV_{t+5\|t}$	36.90			
EURUSD				
1	8.04	7.94	8.00	7.92
2	7.97	7.95	7.95	7.93
3	6.10	6.12	6.11	6.10
4	6.10	6.12	6.11	6.10
$\min_{m=1,\dots,M}(\varepsilon_{t\|t-1}^{2(m)})$	6.91	6.91	6.90	6.91
$\min_{m=1,\dots,M}(z_{t\|t-1}^{2(m)})$	6.92	6.91	6.90	6.91
$AV_{t+5\|t}{}^{(distr)}$	6.86	6.84	6.85	6.83
$AV_{t+5\|t}$	6.84			
EURGBP				
1	5.18	4.96	5.15	5.11
2	5.57	4.97	5.19	5.00
3	4.11	4.14	4.16	4.14
4	4.12	4.16	4.16	4.17
$\min_{m=1,\dots,M}(\varepsilon_{t\|t-1}^{2(m)})$	4.63	4.57	4.63	4.57
$\min_{m=1,\dots,M}(z_{t\|t-1}^{2(m)})$	4.62	4.59	4.64	4.63
$AV_{t+5\|t}{}^{(distr)}$	4.55	4.44	4.51	4.48
$AV_{t+5\|t}$	4.48			
EURJPY				
1	20.90	20.96	20.88	20.70
2	18.85	18.73	18.70	18.57
3	14.04	14.09	13.98	13.90
4	14.47	14.30	13.98	14.19
$\min_{m=1,\dots,M}(\varepsilon_{t\|t-1}^{2(m)})$	14.66	14.87	14.78	14.64

(*continued on next page*)

Table 6 (*continued*)

Model	For $z_t \sim N(0,1)$	For $z_t \sim t(0,1;v)$	For $z_t \sim Ged(0,1;v)$	For $z_t \sim skT(0,1;v,g)$
$\min_{m=1,\dots,M}(z_{t\vert t-1}^{2(m)})$	14.66	14.87	14.77	14.61
$AV_{t+5\vert t}^{(distr)}$	16.11	16.08	15.93	15.89
$AV_{t+5\vert t}$	16.00			

Model 1: ARFIMA(0,d,1)-GARCH(1,1), Model 2: ARFIMA(1,d,1)-GARCH(1,1), Model 3: HAR-RV-GARCH(1,1), Model 4: HAR-RV-GARCH(0,1).

The formulas for multiple-days-ahead realized volatility forecasts ($n \geq 2$) are constructed recursively based on Degiannakis, Floros, and Dent (2014). For example, $\log(RV_{t+n|t})$, and $h_{t+n|t}$ are computed as:

ARFIMA(1,d,1)-GARCH(1,1) model:

$$\log(\sqrt{252RV_{t+n|t}}) = \beta_0^{(t)}(1 - c_1^{(t)}) + c_1^{(t)}\log(\sqrt{252RV_{t+n|t}})$$
$$h_{t+n|t} = \sqrt{a_0^{(t)} + a_1^{(t)}\varepsilon_{t+n-1|t}^2 + b_1^{(t)}h_{t+n-1|t}^2} \tag{8}$$

HAR-RV-GARCH(1,1) model:

$$\log(\sqrt{252RV_{t+n|t}})=$$
$$w_0^{(t)} + w_1^{(t)}\log(\sqrt{252RV_{t+n-1|t}}) + w_2^{(t)}\left(5^{-1}\sum_{j=1}^{5}\log(\sqrt{252RV_{t+n-j}^{(\tau)}})\right) + w_3^{(t)}\left(22^{-1}\sum_{j=1}^{22}\log(\sqrt{252RV_{t+n-j}^{(\tau)}})\right),$$

and

$$h_{t+n|t} = \sqrt{a_0^{(t)} + a_1^{(t)}\varepsilon_{t+n-1|t}^2 + b_1^{(t)}h_{t+n-1|t}^2} \tag{9}$$

6. Combining forecasts

In this section, we will investigate whether the combination of predictions can provide more accurate volatility forecasts compared to the use of a specific single model. Let us define that we have a set of Mcompeting models. At each point in time we forecast the next day's volatility based on the model with the minimum forecast error. Specifically, we investigate two rules (evaluation functions) for model selection based on the most recent one-step-ahead forecast error, or $\varepsilon_{t|t-1} = \log(\sqrt{252RV_t^{(\tau)}}) - \log(\sqrt{252RV_{t|t-1}})$, and the most recent one-step-ahead standardized forecast error, or $z_{t|t-1} = \frac{\varepsilon_{t|t-1}}{h_{t|t-1}}$. In other words, on day $t-1$ we estimate the M competing models, and for day t we compute the one-step-ahead forecasts. For day $t+1$ we forecast the volatility based on the model m with:

$$\min_{m=1,\dots,M}(\varepsilon_{t|t-1}^{2(m)}) \tag{10}$$

or

$$\min_{m=1,\dots,M}(z_{t|t-1}^{2(m)}). \tag{11}$$

The predicted squared forecast error, in Eq. (10), is the most widely accepted criterion for evaluating forecasting ability. The Eq. (11) is the standardized predicted squared forecast error, whose properties have been investigated by Degiannakis and Xekalaki (2005).

Consider a model with the generic form, which incorporates the models in Eqs. (1) and (2):

$$y_t = \mathbf{x}_{t-1}'\boldsymbol{\beta} + \varepsilon_t$$
$$\varepsilon_t = z_t\sigma_t$$
$$\sigma_t^2(\boldsymbol{\eta}) = g(\varepsilon_{t-j}, \sigma_{t-j}), \tag{12}$$

where $\boldsymbol{\eta}$[12] is a vector of parameters to be estimated, $z_t \overset{i.i.d.}{\sim} N(0,1)$, $\sigma_t^2(.)$ represents the conditional variance of ε_t, and $g(.)$ is the functional form of the conditional variance. Under the assumption of constancy of parameters over time, $\boldsymbol{\eta}_1 = \boldsymbol{\eta}_2 = \dots = \boldsymbol{\eta}_T = \boldsymbol{\eta}$, the $z_{t|t-1}$ has an asymptotic standard normal distribution, where $z_{t|t-1} \equiv (y_t - y_{t|t-1})\sigma_{t|t-1}^{-1}$, $y_{t|t-1} = \mathbf{x}_{t-1}'\boldsymbol{\beta}^{(t-1)}$ and $\sigma_{t|t-1}$ is the one-step-ahead conditional standard deviation.

