RESEARCH ARTICLE

Forecasting realized volatility: The role of implied volatility, leverage effect, overnight returns, and volatility of realized volatility

Dimos S. Kambouroudis[1] | **David G. McMillan[1]** | **Katerina Tsakou[2]**

[1]Department of Accounting and Finance, University of Stirling, Stirling, UK

[2]Department of Accounting and Finance, School of Management, Swansea University, Swansea, UK

Correspondence
Katerina Tsakou, Department of Accounting and Finance, School of Management, Swansea University, Bay Campus, Fabian Way, Swansea SA1 8EN, UK.
Email: katerina.tsakou@swansea.ac.uk

Abstract

We forecast realized volatility extending the heterogeneous autoregressive model (HAR) to include implied volatility (IV), the leverage effect, overnight returns, and the volatility of realized volatility. We analyze 10 international stock indices finding that, although a simple HAR model augmented with IV (HAR-IV) is more accurate than any HAR model excluding it, all markets support further extensions of the HAR-IV model. More accurate forecasts are found using overnight returns in all markets except the UK, the volatility of realized volatility in the US, and the leverage effect in five markets. A value-at-risk exercise supports the economic significance of our findings.

KEYWORDS
HAR modeling and forecasting, implied volatility indices, leverage effect, overnight returns, realized volatility

1 | INTRODUCTION

The volume of academic research on volatility modeling and forecasting confirms both its importance and significance within the finance discipline. Volatility plays an important role in both theoretical and empirical finance, for example, as a key input in asset and derivative pricing as well as in risk management. Therefore, obtaining accurate forecasts is critical.

A recurring question is the ability of *Implied Volatility* (IV) to provide information content for stock return volatility forecasts. The view that IV might provide forecast power for stock volatility lies in the fact that it contains investor expectations of future economic conditions. Implied volatility indices (VIX), which are derived from option prices, reflect market expectations regarding future volatility over the life of the option contract and thus, provide a measure of market uncertainty. Indeed, it is argued that if the option market is efficient, then IV should be the only variable required to forecast subsequent volatility (see, e.g., Chiras & Manaster, 1978; Christensen & Prabhala, 1998). Supportive evidence that IV forecasts subsume an autoregressive conditional heteroscedasticity (ARCH) alternative is provided by Corrado and Miller (2005). However, other work is less supportive, see, for example, Becker et al. (2007). Nonetheless, some evidence does suggest that incorporating VIX into standard volatility models does improve the accuracy of forecasts (Blair et al., 2001; Frijns et al., 2010; Kambouroudis & McMillan, 2016; Kambouroudis et al., 2016; Yu et al., 2010).

At the same time, the availability of high-frequency data has enabled the development of nonparametric daily estimators of volatility, known as realized volatility. The most popular such measure is the sum of squared intraday returns, after the seminal work of Andersen and Bollerslev (1998). For (at least) two reasons, much attention is devoted to realized volatility models over the last two decades. First, realized volatility is a more accurate measure of volatility which, unlike squared daily returns, ably captures the persistence of volatility. Second, well-known reduced form time series models (e.g., ARFIMA) based on realized volatility outperform GARCH and stochastic volatility models in forecasting future volatility (Andersen et al., 2003).[1]

Due to its simplicity and ability to reproduce volatility persistence, the most popular realized volatility modeling approach is the Heterogeneous Autoregressive (HAR) model proposed by Corsi (2009). Several extensions of the standard HAR model are proposed to capture the stylized facts of volatility. First, Corsi and Renò (2012) allowing for the leverage effect, that is, that volatility increases more after a negative shock than an equal positive shock, find that its impact on realized volatility is significant for the S&P500 index.[2] Second, Wang et al. (2015) include overnight returns along with the leverage effect and find significant effects for the nighttime arrival of news and asymmetry in the Chinese stock market.[3] Third, arguing that, apart from volatility itself, the volatility of realized volatility may cluster, Corsi et al. (2008) note that the inclusion of the time-varying volatility of realized volatility leads to improved S&P500 volatility forecasts.

While IV has been widely exploited in low-frequency volatility modeling, its role in realized volatility models has not been considered until the work of Busch et al. (2011) who, employing a standard HAR model, find that IV is a strong predictor of realized volatility. Buncic and Gisler (2016) show that VIX contains a substantial amount of predictive information for forecasting the realized volatility of international equity markets, augmenting each foreign market's HAR model with both forward- and backward-looking US volatility information.[4] Despite such convincing evidence on the role of IV in a standard HAR model, the wider interaction between IV and the stylized facts of volatility in the family of HAR-type models is unknown. While the above-discussed extensions of HAR are widespread, to the best of our knowledge, the corresponding extensions of an IV-augmented HAR (HAR-IV) have not been assessed. In this study we investigate the predictive performance of HAR extensions once IV is included and so consider HAR-IV extensions accounting for the leverage effect, overnight returns, and the volatility of realized volatility. Ultimately, the aim is to find the most accurate predictive specification, while also drawing conclusions on the individual and joint role of such extensions.

Unlike most of the existing studies on the role of IV in realized volatility models, which focus on the United States, we additionally consider stock market indices in Europe and Japan. All data are observed at daily frequency in the sample running from February 2, 2001 to June 27, 2019.

Our main results are as follows. First, the inclusion of IV in the standard HAR model substantially improves volatility forecasts in all 10 international stock markets, confirming the importance of IV information. Not only does the HAR-IV outperform the standard HAR but, in all markets, is also more accurate than any HAR model variant that does not include IV. Furthermore, using the Giacomini and Rossi (2010) testing approach for local superior predictive ability and so assessing how the relative forecasting performance of our models evolve, this result is consistent over our sample period. Second, no single extension of HAR-IV outperforms the others across all the markets. Nonetheless, overnight returns are beneficial for several indices in improving the accuracy of the predictive model. While the inclusion of IV does not impair the evidence of Corsi et al. (2008) in favor of including the volatility of realized volatility in the United States, no other market leads us to the same conclusion.[5] More mixed evidence is found on the leverage effect which improves predictions only in five markets. Third, our main conclusions are unchanged once we apply our volatility forecasts to the practice of risk management within a value-at-risk (VaR) framework.

The remainder of the paper is organized as follows. In Section 2, we outline realized volatility and describe the predictive models together with the forecast evaluation techniques. The data are presented in Section 3. In Section 4, we present the empirical results of our volatility forecasting exercise, while VaR results are in Section 5. Section 6 concludes.

---

[1]See, also, Martens and Zein (2004), Pong et al. (2004), and Koopman et al. (2005).
[2]Similar results are provided by Martens et al. (2009) and Wang et al. (2015) for the S&P500 index and the Chinese stock market, respectively.
[3]Gallo (2001) and Tsiakas (2008) also provide evidence about the predictive content of overnight returns for the subsequent daytime volatility.
[4]The ability of option-implied information to forecast realized volatility has also been examined by Oikonomou et al. (2019) and Jeon et al. (2020).
[5]More precisely, Corsi et al. (2008) find evidence of volatility of realized volatility in the S&P500. Accounting for IV we reach the same conclusion for both S&P500 and Dow Jones Industrial Average (DJIA)—but not for the less liquid Nasdaq100.

2 | EMPIRICAL FRAMEWORK

This section starts with a brief introduction to realized volatility. We then describe the HAR model of Corsi (2009) and its various extensions that will be used to assess the role of IV, leverage effect, overnight returns, and volatility of realized volatility on realized forecasts. We conclude this section with a description of the techniques we implement to evaluate the forecast performance of our models.

2.1 | Realized measure

Consider that the logarithm of an asset price at time t, p_t, is a continuous-time diffusion process

$$dp_t = \mu_t d_t + \sigma_t dW_t, \tag{1}$$

where μ_t is a locally bounded predictable drift, W_t is the Brownian motion, and σ_t is a stochastic process independent of W_t. For this log-price process, the integrated variance is

$$IV_t = \int_0^t \sigma_s^2 \, ds. \tag{2}$$

As shown in Andersen et al. (2001) a natural estimator for the integrated variance is the sum of squared intraday returns, which is commonly known as realized variance (RV). The standard definition of the RV estimator of integrated variance is

$$RV_t = \sum_{i=1}^{m} r_{t,i}^2, \tag{3}$$

where m is the number of intraday returns during day t. Letting $m \to \infty$, that is, in case of continuous sampling, RV_t converges to the true integrated variance.

Subsequently, several alternatives to the standard RV have been proposed (see, Barndorff-Nielsen & Shephard, 2004; Hansen & Lunde, 2006, among others). However, the standard RV, based on 5-min intraday returns, remains the most popular in practical applications, for example, Bollerslev et al. (2009), Sévi (2014), Buncic and Gisler (2016), Bollerslev et al. (2018), Gong and Lin (2018), Jeon et al. (2020), and Zhang et al. (2020). Notably, Liu et al. (2015) consider the problem of comparing the empirical accuracy of over 400 different volatility estimators constructed from high-frequency data across multiple assets and conclude that "it is difficult to significantly beat 5-minute RV." Therefore, following the sampling frequency used in much of the existing RV literature, we choose the 5-min RV.

2.2 | Modeling realized volatility

We adopt the HAR model of Corsi (2009) as our benchmark to model and forecast RV. Corsi (2009), inspired by the Heterogeneous Market Hypothesis of Muller et al. (1997), addresses the heterogeneity that arises from market participants operating across a spectrum of different trading frequencies. The main idea being that investors with different time horizons perceive, react to and cause different types of volatility. The HAR model is an additive cascade model of different volatility components (daily, weekly, and monthly) designed to mimic the trading frequencies of the different agents.

The HAR model is defined as

$$y_{t+1} = \alpha_0 + \alpha_d y_t^{(d)} + \alpha_w y_t^{(w)} + \alpha_m y_t^{(m)} + \varepsilon_{t+1}, \tag{4}$$

where $y_{t+1} = log(RV_{t+1}^2)$, ε_{t+1} is an innovation term and $y_t^{(d)}$, $y_t^{(w)}$, and $y_t^{(m)}$ are the three realized volatility components—daily, weekly, and monthly, respectively. Our HAR benchmark model differs from the original HAR specification of Corsi (2009) in two dimensions. First, our model relies on the logarithmic RV, y_{t+1} instead of realized volatility per se, as in the original HAR formulation of Corsi (2009). This choice is motivated by the

observation of Andersen et al. (2003) who note that the distribution of logarithmic RV is much closer to normal, and Andersen et al. (2007) who report similar results when RV, realized volatility, and their logarithms are used to estimate the HAR model parameters. Second, we use a slightly different lag structure for constructing the weekly and monthly volatility components to the one in Corsi (2009). The original HAR model regresses RV on the previous 1-, 5-, and 22-day average realized volatility. To avoid overlapping terms and, thus, ease inter-pretation, we use the numerically identical reparameterization implemented in Patton and Sheppard (2015). Here, the weekly component is constructed based on $log\,RV$ between lag 2 and 5, with the monthly component based on $log\,RV$ between lag 6 and 22.[6] Thus, to formalize the structure of our HAR model for RV, let $y_t^{(d)} = log\,RV_t^2$, $y_t^{(w)} = \frac{1}{4}\sum_{i=1}^{4} log\,RV_{t-i}^2$, and $y_t^{(m)} = \frac{1}{17}\sum_{i=5}^{21} log\,RV_{t-i}^2$ be the daily, weekly, and monthly components, respectively. One of the reasons for the popularity of the HAR model in Equation (4) is its simplicity. Notably, once the three volatility components have been constructed, the HAR model is estimated by ordinary least squares (OLS).

To explicitly capture the salient features of stock index RV, we extend the benchmark model in Equation (4) to allow for IV, the leverage effect, overnight returns, and the volatility of realized volatility. Hence, we consider a more general HAR specification of the form

$$y_{t+1} = \alpha_0 + \alpha_d y_t^{(d)} + \alpha_w y_t^{(w)} + \alpha_m y_t^{(m)} + \beta' X + \varepsilon_{t+1}, \tag{5}$$

where the model allows for a k-dimensional vector X of noted exogenous variables.[7]

We examine the predictive power of IV on RV by augmenting the HAR model with IV similar to Busch et al. (2011). As the nature of the IV indices appears to be highly persistent, we use a HAR-type structure for IV. Of note, Fernandes et al. (2014) evaluate different HAR models to forecast the VIX index and find that it is difficult to beat the forecasting performance of the simple HAR process. Therefore, we add three components of IV in the HAR model (HAR-IV) by setting $\beta' X = \beta_d log\,IV_t^{(d)} + \beta_w log\,IV_t^{(w)} + \beta_m log\,IV_t^{(m)}$ in Equation (5). This allows us to capture the information content of IV across different maturities beyond that within the RV components. The $log\,IV$ components are defined analogously to the RV components.[8]

Corsi and Renò (2012) extend the HAR model by adding positive and negative daily, weekly, and monthly returns to capture the leverage effect. Thus, the HAR-L model is specified by setting $\beta' X = \beta_1 r_t^+ + \beta_2 r_t^-$. As such, RV reacts asymmetrically to previous positive and negative daily returns of equal magnitude.[9]

We capture the arrival of news at nighttime by adding the overnight return as an explanatory variable in the HAR model. Similar to Wang et al. (2015) we refer to HAR-O when we set $\beta' X = \beta_1 r_{over,t}$. The overnight return, $r_{over,t}$, is defined as the difference in the logs of the opening price on day $t + 1$ and the closing price on day t.

We also capture the conditional heteroskedasticity of the innovations of RV. Corsi et al. (2008) observe that the innovations of RV are not identically and independently distributed (iid), but exhibit volatility clustering. To account for the volatility of RV they extend the HAR model by incorporating a GARCH component (HAR-G). So the innovation term in Equation (5) is not a Gaussian white noise, but its variance is time-varying; $\varepsilon_t = h_t z_t$, where $z_t \sim N(0, 1)$ and $h_t^2 = a_0 + a_1 \varepsilon_{t-1}^2 + b_1 h_{t-1}^2$.

---

[6]The HAR model implemented in Patton and Sheppard (2015) has also been used by Prokopczuk et al. (2016) to evaluate the importance of jumps for forecasting RV for four energy futures.

[7]We do not explicitly consider the role and results for jump components in this paper. Important contributions in this area include (Barndorff-Nielsen and Shephard (2005), Andersen et al. (2007), and Corsi and Reno (2012). Of note, Busch et al. (2011) find IV contains additional information about future S&P500 volatility and its continuous path and jump components beyond that in RV and its components. However, more recently, Buncic and Gisler (2017), using aggregate daily realized measures, assessed the role of jumps and leverage in predicting RV for 18 international stock markets. They find that the separation of RV into a continuous and a jump component is beneficial only for the S&P500 index and it has limited value for the non-US markets. Nonetheless, using the bi-power variation, we do consider jumps and find that while accounting for jumps improves performance over the standard HAR model for some indices, it does not improve performance compared with the preferred forecast models reported below. Therefore, we do not report these results in full and instead focus on components that have received less attention in the literature. Nonetheless, the results are available upon request.

[8]Buncic and Gisler (2016) also augment the HAR model with the same three IV components to forecast RV. More recently, Plíhal and Lyócsa (2021) evaluate the role of IV calculated across different maturities (of up to 1 month) to similarly observe additional predictive information in forecasting the EUR/USD exchange rate RV.

[9]In the original analysis, we also included positive and negative returns over the last week and month as in Corsi and Renò (2012). However, their parameters were not significant in most cases and thus were omitted.

Finally, we extend the models by allowing them to capture multiple RV features simultaneously.[10] Thus, we have 16 HAR specifications to model and forecast RV.

2.3 | Forecast evaluation

The ability of the models described in Section 2.2 to accurately forecast RV is assessed using the quasi-likelihood (QLIKE) loss function.

$$L(y_t, \hat{y}_{it}) = \frac{y_t}{\hat{y}_{it}} - log \frac{y_t}{\hat{y}_{it}} - 1, \quad t = N + 1, N + 2, ..., T, \tag{6}$$

where $\hat{y}_{it}$ is the prediction of y_t based on $t - 1$ information obtained with the ith predictive model, T is the sample size, and N is the size of the estimation window. The model that yields the smallest average loss is the most accurate and therefore preferred. Because true volatility is latent, forecasted volatility must be compared against an ex post proxy of volatility that is imperfect in nature. Patton (2011) shows that the QLIKE loss function is robust in the sense that the ranking of the models is the same whether the ranking is done using the true volatility or some unbiased proxy, such as the RV presented in Equation (3). The QLIKE loss function is also used by Brownlees et al. (2011), Sévi (2014), and Tian et al. (2017), among others, to evaluate their forecasts.

To test for significant differences in predictive accuracy of the models, we employ the model confidence set (MCS) approach developed by Hansen et al. (2011) and the pairwise test proposed by Giacomini and White (2006). The MCS procedure determines the set that contains the best models from a full set of models M. Starting with the full set, M, the MCS procedure sequentially eliminates the models that are found to be significantly inferior until the null hypothesis of equal forecast accuracy is no longer rejected at the α significance level. The set of surviving models is the MCS, $\hat{M}^*_{1-\alpha}$, containing the best models from M at the $(1 - \alpha)$ confidence level.

Following Hansen et al. (2011), we test the null hypothesis of equal predictive ability using the range statistics $T_R = \max\limits_{i,j \in M} \frac{|\bar{d}_{ij}|}{\sqrt{v\hat{a}r(\bar{d}_{ij})}}$ and the semiquadratic statistics $T_{SQ} = \sum\limits_{i,j \in M} \frac{(\bar{d}_{ij})^2}{v\hat{a}r(\bar{d}_{ij})}$, where $d_{ij,t} = L(y_t; \hat{y}_{it}) - L(y_t; \hat{y}_{jt})$ is the loss differential between the ith and the jth model, $\bar{d}_{ij} = \frac{1}{n}\sum_{t=N+1}^{T} d_{ij,t}$ is its average value among the $n = T - N$ one-period ahead forecasts computed, and $\widehat{var}(\bar{d}_{ij})$ is an estimate of $var(\bar{d}_{ij})$ obtained by using a block-bootstrap procedure of 10,000 resamples. The MCS procedure yields p-values for each model in the initial set. For a given model $i \in M$, the MCS p-value, $\hat{p}_i$, is the threshold at which $i \in \widehat{M}^*_{1-\alpha}$, if and only if $\hat{p}_i \geq \alpha$.

The Giacomini–White (GW) test of conditional predictive ability evaluates the forecasting performance of two competing models, accounting for parameter uncertainty. The null hypothesis of equal predictive ability $H_0 : E(d_{ij,t}) = 0$ is tested using the Wald statistic

$$CPA = n \left(n^{-1} \sum_{t=N+1}^{T} d_{ij,t} \right)' \hat{\Omega}_n^{-1} \left(n^{-1} \sum_{t=N+1}^{T} d_{ij,t} \right) \sim \chi_1^2, \tag{7}$$

where $\hat{\Omega}_T$ is the Newey and West (1987) HAC estimator of the asymptotic variance of $d_{ij,t}$.

The MCS procedure and the GW test help evaluate the model with the best forecast performance on average over the out-of-sample period. However, in unstable environments, the relative forecasting performance of the models may

---

[10]Specifically, HAR-IVL, HAR-IVO, and HAR-IVG are extensions of the HAR-IV model that seek to capture the leverage effect, overnight returns, and the volatility of RV, respectively. HAR-LO and HAR-LG are extensions of the HAR-L model that capture overnight returns and the volatility of RV, respectively. HAR-OG extends the HAR model to account for both overnight returns and the volatility of RV. HAR-LOG is an extension of the HAR model to include simultaneously the leverage effect, overnight returns, and the volatility of RV. HAR-IVLO and HAR-IVLG are extensions of the HAR-IVL model to also include overnight returns and the volatility of RV, respectively. HAR-IVOG extends the HAR-IV specification to capture both overnight returns and the volatility of RV. Finally, HAR-IVLOG is the most sophisticated specification in this study that extends the simple HAR model to include simultaneously the IV, leverage effect, overnight returns, and the volatility of RV.

change over time. To evaluate whether the relative predictive performance of the models is indeed time-varying we consider the cumulative forecast error of the ith predictive model (based on QLIKE loss)

$$cumFE_{it} = \sum_{\tau=N+1}^{t} L(y_\tau, \hat{y}_{i\tau}) \qquad (8)$$

for each prediction $t = N + 1, N + 2, ..., T$. In addition to $cumFE_{it}$ we apply the Giacomini and Rossi (2010) fluctuation test to formally test whether the models′ relative performance is time-varying. We test the null hypothesis that the ith predictive model has equal predictive ability with HAR-IV using the test statistic:

$$F_{t,k} = \hat{\sigma}^{-1}k^{-1/2} \sum_{s=t-k/2}^{t+k/2-1} d_{oi,s} \qquad (9)$$

for $t = N + 1 + k/2, ..., T - k/2 + 1$, where o is the HAR-IV model and $\hat{\sigma}^2$ is the HAC standard error estimator of the variance of $d_{oi,t}k$. Hence, the fluctuation test is similar to the GW test described before, but is computed over a rolling out-of-sample window of size k. If the $max_t F_{t,k}$ is higher than the critical value the null hypothesis of local equal predictive ability between the two models is rejected.[11]

3 | DATA

Our data set consists of daily RV data, open and close prices, and IV indices for 10 international equity markets. The daily RV measures that we employ are based on 5-min intraday returns and are obtained from the Oxford-Man Institute's Quantitative Finance Realized Library (Heber et al., 2009).[12] The daily open and close prices, and the IV indices are collected from Thomson Reuters Datastream (2019). Since the various IV indices have been listed on different dates, we consider the period from February 2, 2001 to June 27, 2019 to study the indices over the same time period.