If we have $m = 1, 2, \dots, M$ competing models, we may compute the $z_{t|t-1}^{(m)}$. Krishnamoorthy and Parthasarathy (1951) showed that if M variables jointly follow the standard normal distribution, then the joint distribution of $(z_{t|t-1}^{(1)}, z_{t|t-1}^{(2)}, \dots, z_{t|t-1}^{(M)})$ is the Multivariate Gamma. Then the distribution function of $z_{(1)} \equiv \min_{m=1,\dots,M}(z_{t|t-1}^{(1)}, z_{t|t-1}^{(2)}, \dots, z_{t|t-1}^{(M)})$ can be used to compare the predictability of

[12] $\boldsymbol{\beta}$ belongs to $\boldsymbol{\eta}$.

Table 7
The p-values of the model confidence set for the one-week-ahead volatility forecasts.

Model	For $z_t \sim N(0,1)$	For $z_t \sim t(0,1;v)$	For $z_t \sim Ged(0,1;v)$	For $z_t \sim skT(0,1;v,g)$
CAC 40				
1	0.0042	0.0034	0.0034	0.0036
2	0.0088	0.007	0.0065	0.008
3	0.0913	0.0201	0.0142	1.0000[a]
4	0.054	0.0002	0.0142	0.0013
$\min_{m=1,\dots,M}(\varepsilon_{t\mid t-1}^{2(m)})$	0.2779[a]	0.2529[a]	0.0142	0.2779[a]
$\min_{m=1,\dots,M}(z_{t\mid t-1}^{2(m)})$	0.2779[a]	0.1444[a]	0.0116	0.2779[a]
$AV_{t+5\mid t}^{(distr)}$	0.2529[a]	0.0106	0.013	0.1229[a]
$AV_{t+5\mid t}$	0.0156			
DAX30				
1	0.0068	0.0073	0.0068	0.0094
2	0.0072	–	0.0073	0.0073
3	0.1034[a]	0.1263[a]	0.0295	1.0000[a]
4	0.0163	0.0004	0.0295	0.9648[a]
$\min_{m=1,\dots,M}(\varepsilon_{t\mid t-1}^{2(m)})$	0.1462[a]	0.1462[a]	0.1462[a]	0.1462[a]
$\min_{m=1,\dots,M}(z_{t\mid t-1}^{2(m)})$	0.1363[a]	0.1462[a]	0.1462[a]	0.1462[a]
$AV_{t+5\mid t}^{(distr)}$	0.029	0.1363[a]	0.1263[a]	0.1462[a]
$AV_{t+5\mid t}$	0.0586			
FTSE100				
1	0.005	0.005	0.005	0.005
2	0.005	0.005	0.005	0.005
3	0.4261[a]	0.0033	0.0041	0.4261[a]
4	1.0000[a]	0.0006	0.0041	0.005
$\min_{m=1,\dots,M}(\varepsilon_{t\mid t-1}^{2(m)})$	0.005	0.0031	0.0031	0.005
$\min_{m=1,\dots,M}(z_{t\mid t-1}^{2(m)})$	0.005	0.0031	0.0031	0.005
$AV_{t+5\mid t}^{(distr)}$	0.0284	0.005	0.005	0.0072
$AV_{t+5\mid t}$	0.005			
EURUSD				
1	0.0000	0.0000	0.0000	0.0000
2	0.0000	0.0000	0.0000	0.0000
3	1.0000[a]	0.0000	0.0233	0.8058[a]
4	0.0315	0.0000	0.0233	0.5904[a]
$\min_{m=1,\dots,M}(\varepsilon_{t\mid t-1}^{2(m)})$	0.0000	0.0000	0.0000	0.0000
$\min_{m=1,\dots,M}(z_{t\mid t-1}^{2(m)})$	0.0000	0.0000	0.0000	0.0000
$AV_{t+5\mid t}^{(distr)}$	0.0000	0.0000	0.0000	0.0000
$AV_{t+5\mid t}$	0.0000			
EURGBP				
1	0.0002	0.0001	0.0001	0.0002
2	0.0003	0.0003	0.0003	0.0003
3	1.0000[a]	0.0139	0.0028	0.0228
4	0.1635	0.0041	0.0028	0.0003
$\min_{m=1,\dots,M}(\varepsilon_{t\mid t-1}^{2(m)})$	0.0139	0.0139	0.0139	0.0139
$\min_{m=1,\dots,M}(z_{t\mid t-1}^{2(m)})$	0.0139	0.0139	0.0139	0.0139
$AV_{t+5\mid t}^{(distr)}$	0.0108	0.0139	0.0125	0.0139
$AV_{t+5\mid t}$		0.0128		
EURJPY				
1	0.0017	0.0017	0.0017	0.0017
2	0.0017	0.0018	0.002	0.0038
3	0.0392	0.0326	0.2601[a]	1.0000[a]
4	0	0	0.2601[a]	0.0392
$\min_{m=1,\dots,M}(\varepsilon_{t\mid t-1}^{2(m)})$	0.2601[a]	0.0085	0.2601[a]	0.2601[a]
$\min_{m=1,\dots,M}(z_{t\mid t-1}^{2(m)})$	0.2601[a]	0.0095	0.2601[a]	0.2601[a]
$AV_{t+5\mid t}^{(distr)}$	0.0104	0.0039	0.095	0.0982
$AV_{t+5\mid t}$	0.0326			

[a] Denotes that the model belongs to the confidence set of the best performing models.