The 10 international stock markets we examine are: the S&P500 (US), the DJIA (US), the Nasdaq100 (US), the Euro STOXX50 (Euro area), the CAC40 (France), the DAX (Germany), the AEX (The Netherlands), the SMI (Switzerland), the FTSE100 (United Kingdom), and the Nikkei225 (Japan). The IV indices constructed from the market prices of options on these stock indices are the VIX, VXD, VXN, VSTOXX, VCAC, VDAX-New, VAEX, VSMI, VFTSE, and VXJ, respectively. They represent the expected market volatility over the subsequent 22 trading days. The CBOE VIX index is probably the most widely used example of option-implied information and its popularity spawned the introduction of IV indices around the world.[13]

Table 1 provides the descriptive statistics for the (log) RV and IV as well as for daily returns of all data we use. The last six columns of the table provide the first to third order of the autocorrelation function (ACF) and partial ACF (PACF). The *log RV* appears approximately Gaussian with the skewness being between 0 and 1 and the kurtosis being around 3 for all indices. The only exception is the *log RV* series for STOXX50 where a kurtosis of 5.635 suggests slightly fatter tails than a Gaussian variable. As documented in other studies, the ACF and PACF values highlight the long memory characteristic of (realized) volatility. Similar features emerge for the *log IV* series, being close to a normal distribution, except for the *log IV* series of Nikkei225 that has fatter tails. Also, the *log IV* series are the most persistent series overall. Finally, the daily returns of all indices are negatively skewed and leptokurtic. This suggests that the return distribution is not symmetric and it has heavier tails than the normal.

---

[11]Under the null hypothesis, the asymptotic distribution of the fluctuation test can be approximated by functionals of the Brownian motion. Critical values are computed by simulation and are reported in Giacomini and Rossi (2010).

[12]http://realized.oxford-man.ox.ac.uk/data/download. Our RVs are from Library Version 0.3. The Oxford-Man Institute's Quantitative Finance Realized Library contains a selection of daily nonparametric estimated of volatility.

[13]All IV indices apply the VIX algorithm that relies on a model-free approach rather than a model-based approach that depends on option pricing structure, such as the Black–Scholes model (Jiang and Tian, 2005). As the VIXs are quoted in annualized percentages and to make them comparable to RV measures, we compute the daily IV index as $(IV/(100\sqrt{252}))^2$. The same approach is also followed by Blair et al. (2001) and Becker et al. (2007), among others, to evaluate the information content of IV about future volatility when the alternative is historical models based only on daily returns.

TABLE 1 Summary statistics of (log) RV, (log) IV, and daily returns

	Mean	Std.dev.	Skew.	Kurt.	ACF(1–3)			PACF(1–3)		
log(RV)										
S&P500	−9.906	1.141	0.388	3.349	0.819	0.784	0.755	0.819	0.344	0.177
DJIA	−9.895	1.117	0.439	3.474	0.787	0.757	0.729	0.787	0.362	0.193
Nasdaq100	−9.778	1.028	0.414	3.077	0.843	0.792	0.761	0.843	0.283	0.165
STOXX50	−9.368	1.047	0.051	5.635	0.765	0.725	0.708	0.765	0.336	0.225
CAC40	−9.494	0.984	0.435	3.217	0.830	0.792	0.774	0.830	0.329	0.214
DAX	−9.370	1.043	0.460	3.277	0.820	0.786	0.771	0.820	0.345	0.227
AEX	−9.696	1.028	0.533	3.228	0.845	0.807	0.788	0.845	0.326	0.207
SMI	−9.926	0.908	0.925	3.949	0.858	0.823	0.805	0.858	0.328	0.209
FTSE100	−10.039	0.997	0.608	3.405	0.824	0.790	0.769	0.824	0.348	0.203
Nikkei225	−9.862	0.924	0.323	3.412	0.787	0.735	0.704	0.787	0.304	0.178
log(IV)										
S&P500	−8.965	0.746	0.791	3.485	0.983	0.968	0.955	0.983	0.075	0.057
DJIA	−9.073	0.717	0.864	3.567	0.983	0.970	0.959	0.983	0.106	0.047
Nasdaq100	−8.526	0.819	0.874	2.972	0.989	0.979	0.971	0.989	0.037	0.065
STOXX50	−8.553	0.727	0.625	3.077	0.986	0.972	0.960	0.986	0.035	0.036
CAC40	−8.673	0.712	0.600	3.146	0.984	0.970	0.957	0.984	0.063	0.047
DAX	−8.584	0.713	0.816	3.367	0.987	0.975	0.964	0.987	−0.001	0.045
AEX	−8.729	0.809	0.736	3.135	0.987	0.976	0.967	0.987	0.076	0.049
SMI	−9.031	0.707	1.036	3.931	0.989	0.976	0.964	0.989	−0.061	0.027
FTSE100	−9.012	0.768	0.736	3.241	0.983	0.970	0.958	0.983	0.089	0.041
Nikkei225	−8.440	0.623	0.674	4.188	0.981	0.964	0.949	0.981	0.046	0.037
Returns										
S&P500	$1.67 * 10^{-4}$	0.012	−0.245	12.287	−0.078	−0.045	0.029	−0.078	−0.052	0.021
DJIA	$1.93 * 10^{-4}$	0.011	−0.099	11.871	−0.074	−0.040	0.041	−0.074	−0.046	0.035
Nasdaq100	$2.44 * 10^{-4}$	0.016	−0.038	8.612	−0.049	−0.061	0.019	−0.049	−0.064	0.013
STOXX50	$−6.46 * 10^{-5}$	0.014	−0.052	8.076	−0.028	−0.039	−0.043	−0.028	−0.040	−0.046
CAC40	$−1.25 * 10^{-5}$	0.014	−0.028	8.541	−0.029	−0.035	−0.048	−0.029	−0.036	−0.050
DAX	$1.32 * 10^{-4}$	0.015	−0.078	7.923	−0.015	−0.019	−0.019	−0.015	−0.019	−0.019
AEX	$−2.63 * 10^{-5}$	0.014	−0.105	9.907	−0.004	−0.008	−0.054	−0.004	−0.008	−0.054
SMI	$4.50 * 10^{-5}$	0.012	−0.179	10.060	0.029	−0.046	−0.039	0.029	−0.046	−0.037
FTSE100	$3.62 * 10^{-5}$	0.012	−0.160	9.839	−0.043	−0.040	−0.051	−0.043	−0.042	−0.055
Nikkei225	$9.81 * 10^{-5}$	0.015	−0.407	9.545	−0.041	−0.006	−0.011	−0.041	−0.008	−0.012

Note: Entries report the summary statistics of the daily (log) realized volatility, the (log) implied volatility, and the returns of the 10 international stock markets. The sample period covers from February 2, 2001 to June 27, 2019. The last six columns of the table provide the first to the third order of the autocorrelation function (ACF) and partial ACF (PACF).

4 | RESULTS

Table 2 presents the out-of-sample forecast evaluation results for all 10 indices using a rolling estimation window with an initial in-sample period from February 2, 2001 to December 31, 2005. To assess the one-step-ahead forecasts we report the average QLIKE loss of the benchmark HAR model in Equation (4) and the 15 extensions of the HAR model presented in Equation (5). The relative forecast error obtained by the ratio of the average QLIKE loss for each model to the average QLIKE loss of the benchmark HAR is provided in parentheses. The model that yields the lowest loss, and thus the model with the best forecast performance is indicated in bold for every index.

TABLE 2 Out-of-sample performance of the model specifications for each of the stock market indices

	S&P500	DJIA	Nasdaq100	STOXX50	CAC40	DAX	AEX	SMI	FTSE100	Nikkei225
HAR	0.2421	0.2756	0.1725	0.2444	0.1549	0.1645	0.1552	0.1187	0.1537	0.2255
HAR-IV	0.1959 (0.8093)	0.2222 (0.8062)	0.1492 (0.8649)	0.1828 (0.7478)	0.1347 (0.8701)	0.1416 (0.8608)	0.1393 (0.8978)	0.1083 (0.9124)	0.1411 (0.9179)	0.2007 (0.8899)
HAR-L	0.2320 (0.9585)	0.2615 (0.9488)	0.1606 (0.9310)	0.2186 (0.8942)	0.1454 (0.9389)	0.1552 (0.9432)	0.1446 (0.9318)	0.1106 (0.9313)	0.1453 (0.9450)	0.2169 (0.9617)
HAR-O	0.2323 (0.9596)	0.2744 (0.9957)	0.1570 (0.9100)	0.2366 (0.9680)	0.1432 (0.9245)	0.1509 (0.9174)	0.1381 (0.8903)	0.1088 (0.9168)	0.1542 (1.0032)	0.2126 (0.9429)
HAR-G	0.2417 (0.9983)	0.2753 (0.9986)	0.1737 (1.0068)	0.2519 (1.0304)	0.1561 (1.0059)	0.1655 (1.0102)	0.1568 (1.0102)	0.1214 (1.0228)	0.1554 (1.0111)	0.2269 (1.0065)
HAR-IVL	0.1978 (0.8171)	0.2251 (0.8165)	0.1476 (0.8554)	0.1816 (0.7431)	0.1331 (0.8596)	0.1403 (0.8526)	0.1353 (0.8721)	0.1054 (0.8883)	**0.1385 (0.9011)**	0.2003 (0.8883)
HAR-IVO	0.1880 (0.7764)	0.2193 (0.7955)	0.1367 (0.7923)	0.1721 (0.7039)	0.1216 (0.7811)	0.1285 (0.7887)	0.1224 (0.7887)	0.0986 (0.8308)	0.1413 (0.9193)	**0.1878 (0.8331)**
HAR-IVG	0.1950 (0.8053)	0.2210 (0.8018)	0.1496 (0.8672)	0.1847 (0.7556)	0.1351 (0.8724)	0.1416 (0.8607)	0.1395 (0.8993)	0.1098 (0.9246)	0.1424 (0.9264)	0.2008 (0.8904)
HAR-LO	0.2225 (0.9189)	0.2589 (0.9392)	0.1467 (0.8503)	0.2098 (0.8582)	0.1331 (0.8592)	0.1422 (0.8642)	0.1281 (0.8254)	0.1018 (0.8575)	0.1457 (0.9477)	0.2048 (0.9084)
HAR-LG	0.2318 (0.9576)	0.2619 (0.9503)	0.1609 (0.9327)	0.2248 (0.9198)	0.1460 (0.9428)	0.1553 (0.9439)	0.1454 (0.9369)	0.1117 (0.9414)	0.1460 (0.9500)	0.2172 (0.9634)
HAR-OG	0.2326 (0.9609)	0.2741 (0.9944)	0.1579 (0.9154)	0.2509 (1.0263)	0.1437 (0.9282)	0.1512 (0.9191)	0.1389 (0.8949)	0.1103 (0.9289)	0.1559 (1.0145)	0.2133 (0.9458)
HAR-IVLO	0.1896 (0.7830)	0.2220 (0.8053)	**0.1350 (0.7825)**	**0.1709 (0.6993)**	**0.1199 (0.7744)**	0.1275 (0.7748)	**0.1189 (0.7663)**	**0.0965 (0.8129)**	0.1387 (0.9026)	0.1880 (0.8339)
HAR-IVLG	0.1969 (0.8133)	0.2242 (0.8134)	0.1480 (0.8578)	0.1851 (0.7571)	0.1332 (0.8604)	0.1403 (0.8525)	0.1358 (0.8749)	0.1064 (0.8966)	0.1395 (0.9075)	0.2002 (0.8877)
HAR-IVOG	**0.1872 (0.7733)**	**0.2181 (0.7911)**	0.1370 (0.7940)	0.1738 (0.7108)	0.1217 (0.7801)	0.1284 (0.7891)	0.1224 (0.7891)	0.0991 (0.8348)	0.1427 (0.9282)	0.1880 (0.8338)
HAR-LOG	0.2221 (0.9174)	0.2596 (0.9420)	0.1468 (0.8507)	0.2183 (0.8931)	0.1333 (0.8611)	0.1421 (0.8635)	0.1283 (0.8268)	0.1026 (0.8644)	0.1465 (0.9531)	0.2049 (0.9088)
HAR-IVLOG	0.1888 (0.7798)	0.2211 (0.8020)	0.1352 (0.7835)	0.1728 (0.7071)	0.1201 (0.7756)	**0.1273 (0.7734)**	0.1191 (0.7676)	0.0969 (0.8166)	0.1397 (0.9089)	0.1882 (0.8347)

Note: The table reports the average forecast performance for each model between January 1, 2006 and June 27, 2019 for all 10 indices that we consider using a rolling estimation window with the initial in-sample estimating period from February 2, 2001 to December 31, 2005. The relative forecast error obtained by the ratio of the average QLIKE loss for each model to the average QLIKE loss of the benchmark heterogeneous autoregressive (HAR) is provided in brackets. The model with the best forecast performance, is indicated in bold for every index.

Several interesting conclusions arise from Table 2 regarding the importance of the various features of realized volatility. In most cases, the relative forecast error is below one, indicating the poor performance of the benchmark HAR model. Thus, accounting for one or more of the volatility features appears to improve forecast accuracy. Augmenting the benchmark HAR model with each of the different stylized facts individually (first part of Table 2) enhances the forecast accuracy across all indices. Explicitly incorporating the implied VIX in the HAR model notably improves its forecast performance across all indices. The HAR-IV specification yields the lowest loss when compared with the HAR, HAR-L, HAR-O, and HAR-G across all indices (except AEX). The presence of the leverage effect and overnight returns also improve the forecast performance of the HAR model. The only exception is for the FTSE100 index. In this case the benchmark HAR performs marginally better than the HAR-O (the relative error is higher than 1). Interestingly, accounting for the volatility of realized volatility adds little or nothing in terms of forecast accuracy. The forecast error of the HAR-G is higher than the benchmark HAR for most indices indicating that the HAR model performs better than the HAR-G.

Across all features, accounting for IV contributes most to the improvement in forecast performance of the HAR model. Looking at the forecast errors of the models that account for each of the various stylized facts separately, that is, HAR-IV, HAR-L, HAR-O, and HAR-G, we find that the biggest predictive gains are driven by IV. In addition, the parsimonious HAR-IV specification yields remarkably lower loss than specifications that exclude IV, but simultaneously include two or more of the other stylized facts, such as HAR-LO, HAR-LG, HAR-OG, and HAR-LOG.

Although the HAR-IV model is more accurate than any HAR-type that does not include IV, the results in Table 2 indicate that extensions of the HAR-IV further improve forecast accuracy. Comparing the forecast error of the HAR-IV with more sophisticated models that include not only IV, but also the leverage effect, overnight returns, and/or the volatility of realized volatility we find that an extension of the HAR-IV provides a more accurate forecast of realized volatility across all indices. The HAR-IVOG performs best for the S&P500 and DJIA, while the HAR-IVLO provides the best model for the Nasdaq100 and most of the European indices. The exceptions being the DAX and FTSE100 index, for which the HAR-IVLO is the second-best model, with the first being the HAR-IVLOG and HAR-IVL, respectively. The HAR-IVO is the best model for forecasting the realized volatility of the Nikkei index.

To compare the relative forecast performance of the models we apply the MCS test. On the basis of the QLIKE loss function Table 3 presents the p-values from the MCS test at the 10% and 25% significance levels. The MCS findings reveal the same picture as Table 2. Across all indices only HAR specifications that include IV belong to the MCS with p-values larger than 10%. This indicates that such models are superior to the other HAR specifications. The results are consistent using both the range and semiquadratic statistics. This indicates that the IV is a strong predictor of future RV, confirming the previous finding about the importance of IV information (Busch et al., 2011). However, the simpler HAR-IV model belongs to the MCS only in one case—the DJIA index. For all the other indices, the HAR-IV is significantly outperformed by an extension of the model that accounts for one or more additional volatility stylized fact. Again, indicating that they contain incremental information about future realized volatility beyond that captured in HAR-IV.

In sum, the key result in Tables 2 and 3 demonstrates that a substantial predictive gain is driven by the inclusion of IV. In most cases, the parsimonious HAR-IV performs better than models that do not include IV even if they are more sophisticated, that is, contain more than one of the other stylized features of RV simultaneously. However, the evidence equally supports the inclusion of further extensions to the HAR-IV model.

4.1 | Does the strong predictive content of IV information change over time?

To further evaluate the role of IV in forecasting RV, we examine whether forecast performance of HAR-IV relative to the models that exclude IV, that is, HAR, HAR-L, HAR-O, HAR-G, HAR-LO, HAR-LG, HAR-OG, HAR-LOG, changes over time. We calculate the cumulative forecast error of these models based on the QLIKE loss function and use the Giacomini and Rossi (2010) test of local forecast performance. The cumulative forecast error, as defined in Equation (8), is plotted in Figure 1.[14] Here, we observe that the HAR-IV specification yields the lowest cumulative sum of the QLIKE errors. This indicates that HAR-IV produces the most accurate forecast consistently over time and, thus, including IV substantially improves the forecast performance of the HAR model. The only exceptions are for the AEX

---

[14]To facilitate visualizations, we only present the cumulative forecast errors of the HAR model (blue dashed line), HAR-IV (red line), HAR-L (orange line), HAR-O (green dashed-dotted line), and HAR-G (yellow line). The quality of the results does not change when the cumulative errors of HAR-LO, HAR-LG, HAR-OG, and HAR-LOG are plotted as well.

TABLE 3 Model confidence set (MCS) test

	S&P500		DJIA		Nasdaq100		STOXX50		CAC40		DAX		AEX		SMI		FTSE100		Nikkei225	
	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}	T_R	T_{SQ}
HAR	0.001	0.000	0.003	0.002	0.000	0.000	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000
HAR-IV	0.019	0.005	0.344**	0.014	0.000	0.000	0.007	0.001	0.002	0.000	0.000	0.000	0.001	0.002	0.003	0.003	0.028	0.023	0.040	0.010
HAR-L	0.001	0.000	0.003	0.002	0.000	0.000	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.007	0.003	0.011	0.001	0.000	0.000
HAR-O	0.001	0.000	0.004	0.003	0.000	0.000	0.007	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.000	0.000	0.000	0.000	0.006
HAR-G	0.001	0.000	0.003	0.002	0.000	0.000	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001
HAR-IVL	0.010	0.003	0.170**	0.014	0.016	0.000	0.007	0.003	0.002	0.001	0.000	0.000	0.003	0.001	0.018	0.004	1.000**	1.000**	0.013	0.009
HAR-IVO	0.030	0.014	0.368**	0.014	0.070	0.075	0.311**	0.219*	0.007	0.008	0.014	0.021	0.030	0.038	0.018	0.004	0.019	0.021	1.000**	1.000**
HAR-IVG	0.026	0.001	0.368**	0.014	0.000	0.000	0.007	0.001	0.002	0.000	0.000	0.000	0.000	0.001	0.003	0.000	0.011	0.007	0.040	0.013
HAR-LO	0.001	0.000	0.003	0.001	0.016	0.004	0.003	0.001	0.002	0.000	0.000	0.000	0.010	0.020	0.018	0.004	0.011	0.005	0.040	0.036
HAR-LG	0.001	0.000	0.003	0.002	0.000	0.000	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.007	0.003	0.011	0.004	0.000	0.000
HAR-OG	0.001	0.000	0.003	0.002	0.000	0.000	0.007	0.001	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.001	0.000	0.000	0.000	0.004
HAR-IVLO	0.026	0.001	0.344**	0.014	1.000**	1.000**	1.000**	1.000**	1.000**	1.000**	0.198*	0.198*	1.000**	1.000**	1.000**	1.000**	0.428**	0.428**	0.949**	0.904**
HAR-IVLG	0.026	0.001	0.344**	0.014	0.000	0.000	0.007	0.001	0.002	0.000	0.000	0.000	0.003	0.004	0.007	0.004	0.033	0.040	0.040	0.010
HAR-IVOG	1.000**	1.000**	1.000**	1.000**	0.070	0.047	0.188*	0.168*	0.007	0.009	0.034	0.040	0.030	0.034	0.018	0.004	0.011	0.006	0.949**	0.904**
HAR-LOG	0.001	0.000	0.003	0.002	0.016	0.003	0.003	0.001	0.002	0.000	0.000	0.000	0.010	0.017	0.007	0.004	0.011	0.003	0.040	0.010
HAR-IVLOG	0.030	0.014	0.368**	0.194*	0.038	0.194*	0.311**	0.219*	0.284**	0.284**	1.000**	1.000**	0.337**	0.337**	0.018	0.004	0.028	0.021	0.928**	0.880**

Note: Hansen et al. (2011) test in terms of the QLIKE criterion for our models across all indices. The *p*-values based on the range statistics (T_R) and the semiquadratic statistics (T_{SQ}) are obtained from 10,000 bootstraps. Low *p*-values indicate that it is unlikely for the model to belong to the set of the best models (MCS). The forecast evaluation period covers January 1, 2006–June 27, 2019.

Abbreviation: HAR, heterogeneous autoregressive.

*Indicates that the model belongs to the MCS at the 75% confidence level.