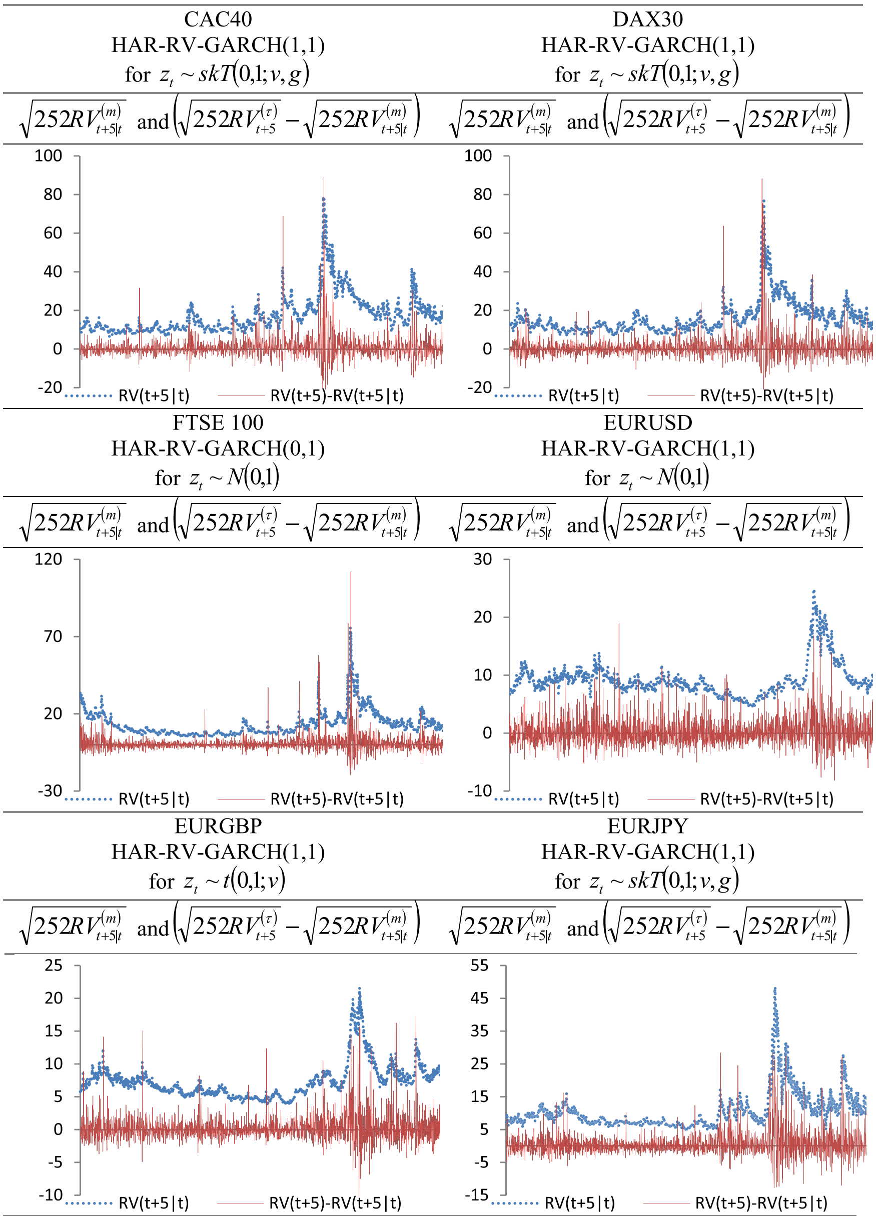

Fig. 3. The $\sqrt{252RV_{t+5|t}^{(m)}}$, the discrepancy between $\sqrt{252RV_{t+5}^{(\tau)}}$ and $\sqrt{252RV_{t+5|t}^{(m)}}$ for the forecasting methods with the minimum value of $MPSE_{(5)}{}^{(m)}$.

the M models. The cumulative distribution function of $z_{(1)}$ is the minimum multivariate gamma (MMG) distribution (see Xekalaki & Degiannakis, 2005, 2010). The single models are based on the statistical assumption that the standardized residuals are i) normally, ii) Student t, iii) GED or iv) skewed Student t distributed. On the other hand, the combined forecasts according to the $\min\limits_{m=1,\ldots,M}(z_{t|t-1}^{2(m)})$ criterion for normally distributed standardized residuals have a known and explicitly derived distribution form; the minimum multivariate gamma. Based on simulated evidence, Degiannakis and Livada (2016) expanded the research on the non-normally distributed standardized residuals. However, the combined forecasts according to the $\min\limits_{m=1,\ldots,M}(\varepsilon_{t|t-1}^{2(m)})$ criterion do not have a known distribution function, despite the fact that almost all the forecasting evaluations conducted in the financial literature, are based on the non-standardized residuals, $\varepsilon_{t|t-1}$. Hence, the combined forecasts according to the $\min\limits_{m=1,\ldots,M}(z_{t|t-1}^{2(m)})$ criterion are compatible with the assumptions behind the each of the models that comprise it, whereas this is not the case for the $\min\limits_{m=1,\ldots,M}(\varepsilon_{t|t-1}^{2(m)})$ criterion.

7. Averaging forecasts

Next, we proceed with model-average forecasts in order to assess whether the average forecast could improve forecasting

Table 8

The mean predictive squared error $10^3 MPSE_{(10)}{}^{(m)}$, of the four models for conditionally i) normally; ii) Student t; ii) GED; and iv) skewed Student t distributed innovations. The $MPSE_{(10)}{}^{(distr)}$ i) from combining the forecasts of the four models under the same distribution according to the criteria $\min_{m=1,\dots,M}(\varepsilon_{t|t-1}^{2(m)})$ and $\min_{m=1,\dots,M}(z_{t|t-1}^{2(m)})$; ii) from averaging the forecasts of the four models under the same distribution ($AV_{t+10|t}{}^{(distr)}$); and iii) of the overall averaging forecast ($AV_{t+10|t}$).

Model	For $z_t \sim N(0,1)$	For $z_t \sim t(0,1;v)$	For $z_t \sim Ged(0,1;v)$	For $z_t \sim skT(0,1;v,g)$
CAC 40				
1	62.42	64.46	64.83	63.37
2	57.24	57.90	58.61	56.98
3	56.57	56.04	56.03	55.55
4	56.36	55.78	56.03	55.21
$\min_{m=1,\dots,M}(\varepsilon_{t|t-1}^{2(m)})$	50.81	51.18	51.51	50.51
$\min_{m=1,\dots,M}(z_{t|t-1}^{2(m)})$	51.05	51.44	51.68	50.75
$AV_{t+10|t}{}^{(distr)}$	52.19	52.85	53.37	52.07
$AV_{t+10|t}$	52.61			
DAX30				
1	66.76	65.87	66.50	65.03
2	63.34	–	61.64	61.73
3	57.81	57.34	57.22	57.14
4	57.56	57.27	57.22	57.06
$\min_{m=1,\dots,M}(\varepsilon_{t|t-1}^{2(m)})$	52.75	53.43	52.10	52.26
$\min_{m=1,\dots,M}(z_{t|t-1}^{2(m)})$	52.76	53.45	52.13	52.27
$AV_{t+10|t}{}^{(distr)}$	57.32	57.43	56.61	56.04
$AV_{t+10|t}$	56.82			
FTSE100				
1	49.51	50.06	50.32	49.48
2	48.21	48.58	49.26	48.10
3	46.77	46.33	46.82	46.15
4	46.54	46.19	46.82	45.97
$\min_{m=1,\dots,M}(\varepsilon_{t|t-1}^{2(m)})$	42.66	43.06	44.06	42.65
$\min_{m=1,\dots,M}(z_{t|t-1}^{2(m)})$	42.79	43.13	44.07	42.83
$AV_{t+10|t}{}^{(distr)}$	44.08	44.35	44.77	44.00
$AV_{t+10|t}$	44.29			
EURUSD				
1	8.15	7.98	8.11	7.95
2	8.04	7.99	8.01	7.95
3	7.82	7.78	7.79	7.78
4	7.82	7.78	7.79	7.78
$\min_{m=1,\dots,M}(\varepsilon_{t|t-1}^{2(m)})$	7.75	7.70	7.72	7.72
$\min_{m=1,\dots,M}(z_{t|t-1}^{2(m)})$	7.75	7.70	7.72	7.70
$AV_{t+10|t}{}^{(distr)}$	7.62	7.57	7.600	7.55
$AV_{t+10|t}$	7.58			
EURGBP				
1	5.38	5.04	5.35	5.27
2	5.86	5.00	5.34	5.05
3	4.82	4.81	4.82	4.80
4	4.80	4.80	4.82	4.79
$\min_{m=1,\dots,M}(\varepsilon_{t|t-1}^{2(m)})$	4.88	4.83	4.89	4.90
$\min_{m=1,\dots,M}(z_{t|t-1}^{2(m)})$	4.88	4.83	4.90	4.90
$AV_{t+10|t}{}^{(distr)}$	4.90	4.70	4.81	4.75
$AV_{t+10|t}$	4.77			
EURJPY				
1	23.90	23.97	23.89	23.60
2	20.10	19.80	19.87	19.55
3	19.60	19.38	19.51	19.32
4	19.70	19.53	19.51	19.44
$\min_{m=1,\dots,M}(\varepsilon_{t|t-1}^{2(m)})$	17.67	17.74	17.90	17.49