**Indicates that the model belongs to the MCS at the 90% confidence level.
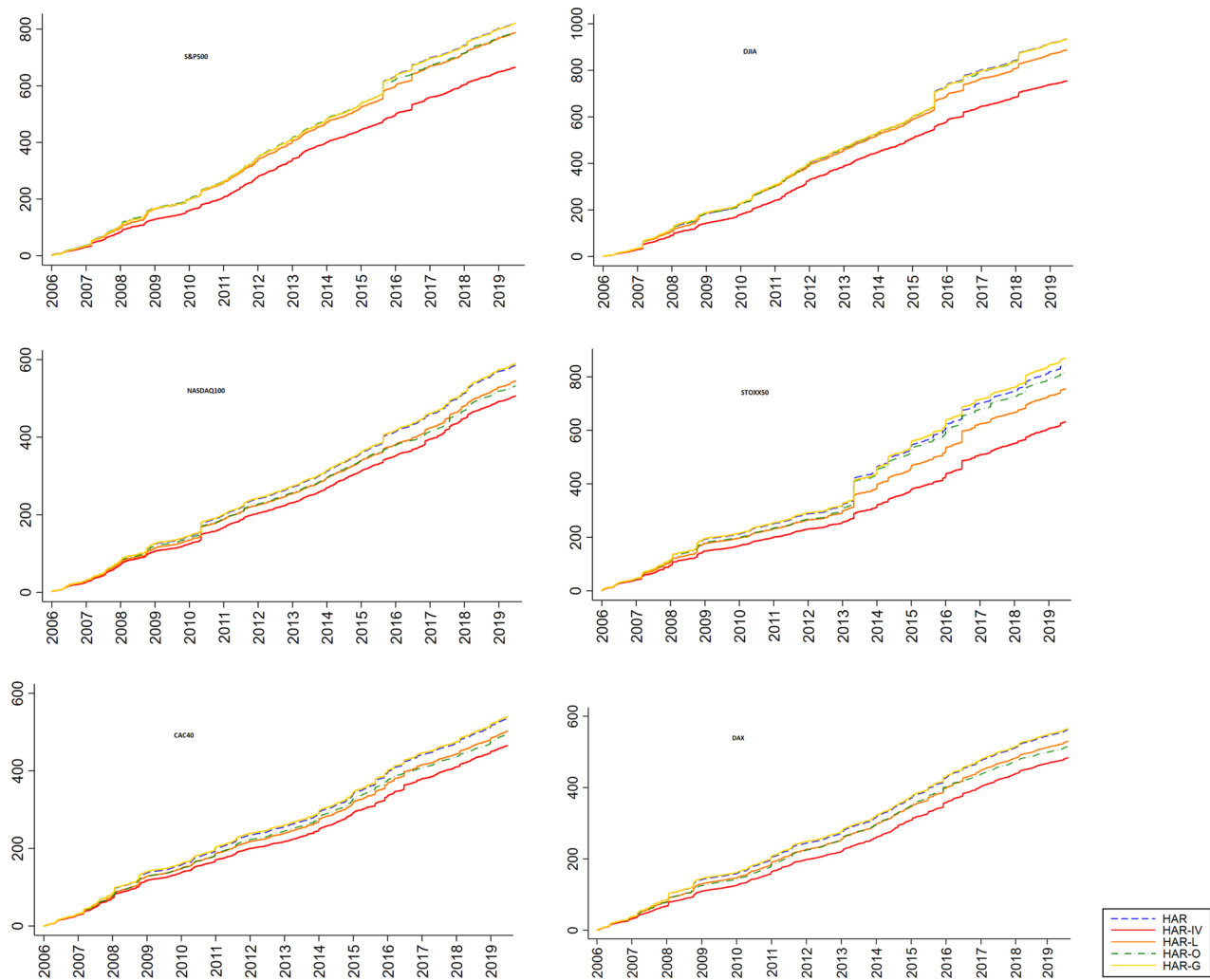
FIGURE 1 Cumulated sum of the QLIKE forecast error. *Note*: The figures present the cumulative QLIKE forecast errors of the HAR model (blue dashed line), HAR-IV (red line), HAR-L (orange line), HAR-O (green dashed-dotted line), and HAR-G (yellow line) for all stock markets as defined in Equation (8) over the forecast period (January 1, 2006–June 27, 2019). The results are computed on a rolling window using an initial in-sample estimating period that covers February 2, 2001–December 31, 2005. The model that yields the lowest cumulative QLIKE forecast errors produces the most accurate forecast over time. HAR, heterogeneous autoregressive; QLIKE, quasi-likelihood [Color figure can be viewed at [wileyonlinelibrary.com]

and SMI indices where the HAR-IV, HAR-L, and HAR-O models yield the lowest QLIKE errors interchangeably over time. We further find that the standard HAR and HAR-G specifications provide the worst forecasts over time.

Further graphical evidence of the superiority of HAR-IV is provided in Figure 2 based on the results of the Giacomini and Rossi (2010) test. This test compares the relative performance of two competing models over time. We evaluate the performance of HAR-IV against the standard HAR and HAR-L, HAR-O, HAR-G. The solid (blue) line is the difference between the QLIKE error of HAR-IV and one of the other HAR specifications. Giacomini and Rossi (2010) use this approach to test the null hypothesis of equal local predictive ability of the two forecasting models. The zero line indicates equal predictive ability and the dashed lines indicate the 5% critical values. The HAR-IV specification performs better (worse) than a competing model at the 5% level of significance when the solid line is below (above) the lower (upper) dashed line. For most indices the continuous line is clearly outside the critical values and below the lower dashed line. This means that the HAR-IV model consistently outperforms the standard HAR model and alternative HAR models that include either leverage effect, overnight returns or the volatility of realized volatility. In some periods of time the null hypothesis of equal predictive ability is not rejected as the solid line lies between the two dashed lines. In a limited number of cases, one of the competing models performs significantly better than the HAR-IV. For example, for the AEX index, the HAR-IV is significantly worse than its competitors between mid-2011 and 2013. Nonetheless, the overall superiority of the HAR-IV model is consistent across time.
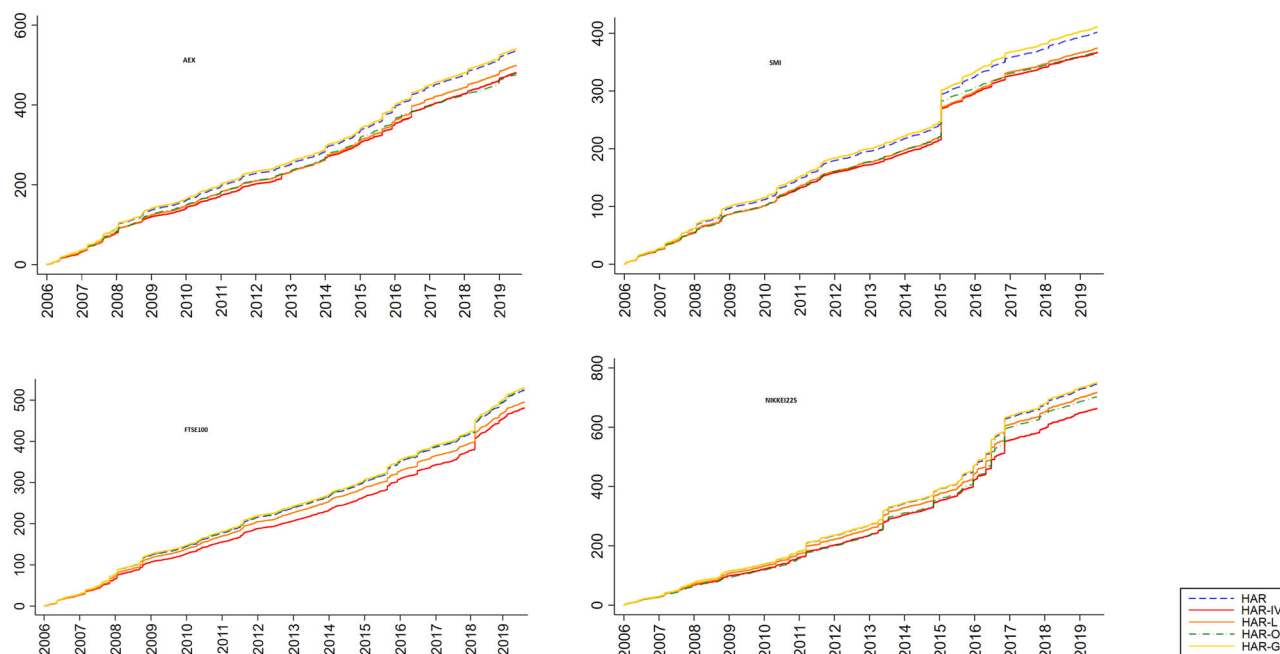
FIGURE 1 (Continued)

4.2 | Do leverage effects, overnight returns, and the volatility of realized volatility enhance the forecast result?

From the results reported in Tables 2 and 3, and Figures 1 and 2 it is evident that IV is a strong predictor of future realized volatility across all the markets we analyze. This result confirms the previous finding about the importance of IV information for the US market (Busch et al., 2011). The HAR-IV model outperforms not only the standard HAR model, but also other HAR models that capture well-known features of realized volatility, such as the leverage effect, overnight returns, and the volatility of realized volatility, but ignore IV. However, the results above provide evidence in support of extensions to the HAR-IV. In this subsection we further investigate the predictive performance of the HAR-IV model and extensions accounting for the noted features.

Table 4 presents the Giacomini and White pairwise test results for all indices. The *p*-values reported in the table are based on the mean difference between the row model and the column model. The null hypothesis of equal forecasting performance between the row and column models is tested in terms of QLIKE forecast error. The sign in parenthesis indicates which model performs best. A positive sign shows that the column model is significantly preferred, that is, the row model forecast yields a larger loss than the column model forecast. Similarly, a negative sign denotes that the row model forecast performs significantly better than the column model forecast, with the latter producing a larger loss.

Several interesting conclusions emerge from Table 4. Augmenting HAR-IV with overnight returns significantly enhances forecasting accuracy for all indices with the exception of the DJIA and FTSE100 indices. Incorporating the volatility of realized volatility in the HAR-IV significantly improves the forecast performance of the model only for the S&P500 and DJIA index. This result is in line with Corsi et al. (2008) who show that allowing for the time-varying volatility of realized volatility leads to an improvement in the predictive performance of the standard HAR for the S&P500. However, the results show that this is not the case for any of the other markets we consider. For all the other indices, the HAR-IV outperforms the HAR-IVG. In some cases, the null hypothesis of equal predictive ability between HAR-IV and HAR-IVG is rejected at the 5% level indicating that the former provides significantly better forecasts than the latter. In other cases, the null hypothesis of equal predictive ability between the two models is not rejected, which indicates that the volatility of realized volatility does not contain any significant additional information.

Mixed results are found on the role of the leverage effect. Adding the leverage effect, one of the most salient features of realized volatility, to the HAR-IV significantly improves the forecasting performance in only two cases—for the SMI and FTSE100 indices. For the S&P500 the HAR-IV is significantly better than HAR-IVL. For all other indices, augmenting HAR-IV with leverage effect does not significantly enhance the predictive power. Thus, the leverage effect does not appear
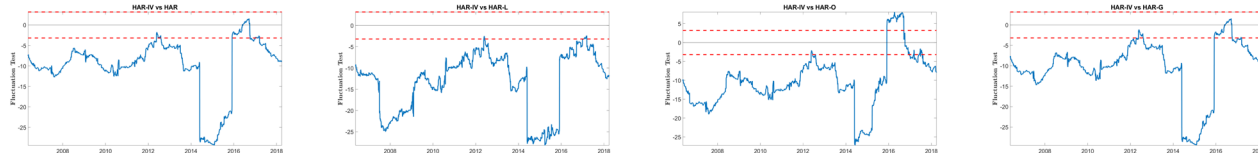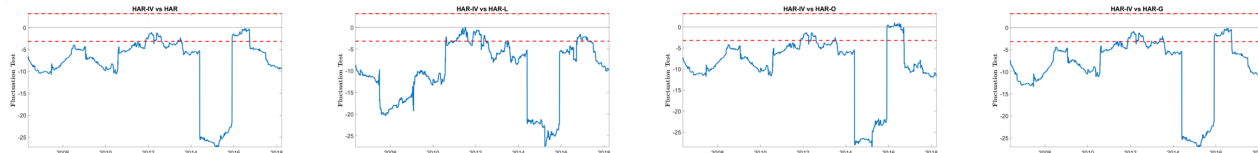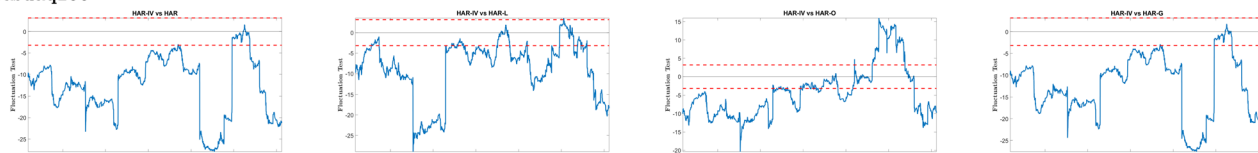
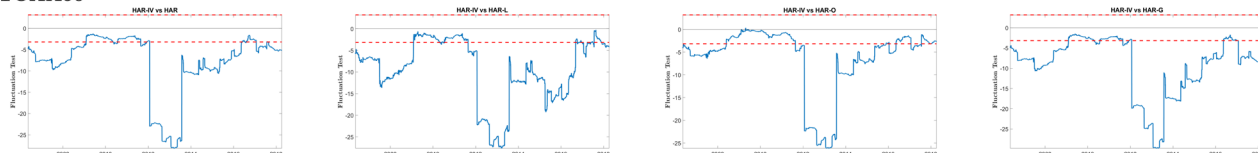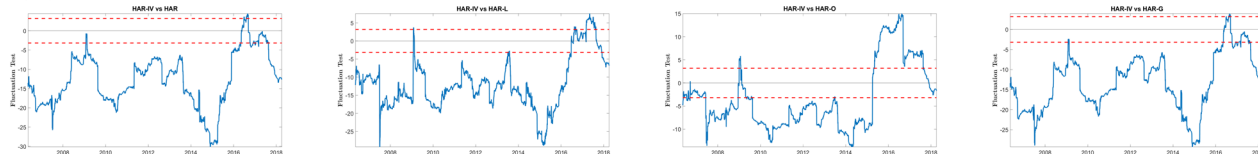**FIGURE 2** Fluctuation test of Giacomini and Rossi (2010). *Note*: The figure reports Giacomini and Rossi's (2010) fluctuation test statistic for comparing forecasts of HAR-IV model relative to either HAR, HAR-L, HAR-O, or HAR-G for all stock markets. The test examines the stability over time of out-of-sample relative forecasting performance of a pair of models in an unstable environment. The null hypothesis is that two models have equal predictive ability at each point in time estimated over a rolling window of data. The zero line indicates equal predictive ability and the dashed lines indicate the 5% critical values. The solid blue lines indicate the fluctuation test statistic. The HAR-IV specification is significantly better (worse) than one of the competing models at the 5% level when the solid line is below (above) the lower (upper) dashed line. HAR, heterogeneous autoregressive [Color figure can be viewed at [wileyonlinelibrary.com]

to play a significant role in forecasting realized volatility. An explanation for this is that the IV indices reflect the leverage effect. IV may capture the leverage effect given the relation between leverage and volatility and changes in leverage on the volatility smile found in the option literature. This is also complemented by the observed negative empirical relation between VIX and stock returns (e.g., Giot, 2005). Notably, several researchers identify a link between leverage and IV. This includes Branger and Schlag (2004) and Ciliberti et al. (2009) who link the shape of the volatility smile to leverage. Further, Choi and Richardson (2016) demonstrate a positive relation between leverage and option IV. They further note that the asymmetric effect in an estimated EGARCH model arises from both a volatility feedback effect as well as leverage.[15]

We further augment the HAR-IV model to account simultaneously for more than one stylized fact. We find that the most sophisticated HAR-IV model, that is, the HAR-IV model that accounts simultaneously for the leverage effect,

---

[15]Black (1976) first observed the negative correlation between stock returns and volatility. While this was subsequently coined the leverage effect (Christie, 1982). The arrival of bad news results in a decline in a firm's stock price and an increase in its leverage ratio. This causes investors to revise upwards their perception of riskiness of the firm and the resultant trading increases volatility. The volatility feedback effect provides an alternative explanation for the asymmetry between returns and volatility (see also, Campbell and Hentschel, 1992). The arrival of news causes stock prices to move, resulting in volatility (an increase in the standard deviation of returns). The increased volatility in turn causes investors to revise upwards their expectations of future volatility and expected returns, putting downward pressure on current prices and returns. This magnifies the effects of negative news and dampens that of positive news, resulting in the negative correlation between returns and volatility.

FIGURE 2 (Continued)

overnight returns, and the volatility of volatility (HAR-IVLOG) does not significantly outperform more restricted specifications across all markets, which means that on average a more parsimonious model is superior. However, no single extension of HAR-IV outperforms all others across the markets we study. Augmenting HAR-IVO with leverage effect (HAR-IVLO) provides the best forecast model for the Nasdaq, CAC40, DAX, and SMI index. That is, for these indices, the HAR-IVLO model performs significantly better than all other competing models. The HAR-IVOG produces the best forecast for the S&P500 and DJIA, while for the remaining indices, an even more parsimonious model provides the best forecast. The HAR-IVO performs best for the STOXX50, AEX, and Nikkei225, and the HAR-IVL is the preferred model for the FTSE100.

Overall, the results regarding the interaction between IV and the stylized facts of volatility indicate that all markets support some extensions of the HAR-IV to include one or more of the stylized fact of volatility. Looking at the best performing HAR-IV specifications we conclude that overnight returns are beneficial for nine out of 10 indices in the sense that its inclusion improves the accuracy of the predictive model. As found by Corsi et al. (2008), allowing for the time-varying volatility of realized volatility helps predict volatility in the United States. However, this result does not extend to any other market we consider. Mixed results are found for the leverage effect which improves predictions only in five markets.

5 | VAR FORECASTS

The above results suggest that HAR models that incorporate IV provide better forecasting performance than any model that excludes IV, even if it does include one or more volatility stylized fact. The results also indicate that, for all markets, an extension of the HAR-IV to include these stylized facts does improve forecasts. In this section, the performance of the different forecasting models is assessed in a risk management context, using the VaR approach.

TABLE 4 Giacomini–White test

	HAR-IVL	HAR-IVO	HAR-IVG	HAR-IVLO	HAR-IVLG	HAR-IVOG	HAR-IVLOG
S&P500							
HAR-IV	0.010(−)	0.014(+)	0.000(+)	0.060(+)	0.243	0.008(+)	0.033(+)
HAR-IVL		0.005(+)	0.000(+)	0.019(+)	0.000(+)	0.003(+)	0.011(+)
HAR-IVO			0.029(−)	0.011(−)	0.011(−)	0.003(+)	0.101
HAR-IVG				0.110	0.011(−)	0.016(+)	0.063(+)
HAR-IVLO					0.037(−)	0.004(+)	0.003(+)
HAR-IVLG						0.006(+)	0.020(+)
HAR-IVOG							0.021(−)
Nasdaq100							
HAR-IV	0.236	0.000(+)	0.003(−)	0.000(+)	0.452	0.000(+)	0.000(+)
HAR-IVL		0.000(+)	0.106	0.000(+)	0.013(−)	0.000(+)	0.000(+)
HAR-IVO			0.000(−)	0.049(+)	0.000(−)	0.014(−)	0.062(+)
HAR-IVG				0.000(+)	0.233	0.000(+)	0.000(+)
HAR-IVLO					0.000(−)	0.050(−)	0.000(−)
HAR-IVLG						0.000(+)	0.000(+)
HAR-IVOG							0.041(+)
CAC40							
HAR-IV	0.121	0.005(+)	0.138	0.001(+)	0.069(+)	0.005(+)	0.001(+)
HAR-IVL		0.016(+)	0.078(−)	0.004(+)	0.498	0.017(+)	0.004(+)
HAR-IVO			0.003(−)	0.013(+)	0.014(−)	0.148	0.015(+)
HAR-IVG				0.001(+)	0.044(+)	0.004(+)	0.001(+)
HAR-IVLO					0.003(−)	0.019(−)	0.241
HAR-IVLG						0.015(+)	0.004(+)
HAR-IVOG							0.013(+)
AEX							
HAR-IV	0.105	0.001(+)	0.696	0.000(+)	0.108	0.001(+)	0.006(+)
HAR-IVL		0.031(+)	0.072(−)	0.002(+)	0.480	0.024(+)	0.001(+)
HAR-IVO			0.001(−)	0.114	0.022(−)	0.489	0.126
HAR-IVG				0.000(+)	0.057(+)	0.001(+)	0.000(+)
HAR-IVLO					0.001(−)	0.075(−)	0.001(+)
HAR-IVLG						0.705	0.018(+)
HAR-IVOG							0.001(+)

	HAR-IVL	HAR-IVO	HAR-IVG	HAR-IVLO	HAR-IVLG	HAR-IVOG	HAR-IVLOG
DJIA							
HAR-IV	0.308	0.677	0.000(+)	0.956	0.399	0.454	0.954
HAR-IVL		0.346	0.125	0.646	0.004(+)	0.216	0.483
HAR-IVO			0.875	0.289	0.467	0.000(+)	0.375
HAR-IVG				0.875	0.247	0.660	0.974
HAR-IVLO					0.793	0.129	0.001(+)
HAR-IVLG						0.302	0.617
HAR-IVOG							0.245
STOXX50							
HAR-IV	0.281	0.000(+)	0.086(+)	0.000(+)	0.208	0.013(+)	0.005(+)
HAR-IVL		0.001(+)	0.160	0.000(+)	0.068(+)	0.047(+)	0.015(+)
HAR-IVO			0.000(−)	0.109	0.000(−)	0.071(−)	0.035(−)
HAR-IVG				0.000(+)	0.965	0.000(+)	0.000(+)
HAR-IVLO					0.000(−)	0.088(−)	0.068(−)
HAR-IVLG						0.000(+)	0.000(+)
HAR-IVOG							0.214
DAX							
HAR-IV	0.105	0.000(+)	0.937	0.000(+)	0.089(+)	0.000(+)	0.000(+)
HAR-IVL		0.000(+)	0.119	0.000(+)	0.752	0.000(+)	0.000(+)
HAR-IVO			0.000(−)	0.077(+)	0.000(−)	0.146	0.024(+)
HAR-IVG				0.000(+)	0.087(+)	0.000(+)	0.000(+)
HAR-IVLO					0.000(−)	0.155	0.613
HAR-IVLG						0.000(+)	0.000(+)
HAR-IVOG							0.051(+)
SMI							
HAR-IV	0.002(+)	0.001(+)	0.000(−)	0.003(+)	0.023(+)	0.002(+)	0.003(+)
HAR-IVL		0.001(+)	0.000(−)	0.020(+)	0.001(−)	0.001(−)	0.006(+)
HAR-IVO			0.001(−)	0.007(+)	0.001(−)	0.011(−)	0.026(+)
HAR-IVG				0.000(+)	0.001(+)	0.001(+)	0.000(+)
HAR-IVLO					0.010(−)	0.002(−)	0.007(−)
HAR-IVLG						0.001(+)	0.004(+)