(*continued on next page*)

Table 8 (*continued*)

Model	For $z_t \sim N(0,1)$	For $z_t \sim t(0,1;v)$	For $z_t \sim Ged(0,1;v)$	For $z_t \sim skT(0,1;v,g)$
$\min_{m=1,\dots,M}(z_{t\|t-1}^{2(m)})$	17.72	17.79	17.96	17.32
$AV_{t+10\|t}^{(distr)}$	19.05	18.94	18.90	18.71
$AV_{t+10\|t}$	18.89			

Model 1: ARFIMA(0,d,1)-GARCH(1,1), Model 2: ARFIMA(1,d,1)-GARCH(1,1), Model 3: HAR-RV-GARCH(1,1), Model 4: HAR-RV-GARCH(0,1).

accuracy. We consider the model-average forecasts of all the models with the same distributional assumption:

$$AV_{t+n|t}^{(distr)} = M^{-1}\sum_{m=1}^{M}\sqrt{252RV_{t+n|t}^{(m)}}, \tag{13}$$

where $M = 4$ and $distr = N, t, Ged, skT$ denotes the conditional distribution of the models.

In addition, we construct the overall average forecast of all the competing models and residual distributions.

$$AV_{t+n|t} = M^{-1}\sum_{m=1}^{M}\sqrt{252RV_{t+n|t}^{(m)}}, \tag{14}$$

where $M = 16$.

8. Evaluating model predictability

The 16 models are re-estimated every trading day t, for $\widetilde{T}$ days, where $\widetilde{T}$=1686, 1784, 2106, 2308, 2091, 2108 for the CAC40, DAX30, FTSE100, EURUSD, EURGBP and EURJPY realized volatility series. The rolling window approach with a fixed window length of $\breve{T} = 1000$ days is utilized for incorporating changes in trading behavior more efficiently. The total number of observations is $T = \widetilde{T} + \breve{T}$. The forecasting accuracy of the models is measured with the mean predictive squared error (MPSE)[13]:

$$MPSE_{(n)}^{(m)} = \widetilde{T}^{-1}\sum_{t=1}^{\widetilde{T}}\left(\sqrt{252RV_{t+n|t}^{(m)}} - \sqrt{252RV_{t+n}^{(\tau)}}\right)^2. \tag{15}$$

The superscript (m) denotes the model $m=1, 2, \dots, M$ and the subscript (n) denotes the n-days-ahead forecast for $n=1, 5, 10$. Patton (2011) argues that the use of proxies for true volatility induces distortions in the model ranking for certain loss functions. He proposes that mean squared error is a loss function which is robust to noisy volatility proxies and will lead to an unbiased model ordering. Therefore, we report the results under the MPSE loss function.

Beyond the 16 models, we have defined 2 methods of combining forecasts (in Section 6). Each method is applied to the models with i) normally; ii) Student t; iii) GED; and iv) skewed Student t distributed innovations. Hence, 8 techniques of combining forecasts are investigated in total.

Additionally, in Section 7, we have constructed 5 model-average forecasts. These are the model-average forecasts of the 4 models with the same distributional assumption, as well as the overall average forecast of all 16 competing models.

Among the statistical methods which evaluate the predictions from a variety of models, the most widely accepted are: The Diebold and Mariano (1995) test for pairwise comparisons, the Equal Predictive Accuracy test of Clark and West (2007) for nested models, and the Reality Check for Data Snooping (White, 2000) and the Superior Predictive Ability test (Hansen, 2005) for multiple comparisons against a benchmark model. Recently, Hansen, Lunde, and Nason (2011) introduced the Model Confidence Set (MCS) test, which evaluates a number of forecasting models simultaneously, not against a benchmark model. The MCS method does not assume the existence of any predefined true data generating process. Its major advantage is the comparison of forecasts, not necessarily estimated by models, which acknowledges the limitations of the data. Thus, uninformative data yield a confidence set with many models whereas informative data yield a set of just a few models. The MCS is employed in order to determine the set of models that is made up of the best ones. The term "best" is defined according to our evaluation function MPSE. The MCS compares the prediction accuracy of an initial set of M^0 models and investigates, at a predefined level of significance, which models survive the elimination algorithm. For $L_t^{(m)}$ denoting the evaluation functions of model m on day t, and $d_t^{(m,m')} = L_t^{(m)} - L_t^{(m')}$ being the evaluation differential for $m, m' \in M^0$, the hypothesis that is being tested is:

$$H_{0,M}: E(d_t^{(m,m')}) = 0, \tag{16}$$

for $\vee m, m' \in M$, $M \subset M^0$ against the alternative $H_{1,M}$:$E(d_t^{(m,m')}) \neq 0$ for some $m,m' \in M$. For example, in the case of the MPSE evaluation function, $L_t^{(m)} = \left(\sqrt{252RV_{t+n|t}^{(m)}} - \sqrt{252RV_{t+n}^{(\tau)}}\right)^2$. The elimination algorithm based on an equivalence test and an elimination rule

[13] The mean predictive absolute error, $MPAE_{(n)}^{(m)} = \widetilde{T}^{-1}\sum_{t=1}^{\widetilde{T}}\left|\sqrt{252RV_{t+n|t}^{(m)}} - \sqrt{252RV_{t+n}^{(\tau)}}\right|$, is also applied. Results are qualitatively similar and available upon request.

Table 9
The *p*-values of the model confidence set for the two-weeks-ahead volatility forecasts.

Model	For $z_t \sim N(0,1)$	For $z_t \sim t(0,1;v)$	For $z_t \sim Ged(0,1;v)$	For $z_t \sim skT(0,1;v,g)$
CAC 40				
1	0.0032	0.0019	0.0013	0.0024
2	0.1185[a]	0.038	0.0135	0.1185[a]
3	0.0000	0.0265	0.0979	0.1353[a]
4	0.0000	0.0039	0.0979	0.1873[a]
$\min_{m=1,\ldots,M}(\varepsilon_{t\|t-1}^{2(m)})$	0.4061[a]	0.0453	0.0514	1.0000[a]
$\min_{m=1,\ldots,M}(z_{t\|t-1}^{2(m)})$	0.0514	0.0244	0.0514	0.2924[a]
$AV_{t+10\|t}^{(distr)}$	0.4061[a]	0.0098	0.0081	0.5175[a]
$AV_{t+10\|t}$	0.008			
DAX30				
1	0.0134	0.0373	0.0208	0.0459
2	0.0346	–	0.0953	0.0953
3	0.0023	0.0056	0.0066	0.0082
4	0.0039	0.0064	0.0066	0.0106
$\min_{m=1,\ldots,M}(\varepsilon_{t\|t-1}^{2(m)})$	0.1087[a]	0.0457	1.0000[a]	0.8506[a]
$\min_{m=1,\ldots,M}(z_{t\|t-1}^{2(m)})$	0.1087[a]	0.0457	0.3532[a]	0.8506[a]
$AV_{t+10\|t}^{(distr)}$	0.0128	0.0457	0.0889	0.3194[a]
$AV_{t+10\|t}$	0.0235			
FTSE100				
1	0.0068	0.0039	0.0057	0.0162
2	0.0475	0.0379	0.0244	0.0475
3	0.0001	0.0231	0.0088	0.0523
4	0.0477	0.0053	0.0088	0.0666
$\min_{m=1,\ldots,M}(\varepsilon_{t\|t-1}^{2(m)})$	0.9371[a]	0.0477	0.1193[a]	1.0000[a]
$\min_{m=1,\ldots,M}(z_{t\|t-1}^{2(m)})$	0.5783[a]	0.016	0.1193[a]	0.7157[a]
$AV_{t+10\backslash t}^{(distr)}$	0.2766[a]	0.0142	0.0084	0.7157[a]
$AV_{t+10\|t}$	0.0065			
EURUSD				
1	0.0022	0.0115	0.0026	0.0196
2	0.0058	0.0106	0.0078	0.0189
3	0.0022	0.0254	0	0.092
4	0.0029	0.4305[a]	0	0.4305[a]
$\min_{m=1,\ldots,M}(\varepsilon_{t\|t-1}^{2(m)})$	0.4514[a]	0.4937[a]	0.4937[a]	0.4937[a]
$\min_{m=1,\ldots,M}(z_{t\|t-1}^{2(m)})$	0.4491[a]	0.4937[a]	0.4937[a]	0.4937[a]
$AV_{t+10\|t}^{(distr)}$	0.0655	0.3283[a]	0.0799	1.0000[a]
$AV_{t+10\|t}$	0.092			
EURGBP				
1	0.0005	0.0005	0.0008	0.0015
2	0.0011	0.0041	0.0065	0.0015
3	0.0173	0.2862[a]	0.1594[a]	0.4597[a]
4	0.4597[a]	0.4597[a]	0.1594[a]	0.4597[a]
$\min_{m=1,\ldots,M}(\varepsilon_{t\|t-1}^{2(m)})$	0.2862[a]	0.2862[a]	0.2862[a]	0.2374[a]
$\min_{m=1,\ldots,M}(z_{t\|t-1}^{2(m)})$	0.2862[a]	0.2862[a]	0.2862[a]	0.2374[a]
$AV_{t+10\|t}^{(distr)}$	0.2082[a]	1.0000[a]	0.2862[a]	0.4597[a]
$AV_{t+10\|t}$	0.2862[a]			
EURJPY				
1	0.0034	0.0037	0.0037	0.0042
2	0.1247[a]	0.1911[a]	0.1911	0.2605[a]
3	0.0016	0.0243	0.0219	0.0383
4	0.0037	0.003	0.0219	0.0243
$\min_{m=1,\ldots,M}(\varepsilon_{t\|t-1}^{2(m)})$	0.3199[a]	0.0546	0.1911[a]	0.3199[a]
$\min_{m=1,\ldots,M}(z_{t\|t-1}^{2(m)})$	0.3199[a]	0.2640[a]	0.1352[a]	1.0000[a]
$AV_{t+10\|t}^{(distr)}$	0.034	0.0066	0.0357	0.3199[a]
$AV_{t+10\|t}$	0.0183			

[a] Denotes that the model belongs to the confidence set of the best performing models.