TABLE 4 (Continued)

	HAR-IVL	HAR-IVO	HAR-IVG	HAR-IVLO	HAR-IVLG	HAR-IVOG	HAR-IVLOG
HAR-IVOG							0.081(+)
FTSE100							
HAR-IV	0.085(+)	0.561	0.002(−)	0.141	0.214	0.004(−)	0.307
HAR-IVL		0.063(−)	0.019(−)	0.378	0.021(−)	0.013(−)	0.029(−)
HAR-IVO			0.052(−)	0.088(+)	0.164	0.001(−)	0.212
HAR-IVG				0.035(+)	0.033(+)	0.454	0.066(+)
HAR-IVLO					0.236	0.018(−)	0.027(−)
HAR-IVLG						0.022(−)	0.464
HAR-IVOG							0.029(+)

	HAR-IVL	HAR-IVO	HAR-IVG	HAR-IVLO	HAR-IVLG	HAR-IVOG	HAR-IVLOG
HAR-IVOG							0.007(+)
Nikkei225							
HAR-IV	0.757	0.001(+)	0.894	0.006(+)	0.696	0.000(+)	0.004(+)
HAR-IVL		0.000(+)	0.814	0.001(+)	0.492	0.000(+)	0.001(+)
HAR-IVO			0.001(−)	0.615	0.000(−)	0.403	0.552
HAR-IVG				0.007(+)	0.737	0.001(+)	0.006(+)
HAR-IVLO					0.002(−)	0.640	0.515
HAR-IVLG						0.000(+)	0.001(+)
HAR-IVOG							0.645

Note: Giacomini and White (2006) test. The null hypothesis that the row model and column model perform equally well is tested in terms of QLIKE forecast error. The superscripts + and − indicate rejection of the null hypothesis, with a positive (negative) sign denoting that the row (column) model is outperformed by the column (row) model. The forecast evaluation period covers from January 1, 2006 to June 27, 2019.

Abbreviations: HAR, heterogeneous autoregressive; QLIKE, quasi-likelihood.

TABLE 5 Results of backtesting the 5%VaR estimates across all indices

		S&P500	DJIA	Nasdaq100	STOXX50	CAC40	DAX	AEX	SMI	FTSE100	Nikkei225
HAR	VR	0.049	0.049	0.054	0.048	0.053	0.053	0.053	0.053	0.042	0.067
	LR_{uc}	0.834	0.834	0.265	0.660	0.374	0.378	0.459	0.441	0.022**	0.000***
	LR_{cc}	0.800	0.800	0.425	0.122	0.057**	0.618	0.027**	0.023**	0.050**	0.000***
HAR-IV	VR	0.051	0.049	0.051	0.047	0.052	0.052	0.050	0.052	0.045	0.067
	LR_{uc}	0.793	0.773	0.675	0.449	0.669	0.620	0.910	0.591	0.221	0.000***
	LR_{cc}	0.553	0.449	0.487	0.374	0.267	0.883	0.481	0.147	0.359	0.000***
HAR-L	VR	0.054	0.052	0.055	0.048	0.054	0.054	0.052	0.051	0.044	0.066
	LR_{uc}	0.336	0.619	0.178	0.660	0.263	0.234	0.558	0.762	0.087*	0.000***
	LR_{cc}	0.390	0.815	0.026**	0.908	0.514	0.493	0.825	0.226	0.226	0.000***
HAR-O	VR	0.045	0.047	0.044	0.044	0.043	0.044	0.045	0.048	0.041	0.052
	LR_{uc}	0.183	0.353	0.135	0.085*	0.060*	0.082*	0.142	0.570	0.018**	0.539
	LR_{cc}	0.298	0.405	0.163	0.099*	0.167	0.217	0.245	0.087*	0.040**	0.828
HAR-G	VR	0.049	0.049	0.055	0.048	0.054	0.054	0.053	0.054	0.043	0.067
	LR_{uc}	0.896	0.834	0.233	0.660	0.297	0.234	0.414	0.316	0.041**	0.000***
	LR_{cc}	0.830	0.800	0.379	0.219	0.059*	0.411	0.013**	0.026**	0.066	0.000***
HAR-IVL	VR	0.051	0.050	0.052	0.049	0.054	0.051	0.052	0.050	0.045	0.066
	LR_{uc}	0.733	0.978	0.565	0.778	0.297	0.854	0.666	0.884	0.221	0.000***
	LR_{cc}	0.521	0.634	0.109	0.929	0.369	0.946	0.876	0.711	0.442	0.000***
HAR-IVO	VR	0.046	0.046	0.044	0.041	0.042	0.045	0.044	0.045	0.045	0.050
	LR_{uc}	0.313	0.242	0.097*	0.014**	0.034**	0.159	0.087*	0.145	0.192	0.827
	LR_{cc}	0.529	0.458	0.136	0.034**	0.099*	0.335	0.229	0.162	0.315	0.614
HAR-IVG	VR	0.050	0.049	0.052	0.047	0.052	0.052	0.050	0.053	0.046	0.066
	LR_{uc}	0.853	0.713	0.675	0.403	0.510	0.620	0.966	0.489	0.326	0.000***
	LR_{cc}	0.583	0.456	0.487	0.481	0.281	0.883	0.350	0.045	0.507	0.000***
HAR-LO	VR	0.047	0.048	0.047	0.043	0.045	0.043	0.045	0.046	0.044	0.052
	LR_{uc}	0.397	0.598	0.353	0.059*	0.219	0.057*	0.142	0.258	0.087*	0.648
	LR_{cc}	0.688	0.869	0.108	0.166	0.423	0.139	0.341	0.294	0.226	0.510

—WILEY—| **1635**

TABLE 5 (Continued)

		S&P500	DJIA	Nasdaq100	STOXX50	CAC40	DAX	AEX	SMI	FTSE100	Nikkei225
HAR-LG	VR	0.054	0.051	0.055	0.047	0.054	0.054	0.052	0.051	0.044	0.066
	LR_{uc}	0.336	0.733	0.204	0.498	0.263	0.234	0.611	0.762	0.103	0.000***
	LR_{cc}	0.390	0.737	0.031	0.793	0.514	0.493	0.853	0.366	0.261	0.000***
HAR-OG	VR	0.045	0.047	0.046	0.043	0.043	0.044	0.046	0.048	0.043	0.052
	LR_{uc}	0.157	0.443	0.276	0.071*	0.072*	0.082	0.221	0.626	0.041**	0.539
	LR_{cc}	0.258	0.498	0.223	0.081*	0.125	0.217	0.175	0.049	0.066*	0.828
HAR-IVLO	VR	0.046	0.046	0.044	0.042	0.043	0.045	0.045	0.044	0.045	0.050
	LR_{uc}	0.313	0.313	0.082*	0.027**	0.050**	0.136	0.142	0.105	0.192	0.952
	LR_{cc}	0.233	0.529	0.063*	0.055*	0.145	0.293	0.251	0.265	0.393	0.425
HAR-IVLG	VR	0.051	0.050	0.052	0.049	0.053	0.051	0.051	0.051	0.045	0.065
	LR_{uc}	0.793	0.959	0.619	0.778	0.374	0.734	0.842	0.823	0.192	0.000***
	LR_{cc}	0.553	0.411	0.121	0.796	0.397	0.944	0.754	0.530	0.393	0.000***
HAR-IVOG	VR	0.046	0.044	0.044	0.041	0.043	0.045	0.044	0.045	0.046	0.051
	LR_{uc}	0.276	0.115	0.135	0.011**	0.072*	0.159	0.087*	0.145	0.288	0.827
	LR_{cc}	0.370	0.227	0.163	0.028**	0.168	0.370	0.229	0.162	0.456	0.614
HAR-LOG	VR	0.047	0.047	0.047	0.043	0.045	0.043	0.045	0.047	0.044	0.052
	LR_{uc}	0.397	0.492	0.353	0.059*	0.219	0.069*	0.192	0.374	0.103	0.648
	LR_{cc}	0.688	0.766	0.108	0.140	0.318	0.189	0.426	0.560	0.261	0.510
HAR-IVLOG	VR	0.046	0.046	0.044	0.041	0.043	0.045	0.045	0.044	0.045	0.050
	LR_{uc}	0.276	0.313	0.135	0.014**	0.050**	0.136	0.142	0.105	0.192	0.952
	LR_{cc}	0.370	0.529	0.163	0.034**	0.142	0.293	0.251	0.265	0.393	0.425

Note: Entries report the violation rate (VR) and the *p*-values of Kupiec's likelihood ratio test of unconditional coverage (LR_{uc}) and Christoffersen's likelihood ratio test of conditional coverage (LR_{cc}) for each model. A violation occurs if the empirical return exceeds the predicted VaR on a particular day. The expected value of the VaR violation ratio is the corresponding tail size, that is, 5%. *, **, and *** denote the significance at 10%, 5%, and 1% levels, respectively.

Abbreviations: HAR, heterogeneous autoregressive; VaR, value-at-risk.

VaR is a popular technique designed to quantify and assess risk in financial markets and its estimation is widespread within financial risk management. The goal of VaR is to provide the most likely loss for a portfolio that could occur with a given probability over a given time horizon. Therefore, VaR can be defined as $Pr(r_{t+1} < -VaR_{t+1|t} = a)$. It also allows for an economic-based evaluation of the alternate realized volatility forecast models. The VaR of a portfolio is estimated as $VaR_{t+1|t} = F(a)RV_{t+1|t}$, where $F(a)$ is the left quantile at $a\%$ (e.g., 5%) for the assumed distribution[16] and $RV_{t+1|t}$ is the one-step-ahead realized volatility forecast.

The out-of-sample performance of the realized volatility models to produce reasonable VaR estimates is assessed by computing the violation rate (VR). That is, the frequency that the actual loss exceeds the predicted VaR. If a VaR model is correctly specified, the VR should be equal to the prespecified VaR level a (in our application a is 5%). The Kupiec (1995) unconditional coverage statistic tests the null hypothesis that the observed VR is statistically equal to the expected one, a. The null hypothesis will be rejected wherever the observed VR is statistically different to the expected VR.

Besides the VR, a relevant VaR model should also consider the cluster of the violation. Christoffersen (1998) introduces a conditional coverage statistic that tests the joint hypothesis of unconditional coverage and the independence of violations. That is, in a correctly specified VaR model not only should the violation occur at a specified rate, but they should also occur independently and spread across the whole estimation period.