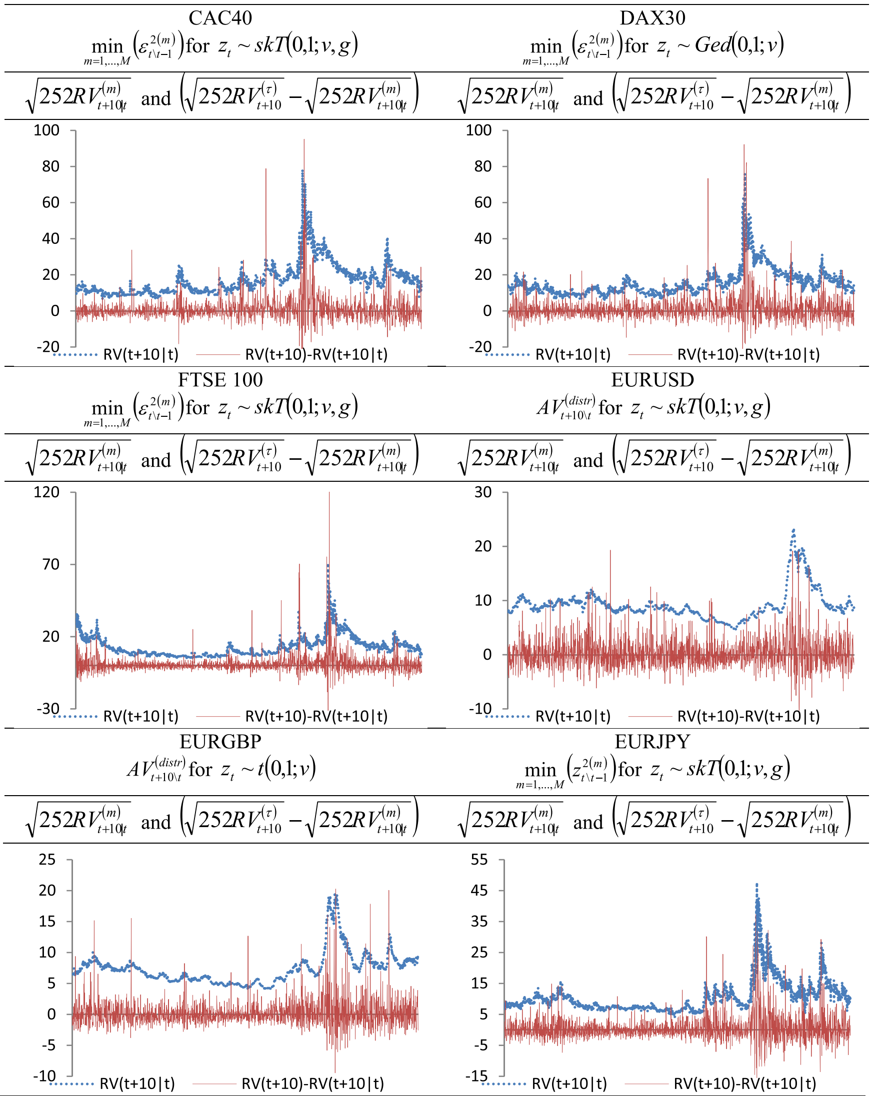

Fig. 4. The $\sqrt{252RV^{(m)}_{t+10|t}}$, the discrepancy between $\sqrt{252RV^{(\tau)}_{t+10}}$ and $\sqrt{252RV^{(m)}_{t+10|t}}$ for the forecasting methods with the minimum value of $MPSE_{(10)}{}^{(m)}$.

employs the former to investigate the $H_{0,M}$ for $\forall M \subset M^0$ and the latter to identify the model m to be removed from M in case $H_{0,\ M}$ is rejected.

9. Investigating predictive accuracy

The main purpose of our study is to explore the possible sources that help us enhance our realized volatility forecasts. Let us keep in mind that we have investigated the predictive accuracy of model frameworks with different autoregressive structures (i.e. long-memory autoregressive against heterogeneous autoregressive), and different distributions for the standardized residuals (i.e. normal against skewed Student *t*). Then, we explore whether the use of a single model can be improved upon by the implementation of a method that combines forecasts (Section 6) or by the averaging of forecasts (Section 7). To sum up, the hypotheses that we investigate are: 1) The heterogeneous autoregressive (HAR) framework is expected to work better than the long memory (ARFIMA) framework. 2) The averaged models are expected to provide superior forecasts compared to those of the single models; either HAR or ARFIMA. 3) Is the modeling of volatility asymmetry crucial in forecasting realized volatility? 4) Does the combination of volatility forecasts according to the statistical properties of forecast errors provide more accurate volatility forecasts? The forecasting evaluation exercise is not limited to the one-day-ahead forecasts, as we also explore the predictive ability for the 5-days and10-days-ahead horizons.

9.1. One-day-ahead predictive accuracy

Table 4 provides the values of the mean predictive squared error, $10^3 MPSE_{(1)}{}^{(m)}$. For each one of the stock indices and the exchange rates, the first four rows provide the $10^3 MPSE_{(1)}{}^{(m)}$ statistics for the four models and the four distributional assumptions. The fifth (sixth) row presents the $MPSE_{(1)}{}^{(distr)}$ statistics from combining the forecasts of the four models under the same distribution according to the criterion $\min_{m=1,\dots,M}(\varepsilon_{t|t-1}^{2(m)})(\min_{m=1,\dots,M}(z_{t|t-1}^{2(m)}))$. The seventh row presents the $AV_{t+1|t}{}^{(distr)}$ statistics resulting from averaging the forecasts of the four models under the same distribution, whereas the last row provides the overall averaging forecast,$AV_{t+1|t}$. The relative *p*-values of the MCS test are presented in Table 5. For each of the realized volatility series, Fig. 2 plots the $\sqrt{252RV_{t+1|t}^{(m)}}$ and the discrepancy between $\sqrt{252RV_{t+1}^{(\tau)}}$ and $\sqrt{252RV_{t+1|t}^{(m)}}$ for the forecast methods with the minimum value of $MPSE_{(1)}{}^{(m)}$. Overall, we cannot infer in favor of a specific method of constructing one-day-ahead realized volatility forecasts. The p-values in Table 5 conclude that most of the prediction methods (single models, combined forecasts and averaged models) belong to the confidence set of the best performing models. The lowest value of the $MPSE_{(1)}{}^{(m)}$ statistic is achieved by a single model, the ARFIMA(1,d,1)-GARCH(1,1), in the case of the DAX30 and the Euro/Pound rate, whereas one of the combined methods has the lowest $MPSE_{(1)}{}^{(m)}$ value for the other four indices.

9.2. Five-days-ahead predictive accuracy

Table 6 shows the mean predictive squared forecast error, or $10^3 MPSE_{(5)}{}^{(m)}$ for the five-trading-days-ahead realized volatility forecasts. According to Table 7, which presents the MCS p-values, the picture is clearer in the case of one-calendar-week-ahead forecasting. A limited number of prediction methods belong to the confidence set of the best performing models. Specifically, for the DAX30 index, just one model, the HAR-RV-GARCH(1,1)-skT, belongs to the set of confidence models (for a 20% level of significance). A similar case holds for the Euro/Pound rate, with the same model under the normal distribution (HAR-RV-GARCH(1,1)-n) constructing the most accurate volatility forecasts. For the FTSE100 stock index and the Euro/Dollar exchange rate, the MCS is comprised of three models, all of which have a heterogeneous autoregressive form. In general, for five-trading-days volatility forecasting, the heterogeneous autoregressive model is superior to the long memory framework. Additionally, the combined forecasts and the averaged models fail to provide superior volatility forecasts for the one-calendar-week-ahead forecasting horizon (only for the CAC40 index and the Euro/Yen rate, the $\min_{m=1,\dots,M}(\varepsilon_{t|t-1}^{2(m)})$ or $\min_{m=1,\dots,M}(z_{t|t-1}^{2(m)})$ methods of combined forecasts belong to the MCS). Fig. 3 plots the $\sqrt{252RV_{t+5|t}^{(m)}}$, and the forecast error $(\sqrt{252RV_{t+5}^{(\tau)}} - \sqrt{252RV_{t+5|t}^{(m)}})$ for the predictive methods with the minimum $MPSE_{(5)}{}^{(m)}$.

9.3. Ten-days-ahead predictive accuracy

Tables 8 and 9 illustrate the relative information for the two-calendar-weeks-ahead forecasts. Overall, in the ten-days-ahead forecasting horizon, the necessity for employing combined forecasts and averaged models arises. According to Table 9, for all the realized volatility series under investigation the combined forecasts according to the $\min_{m=1,\dots,M}(\varepsilon_{t|t-1}^{2(m)})$ and $\min_{m=1,\dots,M}(z_{t|t-1}^{2(m)})$ criteria belong to the confidence set of the best performing methods of forecasting. Moreover, the estimation of the models with skewed Student *t* distributed standardized residuals is crucial in providing superior realized volatility forecasts. The financial literature has delivered strong evidence in favor of modeling the asymmetric and leptokurtic character of the log-returns distribution (see for example Degiannakis et al., 2014). For multiple-steps-ahead forecasting, the asymmetric character of realized volatility must be considered as well. From the different behavior of the leptokurtic distributions (Student *t* and GED) and the asymmetric and leptokurtic one (the skewed Student *t*), we observe that the modeling of volatility asymmetry is important for longer forecasting horizons. For each of the six realized volatility series, Fig. 4 plots the $\sqrt{252RV_{t+10|t}^{(m)}}$, and the discrepancy between $\sqrt{252RV_{t+10}^{(\tau)}}$ and $\sqrt{252RV_{t+10|t}^{(m)}}$ for the forecast methods with the minimum value of $MPSE_{(10)}{}^{(m)}$.

For purposes of robustness, we have investigated the forecasting performance based on the mean predictive absolute error. The results are qualitatively similar. Thus, we do not report the Tables with the values of the evaluation function and the relevant MCS *p*-values, which are available to the readers upon request.

10. Conclusion

Our major task is to investigate whether we can enhance our realized volatility forecasts. The forecasting evaluation is conducted for one-day-ahead, one-calendar-week-ahead and two-calendar-weeks-ahead horizons. The ARFIMA-GARCH and HAR-RV-GARCH models are estimated for the major European Union stock market indices (FTSE100, DAX30, CAC40) and for the exchange rates of the Euro with the British Pound, the US Dollar and the Japanese Yen under the assumption that the standardized innovations are i) normally; ii) Student *t*; ii) GED; and iv) skewed Student *t* distributed. Additionally, we explore whether the use of a single model can be improved upon through the implementation of a method that combines forecasts or by the averaging of the forecasts.

The overall findings can be summarized as follows. For one-day-ahead volatility forecasts, most prediction methods (single models, combined forecasts and averaged models) belong to the confidence set of the best performing models. For five-trading-days-ahead forecasting horizon, the heterogeneous autoregressive model is superior to the long-memory framework model. Moreover, the combined forecasts and the averaged models fail to provide superior volatility forecasts. For the ten-trading-days-ahead forecasting horizon, the $\min_{m=1,\dots,M}(\varepsilon_{t|t-1}^{2(m)})$ and $\min_{m=1,\dots,M}(z_{t|t-1}^{2(m)})$ criteria deliver the most accurate volatility forecasts. Also, the averaged models provide superior forecasts compared to those of single models. Additionally, the modeling of volatility asymmetry (the use of the skewed Student *t* distribution) is important for the ten-days-ahead volatility forecasts.

Thus, for longer forecasting horizons, more complicated forecasting frameworks are required. Combined forecasts and averaged models are methods considered to be adequate for volatility forecasting purposes; a crucial finding for investors, portfolio managers, risk managers, policy makers, etc.

Avenues for future research may include the enrichment of the methods under comparison (i.e. weights anti-proportional to the forecasting errors) or the confirmation of the findings for other datasets, i.e. commodities, non-European stock indices, etc. It would also be interesting to explore whether we can enhance the forecasting accuracy for other measures of volatility, such as the realized kernels and bi-power variation, or from information extracted from futures and options.

Acknowledgements

The author would like thank the editor (Prof. Ali Fatemi) and the anonymous referee for their invaluable comments and suggestions on a previous version of this paper, which helped to significantly improve the quality of the paper. He also thanks Helga Stefansson for editing the manuscript. The author acknowledges the support from the European Community's Seventh Framework Programme (Marie Curie FP7-PEOPLE-IEF & FP7-PEOPLE-RG) funded under grant agreements no. PIEF-GA-2009-237022 & PERG08-GA-2010-276904. The author is solely responsible for any remaining errors and deficiencies.