The results for the one-step-ahead VaR forecasting across all indices for all the HAR specifications are reported in Table 5. The results show that there is not a single model that passes the backtesting tests across all indices. Also, the optimal model varies across indices. This result is consistent with findings in the literature that VaR models are not robust across different markets (Degiannakis et al., 2013). According to the VR, very few models produce a VR equal to the expected one (5%). In some cases the observed VR exceeds the expected VR indicating an excessive underestimation of risk, while in other cases the observed VR is less than the expected one indicating an excessive overestimation of risk. However, and according to the Kupiec (1995) and Christoffersen (1998) test, in many cases the observed VR is not statistically different to the expected VR and the VaR violations are independently distributed. Also, the VR shows that in most cases when IV is added in a HAR specification the VR approaches closer to its expected one. This result indicates that models that include IV appear to improve the forecast accuracy of VaR. Taking into consideration the models that pass the backtesting tests, and the models that produce a VR equal or closer to the expected one, these can be considered to produce the most efficient VaR. Specifically, the HAR-IV model produces the most accurate estimates for the Nasdaq, CAC40 and AEX index, the HAR-IVL performs best for the DJIA, STOXX50, DAX, and SMI index, HAR-IVG produces the most efficient VaR for the S&P500 and FTSE100 index and HAR-IVO is the best performing model for the Nikkei225.

Overall, and consistent with the results presented in Section 4, models that include IV appear to improve the forecast accuracy of VaR. Also, a more parsimonious model seems to produce better VaR estimates than a more sophisticated specification. Thus, a HAR model that accounts for IV and, in most cases, one of the other stylized facts produces the most accurate VaR estimates.

6 | CONCLUSION

This paper is motivated by two strands of literature, the one showing the predictive power of IV over realized volatility, the other proposing extensions to the popular HAR model of Corsi (2009). Considering such extensions within a HAR model that also includes IV, we examine the role of the leverage effect, overnight returns, and the volatility of realized volatility in 10 international stock markets.

We find internationally robust evidence that IV plays an important role in volatility forecasting as the HAR-IV model—that is, the standard HAR augmented with IV—provides significantly better forecasts than any HAR variant model that excludes IV. This result is consistent over time. By considering the above-mentioned extensions to HAR-IV, we shed new light on the way HAR-type models account for well-known volatility stylized facts. No single extension of HAR-IV is more accurate across all markets, but all markets support some extensions. Overnight returns are notably beneficial for the majority of the markets, while the volatility of realized volatility is only beneficial for two US indices. Mixed results are found on the leverage effect, which is beneficial for just five markets. We find similar results in a risk management environment by performing a VaR forecasting exercise.

---

[16]In this exercise, $F(a) = st_{a,\upsilon}$, that is, the left ath quantile of the Student-t distribution with υ degrees of freedom.

ORCID

Dimos S. Kambouroudis https://orcid.org/0000-0002-8230-0028
David G. McMillan https://orcid.org/0000-0002-5891-4193
Katerina Tsakou http://orcid.org/0000-0003-1913-858X

REFERENCES

Andersen, T. G., & Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review, 39*(4), 885–905.

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., & 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.

Barndorff-Nielsen, O. E., & Shephard, N. (2004). Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics, 2*(1), 1–37.

Barndorff-Nielsen, O. E., & Shephard, N. (2005). Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics, 4*(1), 1–30.

Becker, R., Clements, A. E., & White, S. I. (2007). Does implied volatility provide any information beyond that captured in model-based volatility forecasts? *Journal of Banking & Finance, 31*(8), 2535–2549.

Black, F. (1976). Studies of stock market volatility changes. *Proceedings of the American Statistical Association Business and Economic Statistics Section*. 177–181.

Blair, B. J., Poon, S.-H., & Taylor, S. J. (2001). Forecasting S&P 100 volatility: The incremental information content of implied volatilities and high-frequency index returns. *Journal of Econometrics, 105*(1), 5–26.

Bollerslev, T., Hood, B., Huss, J., & Pedersen, L. H. (2018). Risk everywhere: Modeling and managing volatility. *The Review of Financial Studies, 31*(7), 2729–2773.

Bollerslev, T., Tauchen, G., & Zhou, H. (2009). Expected stock returns and variance risk premia. *The Review of Financial Studies, 22*(11), 4463–4492.

Branger, N., & Schlag, C. (2004). Why is the index smile so steep? *Review of Finance, 8*(1), 109–127.

Brownlees, C., Engle, R., & Kelly, B. (2011). A practical guide to volatility forecasting through calm and storm. *Journal of Risk, 14*(2), 3–22.

Buncic, D., & Gisler, K. I. (2016). Global equity market volatility spillovers: A broader role for the United States. *International Journal of Forecasting, 32*(4), 1317–1339.

Buncic, D., & Gisler, K. I. (2017). The role of jumps and leverage in forecasting volatility in international equity markets. *Journal of International Money and Finance, 79*, 1–19.

Busch, T., Christensen, B. J., & Nielsen, M. Ø. (2011). The role of implied volatility in forecasting future realized volatility and jumps in foreign exchange, stock, and bond markets. *Journal of Econometrics, 160*(1), 48–57.

Campbell, J. Y., & Hentschel, L. (1992). No news is good news: An asymmetric model of changing volatility in stock returns. *Journal of Financial Economics, 31*(3), 281–318.

Chiras, D. P., & Manaster, S. (1978). The information content of option prices and a test of market efficiency. *Journal of Financial Economics, 6*(2), 213–234.

Choi, J., & Richardson, M. (2016). The volatility of a firm's assets and the leverage effect. *Journal of Financial Economics, 121*(2), 254–277.

Christensen, B., & Prabhala, N. (1998). The relation between implied and realized volatility. *Journal of Financial Economics, 50*(2), 125–150.

Christie, A. A. (1982). The stochastic behavior of common stock variances: Value, leverage and interest rate effects. *Journal of Financial Economics, 10*(4), 407–432.

Christoffersen, P. F. (1998). Evaluating Interval Forecasts. *International Economic Review, 39*(4), 841–862.

Ciliberti, S., Bouchaud, J.-P., & Potters, M. (2009). Smile dynamics: A theory of the implied leverage effect. *Wilmott Journal*, *1*(2), 87–94.

Corrado, C. J., & Miller, T. W., Jr. (2005). The forecast quality of CBOE implied volatility indexes. *Journal of Futures Markets*, *25*(4), 339–373.

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 realized volatility. *Econometric Reviews*, *27*(1–3), 46–78.

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., Floros, C., & Dent, P. (2013). Forecasting value-at-risk and expected shortfall using fractionally integrated models of conditional volatility: International evidence. *International Review of Financial Analysis*, *27*, 21–33.

Fernandes, M., Medeiros, M. C., & Scharth, M. (2014). Modeling and predicting the CBOE market volatility index. *Journal of Banking & Finance*, *40*, 1–10.

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.

Gallo, G. M. (2001). Modelling the impact of overnight surprises on intra-daily volatility. *Australian Economic Papers*, *40*(4), 567–580.

Giacomini, R., & Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, *25*(4), 595–620.

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*(6), 1545–1578.

Giot, P. (2005). Relationships between implied volatility indexes and stock index returns. *The Journal of Portfolio Management*, *31*(3), 92–100.

Gong, X., & Lin, B. (2018). Structural breaks and volatility forecasting in the copper futures market. *Journal of Futures Markets*, *38*(3), 290–339.

Hansen, P. R., & Lunde, A. (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics*, *24*(2), 127–161.

Hansen, P. R., Lunde, A., & Nason, J. M. (2011). The model confidence set. *Econometrica*, *79*(2), 453–497.

Heber, G., Lunde, A., Shephard, N., & Sheppard, K. (2009). *Oxford-Man Institute's realized library, version 0.3*. Oxford-Man Institute, University of Oxford. http://realized.oxford-man.ox.ac.uk/data/download

Jeon, B., Seo, S. W., & Kim, J. S. (2020). Uncertainty and the volatility forecasting power of option-implied volatility. *Journal of Futures Markets*, *40*(7), 1109–1126.

Jiang, G. J., & Tian, Y. S. (2005). The model-free implied volatility and its information content. *The Review of Financial Studies*, *18*(4), 1305–1342.

Kambouroudis, D. S., & McMillan, D. G. (2016). Does VIX or volume improve GARCH volatility forecasts? *Applied Economics*, *48*(13), 1210–1228.

Kambouroudis, D. S., McMillan, D. G., & Tsakou, K. (2016). Forecasting stock return volatility: A comparison of GARCH, implied volatility, and realized volatility models. *Journal of Futures Markets*, *36*(12), 1127–1163.

Koopman, S. J., Jungbacker, B., & Hol, E. (2005). Forecasting daily variability of the S&P100 stock index using historical, realised and implied volatility measurements. *Journal of Empirical Finance*, *12*(3), 445–475.

Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, *3*(2), 73–84.

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.

Martens, M., van Dijk, D., & de Pooter, M. (2009). Forecasting S&P500 volatility: Long memory, level shifts, leverage effects, day-of-the-week seasonality, and macroeconomic announcements. *International Journal of Forecasting*, *25*(2), 282–303.

Martens, M., & Zein, J. (2004). Predicting financial volatility: High-frequency time-series forecasts vis-à-vis implied volatility. *Journal of Futures Markets*, *24*(11), 1005–1028.

Muller, U. A., Dacorogna, M. M., Dave, R. D., Olsen, R., Pictet, O. V., & von Weizsacker, J. (1997). Volatilities of different time resolutions—Analyzing the dynamics of market components. *Journal of Empirical Finance*, *4*(2–3), 213–239.

Newey, W. K., & West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, *55*, 703–708.

Oikonomou, I., Stancu, A., Symeonidis, L., & Wese Simen, C. (2019). The information content of short-term options. *Journal of Financial Markets*, *46*, 100504. https://doi.org/10.1016/j.finmar.2019.07.003

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

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.

Plíhal, T., & Lyócsa, Š. (2021). Modeling realized volatility of the EUR/USD exchange rate: Does implied volatility really matter? *International Review of Economics & Finance*, *71*, 811–829.

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. (2016). Do jumps matter for volatility forecasting? Evidence from energy markets. *Journal of Futures Markets*, *36*(8), 758–792.

Sévi, B. (2014). Forecasting the volatility of crude oil futures using intraday data. *European Journal of Operational Research*, *235*(3), 643–659.

Thomson Reuters Datastream. (2019). *Daily open and close prices, and implied volatility indices from February 2001 to June 2019*.

Tian, F., Yang, K., & Chen, L. (2017). Realized volatility forecasting of agricultural commodity futures using the HAR model with time-varying sparsity. *International Journal of Forecasting*, *33*(1), 132–152.

Tsiakas, I. (2008). Overnight information and stochastic volatility: A study of European and US stock exchanges. *Journal of Banking & Finance*, *32*(2), 251–268.

Wang, X., Wu, C., & Xu, W. (2015). Volatility forecasting: The role of lunch-break returns, overnight returns, trading volume and leverage effects. *International Journal of Forecasting*, *31*(3), 609–619.

Yu, W. W., Lui, E. C., & Wang, J. W. (2010). The predictive power of the implied volatility of options traded OTC and on exchanges. *Journal of Banking & Finance*, *34*(1), 1–11.

Zhang, Y., Ma, F., & Liao, Y. (2020). Forecasting global equity market volatilities. *International Journal of Forecasting*, *36*(4), 1454–1475.

How to cite this article: Kambouroudis, D. S., McMillan, D. G., & Tsakou, K. (2021). Forecasting realized volatility: The role of implied volatility, leverage effect, overnight returns, and volatility of realized volatility. *Journal of Futures Markets*, 41, 1618–1639. https://doi.org/10.1002/fut.22241