References

Alquist, R., & Kilian, L. (2010). What do we learn from the price of crude oil futures? *Journal of Applied Econometrics, 25*(4), 539–573.
Amendola, A., & Storti, G. (2008). A GMM procedure for combining volatility forecasts. *Computational Statistics and Data Analysis, 52*(6), 3047–3060.
Andersen, T., Bollerslev, T., Christoffersen, P., & Diebold, F. X. (2006). Volatility and correlation forecasting. In G. Elliott, C. W. J. Granger, & A. Timmermann (Eds.). *Handbook of economic forecasting*. Amsterdam: North Holland Press.
Andersen, T., Bollerslev, T., & Diebold, F. (2007). Roughing it up: Including jump components in the measurement, modelling and forecasting of return volatility. *The Review of Economics and Statistics, 89*(4), 701–720.
Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica, 71*(2), 579–625.
Baillie, R. T., Bollerslev, T., & Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics, 74*, 3–30.
Bates, J., & Granger, C. W. J. (1969). The combination of forecasts. *Operations Research Quarterly, 20*, 451–468.
Blair, B. J., Poon, S.-H., & Taylor, S. J. (2001). Forecasting S&P100 volatility: The incremental information content of implied volatilities and high-frequency index returns. *Journal of Econometrics, 105*, 5–26.
Box, G. E. P., & Tiao, G. C. (1973). *Bayesian inference in statistical analysis. Vol. 40*. Reading: Addison-Wesley.
Chen, X., & Ghysels, E. (2011). News-good or bad-and its impact on volatility forecasts over multiple horizons. *Review of Financial Studies, 24*, 46–81.
Chiriac, R., & Voev, V. (2011). Modelling and forecasting multivariate realized volatility. *Journal of Econometrics, 26*(6), 922–947.
Christensen, B. J., & Prabhala, N. R. (1998). The relation between implied and realised volatility. *Journal of Financial Economics, 50*, 125–150.
Clark, T. E., & West, K. D. (2007). Approximately normal tests for equal predictive accuracy in nested models. *Journal of Econometrics, 138*, 291–311.
Clements, M. P., Galvão, A. B., & Kim, J. H. (2008). Quantile forecasts of daily exchange rate returns from forecasts of realized volatility. *Journal of Empirical Finance, 15*(4), 729–750.
Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics, 7*(2), 174–196.
Corsi, F., Mittnik, S., Pigorsch, C., & Pigorsch, U. (2008). The volatility of realised volatility. *Econometric Reviews, 27*(1–3), 46–78.
Corsi, F., & Reno, R. (2012). Discrete-time volatility forecasting with persistent leverageeffect and the link with continuous-time volatility modeling. *Journal of Business & Economic Statistics, 30*(3), 368–380.
Degiannakis, S. (2004). Volatility forecasting: Evidence from a fractional integrated asymmetric power ARCH skewed-t model. *Applied Financial Economics, 14*, 1333–1342.
Degiannakis, S. (2008). Forecasting VIX. *Journal of Money, Investment and Banking, 4*, 5–19.
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., & Floros, C. (2015). *Modelling and forecasting high frequency financial data.* Hampshire: Palgrave - MacMillan Ltd.
Degiannakis, S., Floros, C., & Dent, P. (2014). A Monte Carlo simulation approach to forecasting multi-period value-at-risk and expected shortfall using the FIGARCH-skT specification. *The Manchester School, 82*(1), 71–102.
Degiannakis, S., & Livada, A. (2016). Evaluation of realized volatility predictions from models with leptokurtically and asymmetrically distributed forecast errors. *Journal of Applied Statistics, 43*(5), 871–892.
Degiannakis, S., & Xekalaki, E. (2005). Predictability and model selection in the context of ARCH models. *Journal of Applied Stochastic Models in Business and Industry, 21*, 55–82.
Deo, R., Hurvich, C., & Lu, Y. (2005). Forecasting realized volatility using a long-memory stochastic volatility model: Estimation, prediction and seasonal adjustment. *Journal of Econometrics, 131*, 29–58.
Diebold, F. X., & Mariano, R. S. (1995). Comparing predictive accuracy. *Journal of Business and Economic Statistics, 13*, 253–263.
Fernandez, C., & Steel, M. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association, 93*, 359–371.
Fleming, J. (1998). The quality of market volatility forecast implied by S&P 100 index option prices. *Journal of Empirical Finance, 5*, 317–345.
Fleming, J., Ostdiek, B., & Whaley, R. E. (1995). Predicting stock market volatility: A new measure. *Journal of Futures Markets, 15*, 265–302.
Frijns, B., Tallau, C., & Tourani-Rad, A. (2010). The information content of implied volatility: Evidence from Australia. *Journal of Futures Markets, 30*(2), 134–155.
Ghysels, E., Sinko, A., & Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews, 26*(1), 53–90.
Giot, P. (2003). The information content of implied volatility in agricultural commodity markets. *Journal of Futures Markets, 23*, 441–454.
Giot, P., & Laurent, S. (2003). Value-at-risk for long and short trading positions. *Journal of Applied Econometrics, 18*, 641–664.
Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics, 14*, 227–238.
Granger, C. W. J., & Joyeux, R. (1980). An introduction to long memory time series models and fractional differencing. *Journal of Time Series Analysis, 1*, 15–39.
Granger, C. W. J., & Newbold, P. (1977). *Forecasting economic time series.* New York: Academic Press.
Granger, C. W. J., & Ramanathan, R. (1984). Improved methods of combining forecasts. *Journal of Forecasting, 3*(2), 197–204.
Hansen, P. R. (2005). A test for superior predictive ability. *Journal of Business & Economic Statistics, 23*, 365–380.
Hansen, P. R., & Lunde, A. (2005). A realized variance for the whole day based on intermittent high-frequency data. *Journal of Financial Econometrics, 3*(4), 525–554.
Hansen, P. R., & Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business and Economic Statistics, 24*(2), 127–161.
Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica, 79*, 456–497.
Hu, M. Y., & Tsoukalas, C. (1999). Combining conditional volatility forecasts using neural networks: An application to the EMS exchange rates. *Journal of International Financial Markets Institutions and Money, 9*, 407–422.
Hua, J., & Manzan, S. (2013). Forecasting the return distribution using high-frequency volatility measures. *Journal of Banking & Finance, 37*(11), 4381–4403.
International Monetary Fund (2007). *World economic outlook.* DC: Washington.
Johnson, N. L., Kotz, S., & Balakrishnan, N. (1995). *Continuous univariate distributions* (2nd edition). *Vol. 2*. New York: John Wiley and Sons.
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*, 445–475.
Krishnamoorthy, A. S., & Parthasarathy, M. (1951). A multivariate gamma - type distribution. *Annals of Mathematical Statistics, 22*(4), 549–557.
Lambert, P., & Laurent, S. (2001). *Modeling financial time series using GARCH-type models and a skewed student density.* Mimeo: Universite de Liege.

Liu, C., & Maheu, J. M. (2009). Forecasting realized volatility: A Bayesian model-averaging approach. *Journal of Applied Econometrics, 24*(5), 709–733.
Martens, M., & Zein, J. (2002). Predicting financial volatility: High-frequency time-series forecasts vis-`a-vis implied volatility. *Journal of Futures Markets, 24*(11), 1005–1028.
Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V., & VonWeizsäcker, J. E. (1997). Volatilities of different time resolutions – Analyzing the dynamics of market components. *Journal of Empirical Finance, 4*, 213–239.
Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics, 160*(1), 246–256.
Pong, S., Shackleton, M. B., Taylor, S. J., & Xu, X. (2004). Forecasting currency volatility: A comparison of implied volatilities and AR(FI)MA models. *Journal of Banking & Finance, 28*(10), 2541–2563.
Prokopczuk, M., Symeonidis, L., & Wese Simen, C. (2015). Do jumps matter for volatility forecasting? Evidence from Energy Markets. *Journal of Futures Markets,* 1–35.
Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics, 6*, 461–464.
Sévi, B. (2014). Forecasting the volatility of crude oil futures using intraday data. *European Journal of Operational Research, 235*(3), 643–659.
Shephard, N., & Sheppard, K. (2010). Realising the future: Forecasting with high frequency based volatility (HEAVY) models. *Journal of Applied Econometrics, 25*, 197–231.
Svensson, L. E. O. (2005). *Oil prices and ECB monetary policy.* Mimeo: Department of Economics, Princeton University.
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.
Wasserfallen, W., & Zimmermann, H. (1985). The behavior of intra-daily exchange rates. *Journal of Banking & Finance, 9*, 55–72.
White, H. (2000). A reality check for data snooping. *Econometrica, 68*, 1097–1126.
Xekalaki, E., & Degiannakis, S. (2005). Evaluating volatility forecasts in option pricing in the context of a simulated options market. *Computational Statistics and Data Analysis, 49*(2), 611–629.
Xekalaki, E., & Degiannakis, S. (2010). *ARCH models for financial applications.* New York: John Wiley & Sons Ltd.